

**Títol del treball:**

**Càlcul acurat de constants d'acoblament spin-spin i aplicació a la determinació de l'estructura secundària de pèptids**

---

Estudiant: Gerard Comas i Vilà

Grau en Química

Correu electrònic: gerard.comas7@gmail.com

Tutor: Pedro Salvador Sedano

Empresa / institució: Institut de Química Computacional i Catàlisi  
Universitat de Girona

Vistiplau tutor:

Nom del tutor: Pedro Salvador Sedano

Empresa / institució: IQCC (Universitat de Girona)

Correu(s) electrònic(s): pedro.salvador@udg.edu

## Resum

En aquest treball es fan servir 12 constants d'acoblament spin-spin ( $J$ ) obtingudes experimentalment pel pèptid Ala<sub>3</sub> per tal de determinar l'estructura secundària del seu aminoàcid central. Per a això, es fa servir un procediment que no assumeix cap conformació principal, sinó que considera tot el diagrama de Ramachandran enter sense limitar-lo a zones concretes. A més, el algorisme considera la possibilitat de que el valor de cada  $J$  observat sigui resultat de la contribució de diferents estructures/conformacions, amb els pesos corresponents. Per a poder aplicar aquest procediment és necessari conèixer el valor de cada  $J$  per qualsevol combinació dels angles diedres  $\phi$  i  $\psi$  de l'aminoàcid central. Això s'aconsegueix mitjançant la Teoria del Funcional de la Densitat.

Primerament s'optimitzen les 625 estructures que es generen fixant els angles  $\phi$  i  $\psi$  a intervals de 15° a nivell B3LYP/D95(d,p). Els valors de les 12  $J$  es calculen per a cada estructura amb diferents funcionals i bases per veure quin efecte té cadascun, i es representen en funció de  $\phi$  i  $\psi$  obtenint les quadrícules de  $J$  de dues dimensions. Amb aquestes quadrícules es pot observar que les equacions monodimensionals de Karplus no són aplicables en general, ja que amb excepció de les constants  ${}^3J(C_6-H_{14})$ ,  ${}^3J(C_6-C_{15})$  i  ${}^3J(C_6-C_{16})$ , els valors de  $J$  mostren dependència amb els dos angles diedres. També s'introdueix l'efecte del solvent en els càlculs de  $J$ , tot i que ràpidament es descarta degut a les grans desviacions obtingudes respecte les dades experimentals.

La combinació del funcional MPW1WK i la base 6-31G-J és la que dona uns millors resultats tenint en compte la desviació mitjana (RMSD) respecte les  $J$  experimentals, però les estructures que es troben no es corresponen a conformacions esperables. En canvi calculant les  $J$  amb el funcional MPW1WK i la base D95(d,p) descontreta, l'error és lleugerament més elevat que amb la base 6-31G-J s'obtenen unes estructures que es corresponen totalment amb conformacions d'alfa hèlix, làmina  $\beta$  i l'hèlix de poliprolina II. D'aquesta manera el funcional MPW1WK millora molt els resultats obtinguts respecte càlculs anteriors de referència, a on s'havia fet servir el funcional B3LYP amb la base D95(d,p).

## Resumen

En este trabajo se usan 12 constantes de acoplamiento spin-spin ( $J$ ) obtenidas experimentalmente para el péptido Ala<sub>3</sub> para determinar la estructura secundaria de su aminoácido central. Para ello se usa un procedimiento que no asume ninguna conformación principal, sino que considera todo el diagrama de Ramachandran entero sin limitarlo a zonas concretas. Además, el algoritmo considera la posibilidad de que el valor de cada  $J$  observado sea resultado de la contribución de diferentes estructuras/conformaciones, con los pesos correspondientes. Para poder aplicar este procedimiento es necesario conocer el valor de cada  $J$  para cualquier combinación de los ángulos diedros  $\phi$  y  $\Psi$  del aminoácido central. Esto se consigue mediante la Teoría del Funcional de la Densidad.

Primeramente se optimizan las 625 estructuras que se generan fijando los ángulos  $\phi$  y  $\Psi$  a intervalos de 15º a nivel B3LYP/D95(d,p). Los valores de las 12  $J$  se calculan para cada estructura con diferentes funcionales y bases para determinar el efecto de cada uno, y se representan en función de  $\phi$  y  $\Psi$  obteniendo cuadrículas de  $J$  de dos dimensiones. Con estas cuadrículas se puede observar que las ecuaciones monodimensionales de Karplus no son aplicables en general, puesto que a excepción de las constantes  ${}^3J(\text{C}_6\text{-H}_{14})$ ,  ${}^3J(\text{C}_6\text{-C}_{15})$  y  ${}^3J(\text{C}_6\text{-C}_{16})$ , los valores de  $J$  muestran dependencia con los dos ángulos diedros. También se introduce el efecto del solvente en los cálculos de  $J$ , a pesar de que rápidamente se descarta debido a las grandes desviaciones obtenidas respecto a los datos experimentales.

La combinación del funcional MPW1WK y la base 6-31G-J es la que da unos mejores resultados teniendo en cuenta la desviación media (RMSD) respecto a los valores de  $J$  experimentales, pero las estructuras que se encuentran no se corresponden a conformaciones esperables. En cambio, calculando las  $J$  con el funcional MPW1WK y la base D95(d,p) descontrada, el error es ligeramente más elevado pero se obtienen unas estructuras que se corresponden totalmente con conformaciones de hélice alfa, lámina beta y hélice de poliprolina II. De este modo el funcional MPW1WK mejora significativamente los resultados obtenidos respecto a cálculos anteriores de referencia, donde se había usado el funcional B3LYP con la base D95(d,p).

## Summary

In this work, 12 spin-spin coupling constants ( $J$ ) obtained experimentally for the Ala<sub>3</sub> peptide are used to determine the secondary structure of its central amino acid. For this purpose, a procedure is used that does not assume any main conformation, but considers the entire Ramachandran diagram without limiting it to specific areas. In addition, the algorithm considers the possibility that the value of each observed  $J$  is the result of the contribution of different structures/conformations, with the corresponding weights. In order to apply this procedure, it is necessary to know the value of each  $J$  for any combination of the dihedral angles  $\phi$  and  $\psi$  of the central amino acid. This is achieved through the Density Functional Theory.

Firstly, the 625 structures generated by fixing the dihedral angles  $\phi$  and  $\psi$  are optimized at intervals of 15° at the B3LYP/D95(d,p) level of theory. The values of the 12  $J$  are calculated for each structure with different functionals and basis sets in order to determine their possible effects, and the results are represented as a function of  $\phi$  and  $\psi$ , thus obtaining two-dimensional plots for each  $J$ . With these grids it can be observed that one-dimensional Karplus equations are not applicable in general, since with the exception of the constants  $^3J(\text{C}_6\text{-H}_{14})$ ,  $^3J(\text{C}_6\text{-C}_{15})$  and  $^3J(\text{C}_6\text{-C}_{16})$ , the values of  $J$  show dependence with the two dihedral angles. The effect of the solvent in the  $J$  calculations is also introduced, although it is quickly discarded due to the large deviations obtained with respect to the experimental data.

The combination of the functional MPW1WK and the basis set 6-31G-J gives the best results according to the mean deviation (RMSD) with respect to the experimental  $J$  values, but the structures that are obtained do not correspond to expected conformations. On the other hand, by calculating the  $J$  with the functional MPW1WK and the uncontracted basis set D95 (d,p), the error is slightly higher but the structures obtained fully correspond with alpha helix, beta sheet and polyproline II helix conformations. In this way the functional MPW1WK significantly improves the results obtained with respect to previous reference calculations, where the functional B3LYP with the base D95 (d, p) had been used.

# Índex

<b>1. Introducció</b> .....	<b>1</b>
1.1 Ressonància magnètica nuclear.....	1
1.2 Elucidació estructura secundària .....	4
1.3 Antecedents .....	6
<b>2. Objectives</b> .....	<b>7</b>
<b>3. Metodologia</b> .....	<b>8</b>
3.1 Teoria del funcional de la densitat .....	8
3.2 Optimització parcial de la geometria.....	9
3.3 Càlculs <i>ab initio</i> de constants d'acoblament spin-spin.....	9
3.4 Conjunt de bases.....	10
3.5 Mètode de cerca exhaustiva.....	12
3.6 Criteris d'ètica i sostenibilitat .....	14
<b>4. Resultats i discussió</b> .....	<b>15</b>
4.1 Validació del mètode de càlcul .....	15
4.2 Obtenció de quadrícules de J.....	17
4.3 Estructura secundària del pèptid Acetil(Ala) <sub>3</sub> NH <sub>2</sub> .....	21
<b>5. Conclusions</b> .....	<b>28</b>
<b>6. Referències</b> .....	<b>29</b>

# 1. Introducció

## 1.1 Ressonància magnètica nuclear

L'espectroscòpia de ressonància magnètica nuclear, NMR, és una tècnica que explota les propietats magnètiques de certs nuclis atòmics per tal de proporcionar informació detallada respecte a l'estructura de les molècules. Aquesta tècnica s'utilitza, per exemple, per a identificar compostos orgànics.

Molts nuclis posseeixen un moment angular de spin caracteritzat per un número quàntic de spin nuclear  $I$  (l'anàleg del  $s$  pels electrons) que pot ser un nombre enter o semienter.<sup>1</sup> A la taula 1 es pot veure els diferents valors de  $I$  per a nuclis concrets. Isòtops del mateix element poden tenir diferent  $I$  i nuclis molt comuns com el  $^{12}\text{C}$  i el  $^{16}\text{O}$  tenen un  $I=0$ , per tant no tenen moment angular i no tenen moment magnètic. La tècnica de NMR és aplicable a qualsevol mostra que contingui nuclis que tinguin spin diferent de 0.

Si  $I>0$  apareix un moment magnètic en forma de vector. Les possibles orientacions venen determinades pel valor de  $m_I$ .

$$m_I = I, I - 1, I - 2, \dots - I + 1, -I. \quad (1)$$

La interacció d'aquest moment magnètic del nucli amb un camp magnètic extern aplicat ocasiona el desdoblament dels nivells energètics nuclears. Es poden generar transicions entre aquests diferents estats i l'absorció d'energia es pot detectar i enregistrar com una línia espectral o banda, anomenada senyal de ressonància de freqüència igual a:

$$\nu = \frac{\gamma_N \sigma B_0}{2\pi} \quad (2)$$

On  $\gamma_N$  és la constant giromagnètica del nucli,  $B_0$  és el camp magnètic extern aplicat i  $\sigma$  és una constant que té en compte el diferent apantallament que pateix cada nucli degut als electrons. D'aquesta manera, nuclis amb diferent entorn químic donen diferent senyal. Per això s'utilitza el NMR per l'elucidació estructural.

Taula 1: número quàntic de spin nuclear ( $I$ ) per diferents àtoms actius en NMR, constants giromagnètiques i abundància natural.

Nucli	$I$	$\gamma_N/10^7 T^{-1} s^{-1}$	Abundància natural/%
$^1\text{H}$	1/2	26.75	99.98
$^{19}\text{F}$	1/2	25.18	100.00
$^{13}\text{C}$	1/2	6.73	1.11
$^{14}\text{N}$	1	1.93	99.63
$^{15}\text{N}$	1/2	-2.71	0.37
$^{17}\text{O}$	5/2	-3.63	0.04

No totes les bandes que donen senyal NMR són senzilles (singlets) sinó que algunes es desdoblen en doblets, triplets o altres multiplicitats. El desdoblament de les diferents bandes és degut a l'acoblament spin-spin. Això és així perquè cada nucli magnètic pot contribuir el camp que rep un altre nucli i d'aquesta manera pot modificar lleugerament les seves freqüències de ressonància. L'energia deguda a l'acoblament spin-spin de dos nuclis A i X és

$$E_{m_A m_X} = hJ_{AX} m_A m_X, \quad (3)$$

on la  $J_{AX}$  és la constant d'acoblament spin-spin entre els nuclis A i X (en Hertz) i pot ser positiva o negativa. Quan els moments magnètics nuclears estan en posició paralela els valors de J són negatius ( $J < 0$ ) i quan estan en posició antiparalela els valors de J són positius ( $J > 0$ ). Tot i això, experimentalment, es sol mesurar la J en valor absolut. El diagrama resultant de nivells d'energia pel cas AX a on  $I_A = 1/2$  i  $I_X = 1/2$  es pot observar a la figura 1; on podem veure la interacció de A amb el camp magnètic (a), la interacció de X amb el camp magnètic (b) i finalment l'acoblament on  $J_{AX} > 0$  (c).

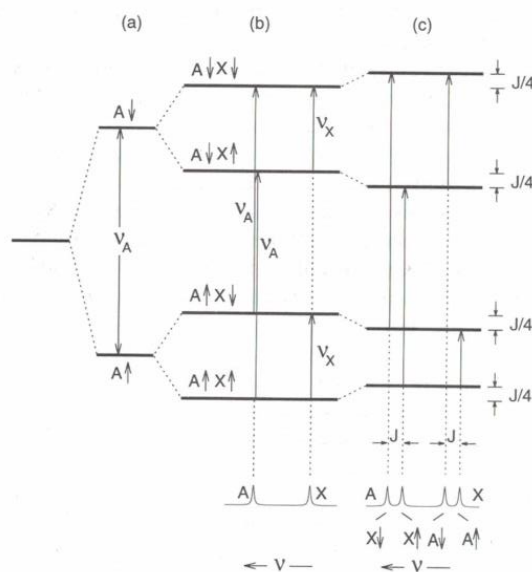


Figura 1: Diagrama nivells d'energia i senyals esperades amb i sense acoblament.<sup>1</sup>

Les constants d'acoblament spin-spin s'expressaran habitualment amb la notació  ${}^nJ(X,Y)$  on n és el nombre d'enllaços que separen els dos àtoms.

Per altra banda, la teoria de Karplus descriu la variació del valor de la constant d'acoblament entre dos àtoms separats per tres enllaços i l'angle diedre que en determina la seva posició relativa (Figura 2). L'equació de Karplus afirma el valor de J només depèn d'aquest angle segons:<sup>2</sup>

$${}^3J(\phi) = C \cos^2 \phi + B \cos \phi + A, \quad (4)$$

on A B i C són constants que depenen dels substituents i les seves electronegativitats.

L'equació de Karplus original ha estat subjecte a diferents modificacions al llarg del temps per tal d'aconseguir un major correspondència entre els resultats experimentals i teòrics. També s'han calculat empíricament les diferents constants en molècules concretes.

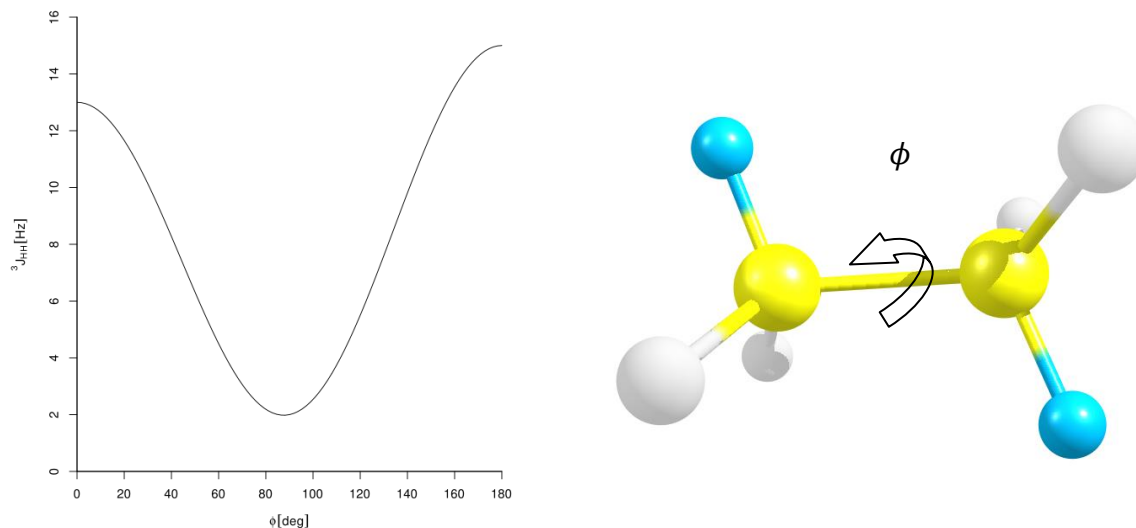


Figura 2: Representació de l'equació de Karplus per J veïnals en l'età (esquerre) i la molècula d'età on s'observa la connexió entre els hidrògens a tres enllaços (àtoms color blau) i l'angle diedre (dreta).<sup>2</sup>

La teoria de Karplus va demostrar que l'acoblament spin-spin es podia utilitzar per l'estudi conformacional de molècules ja que el valor de les J permet definir les posicions relatives dels nuclis, donat que la magnitud de l'interacció spin-spin (constant d'acoblament, J) depèn del nombre i tipus d'enllaços que separen els diferents nuclis. Als recents anys s'han fet accessibles diferents programes de predicció i estimació dels valors de les constants d'acoblament, tot i que, el càlcul dels valors exactes de la constants és encara ara un repte pels químics teòrics. L'importància de poder tenir aquests valors teòrics és que la seva comparació amb el valor experimental ens verifica que la nostra estructura en qüestió és la correcta.

L'equació de Karplus aplica les constants d'acoblament spin-spin de tipus  $^3J$ . Les  $^1J$  i  $^2J$  no s'utilitzen tant per la determinació estructural perquè depenen d'altres paràmetres geomètrics diferents com la distància d'enllaç i no en l'angle diedre.



## 1.2 Elucidació estructura secundària

Karplus amb la seva afirmació va demostrar que els valors de les  $J$  depenen de la conformació i és per això que s'utilitza el valor de les constants d'acoblament spin-spin per determinar l'estructura secundària de pèptids i proteïnes.

L'estructura secundària ve caracteritzada pel rang de valors que prenen de forma simultània dos angles díedres ( $\phi$  i  $\psi$ ) per cada aminoàcid definit a partir d'àtoms que formen l'esquelet de la proteïna o pèptid. Com es pot veure a la Figura 3, es defineix l'angle díedre  $\phi$  a partir dels àtoms N-C'-C $\alpha$ -N i l'angle  $\psi$  entre els àtoms C'-C $\alpha$ -N-C'. Una manera visual de representar l'estructura secundària d'una proteïna és per mitjà d'un diagrama de Ramachandran. Es tracta d'un gràfic bidimensional on es representen els parells de valors ( $\phi$  i  $\psi$ ) que pren cada aminoàcid de la proteïna. La densitat de punts en zones del diagrama dóna una idea del nombre d'aminoàcids que adopta una estructura secundària determinada.

S'ha d'indicar que el moviment dels àtoms és més ràpid (mil·lisegons) que l'obtenció de les dades NMR ja que aquest és un procés lent i el que es mesura experimentalment és un promig de totes les conformacions rellevants. Això és el que explica que àtoms magnèticament equivalents no donin multiplicitat de senyals com pot ser la rotació dels hidrògens en els metils que donen un únic senyal. En el cas de molècules flexibles, la constant d'acoblament observada,  $J_{obs}$ , es pot escriure com

$$J_{obs} = \sum_i^N p_i \langle J \rangle_i \quad (5)$$

a on  $i$  indica cada conformació i  $p_i$  és el pes que té cada conformació individual.

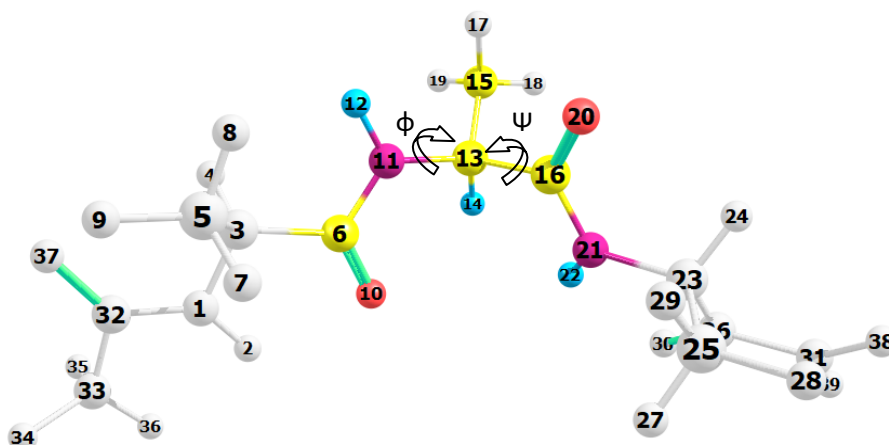


Figura 3: molècula d'acetil(Ala)<sub>3</sub>NH<sub>2</sub> on els àtoms en color són el residu central de la molècula i també es pot observar l'enllaç peptídic.

Kallenbach<sup>5</sup> va afirmar que pèptids d'alanina tenien com a estructura secundària dominant la poliprolina II (PPII) a 2°C i el contingut de làmina  $\beta$  creixia un 10% a 55°C. Basat en aquesta dependència de temperatura va arribar a la conclusió de que hi havia una transició des de la PPII a temperatures baixes fins a la làmina  $\beta$  a temperatures altes; en altres paraules la conformació de PII és la més habitual a temperatura ambient. Per altra banda Vila *et al.*<sup>6</sup> va suggerir que la conformació de PPII no és la principal del pèptid d'Ala<sub>3</sub> en solució aquosa ja que els seus càlculs indiquen que té un pes del 30%.

Més tard, tant Schwalbe<sup>3</sup> com Schweitzer-Stenner<sup>4</sup> van publicar estudis de conformacions de l'Ala<sub>3</sub> basats en l'ajust de diferents constants d'acoblament spin-spin experimentals i d'altres dades experimentals. Schwalbe va assumir l'existència de tres estructures representatives, l'alfa hèlix, la làmina  $\beta$  i la PPII. Va utilitzar les equacions de Karplus per calcular les <sup>3</sup>J per cada una de les conformacions. El millor ajust de les constants d'acoblament per el residu central de l'Ala<sub>3</sub> va donar una població de 0.92, 0.08 i 0.0 per la poliprolina II, làmina  $\beta$  i alfa hèlix respectivament. Va obtenir una RMSD (desviació quadràtica mitjana) de 0.255 Hz per el seu ajust amb les conformacions assumides amb una J que donava unes desviacions bastant grans, la <sup>3</sup>J(6,16) (0.25 experimental versus el 0.6 ajustat)<sup>3</sup>. Schweitzer-Stenner va afegir una nova conformació, el gir, que no tenia cap pes per Schwalbe i d'aquesta manera va obtenir quatre conformacions model. L'anàlisi també va predir que la poliprolina II és la que tenia un pes més elevat (0.84).<sup>4</sup>

El sistema d'Ala<sub>3</sub> s'ha fet servir com a model per determinar el pes que poden tenir diferents conformacions de l'Ala central. Una de les maneres de determinar-ho és a partir de dades experimentals de NMR i és el que s'intentarà realitzar.

### 1.3 Antecedents

En aquest treball es parteix de l'article de la referència 7 on es va intentar ajustar l'estructura secundària del aminoàcid central del pèptid d'Ala<sub>3</sub> utilitzant un mètode que va comparar deu constants d'acoblament spin-spin calculades teòricament amb el seu valor experimental utilitzant la DFT (Teoria del Funcional de la Densitat).

Aquest treball diferia d'altres ja escrits,<sup>3,4</sup> en el fet de que no es va partir de les estructures representatives específiques com podien ser l'hèlix de poliprolina II o la làmina  $\beta$ , sinó que es va usar tot el diagrama de Ramachandran sencer i es va determinar la combinació d'estructures de cada zona que van fer que les diferents constants d'acoblament calculades teòricament tinguessin un error més petit respecte a les calculades experimentalment.

Es va aplicar un algorisme (s'explica en detall a l'apartat 3.5) de cerca exhaustiva per tot el diagrama de Ramachandran que va donar com a resultat les estructures de la figura 4, amb el seu corresponent pes que s'indica a la llegenda. Per una banda s'observa l'estructura secundària d'alfa hèlix a la regió 2 però l'estructura de la regió 1 es troba situada entre la làmina  $\beta$  (regió 4) i la poliprolina II (regió 3) sense ser cap de les dues i es tracta d'una estructura difícil de justificar. Pot ser degut a que a l'hora de fer l'ajust es van descartar dues J problemàtiques ja que presentaven un error molt elevat, la  $^1J(N_{11},C_{13})$  i la  $^2J(C_{13},N_{21})$ . Per tant doncs la metodologia utilitzada a la referència 7 no va donar els resultats esperats i es proposa utilitzar-ne una altra per observar quins resultats s'obtidrien i si es millorarien.

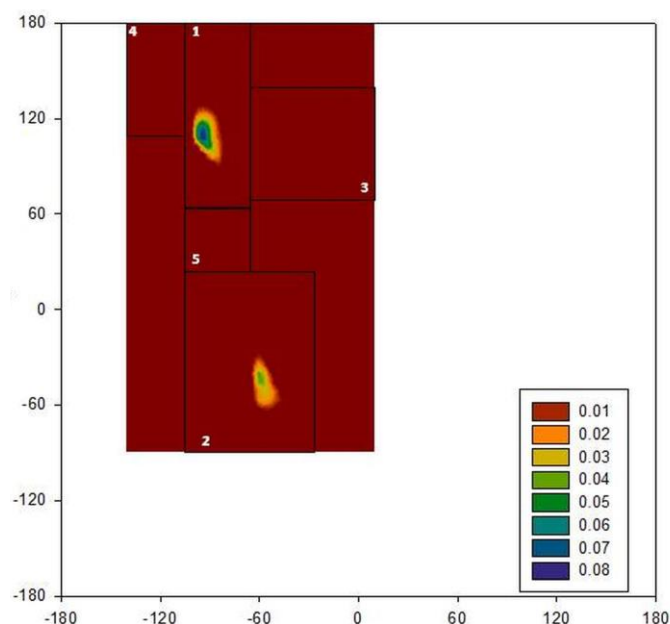


Figura 4: probabilitat de distribució derivat de la millor aproximació en termes de RMSD (*Root-Mean-Square-Deviation*). Extreta de la referència 7.

## 2. Objectives

- Our main goal is comparing 12 measured J-coupling (using the weight of different conformations) to the corresponding values and see which is the effect of the functional DFT and the basis set in the J's. We do not attempt to use the best possible basis set; we want to use a basis set that gives us acceptable results while still being reasonably small.
- For each J coupling considered, show the values of J in a two dimensional grid (Ramachandran plot) and question the Karplus equation.
- Improve the results of the article 7 in which we are based to see that the functional and the basis set are very important in the DFT.
- Observe that  $^1J$  and  $^2J$  are also useful for the structural elucidation of the peptide.
- Develop a method that fit the structure of Ala<sub>3</sub> via NMR spectroscopy using a procedure that does not assume any peptide conformations, using the entire Ramachandran space.

## 3. Metodologia

### 3.1 Teoria del funcional de la densitat

El mètode DFT (*Density Functional Theory*) és un procediment variacional alternatiu a la solució de l'equació de Schrödinger, on el funcional de l'energia de l'energia electrònica és minimitzat respecte a la densitat electrònica. S'obté l'energia electrònica treballant amb el funcional de la densitat electrònica i no amb les funcions d'ona com podria ser amb el mètode de Hartree-Fock. Dins l'aproximació de Born-Oppenheimer l'energia del sistema és la suma de sis components:

$$E_{DFT} = E_{NN} + E_T + E_v + E_{Coul} + E_{Bes} + E_{Corr} \quad (6)$$

Les expressions per l'energia de repulsió nucli-nucli  $E_{NN}$ , la cinètica  $E_T$ , l'atracció nucli-electró  $E_v$ , i la repulsió de Coulomb electró-electró  $E_{Coul}$ , són les mateixes que es fan servir en la Teoria de Hartree-Fock. Tanmateix l'energia de bescanvi electró-electró  $E_{bes}$  és diferent que la que es fa servir en la teoria de Hartree-Fock. L'últim terme,  $E_{corr}$  descriu els efectes de correlació dels electrons de diferents spin i no es té en compte en la teoria de Hartree-Fock.

Els primers passos en la DFT van ser donats per Thomas-Fermi-Dirac que van fer un model estadístic per una distribució homogènia d'electrons on l'energia cinètica s'obté a partir de la densitat electrònica. La base teòrica per la DFT va ser donada però per Hohenberg i Kohn l'any 1964 quan van demostrar que, per l'estat fonamental, qualsevol observable pot ser calculat de forma exacte a partir de la densitat electrònica  $\rho_0(\vec{r})$  (un observable, que és una funció de 3 coordenades únicament). Els observables, i en particular l'energia són un funcional de la densitat electrònica.

$$E_0 = E[\rho_0(\vec{r})] \quad (7)$$

A més, van demostrar també que la densitat exacta minimitza aquest funcional, és a dir proporciona l'energia més baixa (que coincideix amb l'exacta de l'estat fonamental). El desenvolupament més important però, va ser donat l'any següent, quan Kohn i Sham van demostrar que a partir de la teoria del funcional de la densitat és possible escriure una equació per orbitals d'una partícula, dels quals s'obté la densitat de forma molt senzilla i anàloga al mètode Hartree-Fock.

Una de les avantatges de la DFT és que la densitat és una magnitud molt més simple que les funcions d'ona i per tant és més fàcil de calcular i sistemes molt més complexes són més accessibles. La desavantatge és que no es coneix de manera exacte el funcional de correlació-bescanvi. Hi ha diferents famílies d'aproximació de funcionals com són els GGA (*Generalized Gradient Aproximation*), meta-GGA (*Meta-Generalized Gradient Aproximation*) i híbrids.

Els funcionals que s'usen en aquest treball són: B3LYP, tHCTH, OPBE, MPwPW91 i MPW1WK.

El funcional OPBE és un exemple de GGA. El funcional de correlació-bescanvi depèn de la densitat electrònica i del seu gradient. El funcional tHCTH és de la família dels meta-GGA. El funcional de correlació-bescanvi depèn de la densitat electrònica, gradient i Laplaciana.

Els funcionals B3LYP, MPwPW91 i MPW1WK són funcionals híbrids. Aquests inclouen una part d'intercanvi que prové de Hartree-Fock, típicament controlat per paràmetres ajustables. Això és així perquè el mètode Hartree-Fock té un error sistemàtic. La falta de correlació de Coulomb implica que els electrons se'ls permet apropar més del que hauria de ser, per tant les longituds d'enllaç tendeixen a ser més petites i les energies d'enllaç massa grans. En DFT, pel contrari, els errors sistemàtics passen en direcció oposada. Així doncs una combinació de l'intercanvi de Hartree-Fock i de la DFT hauria de millorar considerablement els resultats i per això es van crear aquests funcionals híbrids.

## 3.2 Optimització parcial de la geometria

Tots els càlculs sobre les optimitzacions de la geometria i les constants d'acoblament spin-spin s'han dut a terme mitjançant el programa GAUSSIAN 09.<sup>8</sup>

Es procedeix a dividir el diagrama de Ramachandran de manera que els dos angles diedres,  $\phi$  i  $\Psi$  (definitos a la figura 3), es van fixant amb intervals de 15 graus (de  $-180^\circ$  a  $+180^\circ$ ) mentre que totes les altres coordenades dels àtoms es relaxen creant 625 estructures. Aquestes 625 estructures s'optimitzen utilitzant la DFT a nivell B3LYP/D95(d,p). Per tal que la molècula optimitzada no modifiqui el valor dels dos angles  $\phi$  i  $\Psi$ , s'introdueix l'opció *opt=modredundant*. Amb aquesta indicació és buscarà la geometria optimitzada però sense modificar els valors de  $\phi$  i  $\Psi$  que ja estaran preestablerts per nosaltres.

## 3.3 Càlculs *ab initio* de constants d'acoblament spin-spin

La primera teoria de mecànica quàntica de les constants d'acoblament spin-spin la va realitzar Ramsey qui va afirmar que quatre mecanismes diferents podien comptar a l'hora de calcular les constants d'acoblament spin-spin. Per tant doncs, tenint en compte la teoria no relativista de Ramsey,<sup>9</sup> les  $J$  són la suma de quatre contribucions (veure equació 8): contacte de Fermi (FC), spin-òrbita diamagnètic (DSO), spin-òrbita paramagnètic (PSO) i spin-dipol (SD).

$$J_{AX} = J_{AX}^{FC} + J_{AX}^{DSO} + J_{AX}^{PSO} + J_{AX}^{SD} \quad (8)$$

Ramsey va desenvolupar expressions de perturbacions per cada un dels mecanismes que hi contribueixen. En el primer terme (SO), un moment magnètic nuclear indueix corrents electròniques orbitals que, en el mateix temps, produeixen camps magnètics en al lloc del segon nucli. El segon mecanisme (SD) involucra la interacció del dipol magnètic del spin electrònic amb el dipol magnètic del spin nuclear. La interacció del dipol entre el moment magnètic d'un nucli i el spin electrònic produeix una "polarització del spin electrònic" que no elimina el camp magnètic que es produeix en l'altre nucli. La interacció FC és entre els moments nuclears i els spins electrònics. Ramsey va demostrar que la FC és la contribució dominant.

Finalment per avaluar l'efecte del solvent en la determinació de les J, es fa servir el mètode *Self-Consistent Reaction Field*. Es tracta el solvent que en aquest cas és l'aigua com un continu polaritzable amb una constant dielèctrica determinada ( $\epsilon$ ). El solut es col·loca en una cavitat dins del dissolvent, que pot presentar càrregues parcials a la seva superfície simulant l'efecte electrostàtic dels solvents polars. Es tracta doncs d'un tractament real del solut dins d'un dissolvent model.

### 3.4 Conjunt de bases

El conjunt de bases és el conjunt de funcions monoelectròniques en les que s'expandeixen els orbitals moleculars. Com que generalment són funcions centrades en els nuclis, podem entendre-les com orbitals atòmics dels àtoms que componen la molècula. Aquests orbitals atòmics, dins l'aproximació OM-CLOA són els que es combinaran per formar els orbitals moleculars. Un dels factors que més influeix en la qualitat dels resultats d'un càlcul d'estructura electrònica és el conjunt d'orbitals atòmics que s'introdueixen. N'hi ha de dos tipus principals: orbitals tipus Slater, (STO) (equació 9) i Orbitals tipus Gaussiana, (GTO) (equació 10). A la figura 5 es poden veure les corresponents representacions gràfiques.

$$G_u(\vec{r}) = N_u (x - x_A)^{l_u} (y - y_A)^{m_u} (z - z_A)^{n_u} e^{-\alpha_u \sqrt{(x-x_A)^2 + (y-y_A)^2 + (z-z_A)^2}} \quad (9)$$

$$G_u(\vec{r}) = N_u (x - x_A)^{l_u} (y - y_A)^{m_u} (z - z_A)^{n_u} e^{-\alpha_u ((x-x_A)^2 + (y-y_A)^2 + (z-z_A)^2)} \quad (10)$$

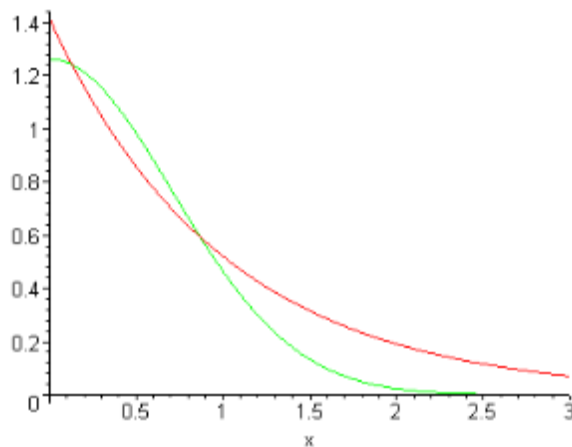


Figura 5: funció d'ona STO (verd) i funció d'ona GTO (vermell).

Els conjunts de funcions de base tipus STO descriuen millor la densitat electrònica a prop dels nuclis, malgrat això la seva utilització comporta un cost computacional molt elevat ja que la solució de les integrals que s'han de resoldre amb funcions tipus STO no és analítica. En canvi la utilització de conjunts tipus GTO comporta un cost computacional més baix, ja aquestes integrals sí són analítiques. El problema és que, aquest tipus de funcions no descriuen tan bé la densitat electrònica propera als nuclis. Per això, és necessari els electrons interns dels orbitals atòmics s'expressen com a combinació lineal de gaussianes del mateix tipus de la forma

$$\chi_i(\vec{r}_1) = \sum_{\mu}^{M_i} d_{\mu i} G_{\mu}(\vec{r}_1) \quad (11)$$

on el coeficient de la combinació lineal de primitives  $d_{\mu i}$  determina el tipus de conjunt de funcions de base utilitzat.

És sap que les bases estàndards que s'utilitzen per determinar l'energia o per optimitzacions de geometria no són les més adequades per calcular les constants d'acoblament spin-spin.<sup>11</sup> Els càlculs de J requereixen bases amb uns exponents més grans que els que hi ha en les bases estàndards. En particular funcions tipus s amb exponents elevats són necessàries per la contribució de FC que és la més important i depèn de la densitat electrònica al nucli. El terme PSO requereix funcions tipus p addicionals i el terme SD funcions p, d i f.<sup>11</sup> Les bases intenten reproduir bastant bé el terme FC ja que és el que contribueix de manera més gran en el càlcul de J. Hi ha diferents bases especialitzades en el càlcul de constants d'acoblament spin-spins com són les pcJ-n i aug-pcJ-n. Desafortunadament, aquestes bases són massa grans sobretot si  $n > 1$ . Algunes d'aquestes bases de mida petita específiques pel càlculs de J són la 6-31G-J<sup>11</sup> o ccJ-pVDZ<sup>12</sup>.



Es pot descontretar la base de manera que en comptes d'expandir els orbitals moleculars sobre el conjunt de funcions de base que són combinacions lineals de primitives, es fa l'expansió directament sobre les primitives de manera que es dona més llibertat a l'orbital molecular. D'aquesta manera, el que s'aconsegueix és tenir un major número de funcions de base amb un valor del exponent  $\alpha$  molt alt, per tant més semblants a funcions tipus STO, que ens permetrà una millor descripció dels electrons propers als nuclis. Per indicar que s'utilitza una base deconstreta s'introduirà la lletra "u" després del nom de la base.

A la taula 2 s'indiquen el nombre de funcions de base i de primitives pel conjunt de bases utilitzats..

Taula 2: Composició de les bases utilitzades en el càlcul de J.

	hidrogen		C,N,O	
	primitives	concretes	primitives	concretes
D95(d,p)	4s1p	2s1p	9s5p1d	3s2p1d
cc-pVDZ	4s1p	2s1p	9s4p1d	3s2p1d
cc-pVTZ	5s2p1d	3s2p1d	10s5p2d1f	4s3p2d1f
6-31G-J	7s	4s	13s5p	6s2p
ccJ-pVDZ	6s2p	4s2p	11s5p2d	6s3p2d

Quan s'utilitzi una base en forma deconstreta, la mida de la base vindrà donada per les funcions primitives.

### 3.5 Mètode de cerca exhaustiva

Les dades de les constants d'acoblament spin-spin es poden fer servir per determinar la barreja d'estructures del pèptid d'alanina Ala<sub>3</sub> ajustant a mitjanes ponderades dels valors de J obtinguts computacionalment .

També s'utilitza un algorisme que interpola valors de J per qualsevol parell d'angles,  $\phi$  i  $\Psi$ , arbitraris que difereixen dels 625 punts calculats fent servir l'equació del pla definida pels tres punts més propers de la quadrícula de dos dimensions. Aquesta triangulació té l'avantatge de produir una quadrícula continua de dos dimensions de valors d'acoblament. La quadrícula 2D s'estén en el rang de  $-210^\circ$  a  $+210^\circ$  en cada coordenada per tenir en compte adequadament la periodicitat de les variables angulars i per evitar discontinuïtats en la interpolació de les variables  $\phi$  i  $\Psi$ . L'exactitud d'aquesta interpolació es pot trobar a la referència 13.

Primerament s'ha de dividir l'espai ( $\phi$  i  $\Psi$ ) en diferents regions, N. Aquesta divisió de l'espai en regions pot cobrir o no tot el diagrama de Ramachandran. Per cada regió, k, s'ha d'introduir la resolució pels valors de  $\phi$  i  $\Psi$  que permeten  $n_k$  estructures possibles dins la regió que considerem. S'interpolen els diferents valors de J calculats i es compara

amb l'experimental. Si la diferència excedeix el límit donat, per cada valor de  $J$ ,  $\delta J$ , l'estructura es descarta. Si els errors de totes les  $J$  calculades està per sota  $\delta J$  s'obté l'estructura i es grava l'error RMSD (desviació quadràtica mitjana) entre totes les  $J$  calculades i l'experimental.

Com a segon pas es repeteix el procediment però considerant que les  $J$  calculades provenen d'un pes ponderat de dues estructures, cada una associada a una regió diferent. Per avaluar les dues possibles conformacions s'introdueix un paràmetre que determina el pes  $w$ , necessari per veure la ponderació de cada una de les conformacions. Altra cop si totes les  $J$  calculades estan per sota de  $\delta J$  s'obté la solució i es veu l'error RMSD amb els pesos associats a cada estructura que donen el millor valor RMSD. El procés es pot repetir per cada possible parell d'estructures de diferents regions. Si el resultat de la barreja de dues estructures és millor que el d'una estructura (utilitzant com a criteri el RMSD), es continua buscant afegint una nova (tercera) estructura amb un pes determinat que ens ho dirà el resultat. Novament s'avalua el RMSD per determinar si incloure una nova estructura millora o perjudica el resultat anterior.

Com més estructures possibles es consideren més solucions possibles hi ha i per això s'ha de limitar aquest nombre de solucions introduint un  $\text{RMSD}_{\text{max}}$ . Només les solucions que estiguin per sota del  $\text{RMSD}_{\text{max}}$  es retindran per fer les mitjanes.

Hi ha factors importants que es poden anar variant per veure de quina manera la solució també es modifica, com la resolució dels angles diedres i els factors del pes de cada estructura. La resolució dels angles és la diferència que es requereix entre un dels angles ( $\delta\phi$  o  $\delta\psi$ ) per indicar que dues estructures són diferents.

L'estratègia que es segueix és la de començar amb una resolució dels dos angles i els factors de ponderació una mica grans i es considera tot el diagrama de Ramachandran sencer per veure quines són les zones amb un pes més elevat a l'hora de calcular les  $J$ . Després es segueix el mateix protocol però per zones més petites per cada estructura i resolució més afinada per anar ajustant les zones a on podem veure-hi més interès. També s'ha de tenir en compte de no triar una zona nova que perjudiqui el resultat anterior. Si les regions escollides són massa petites les conformacions es superposaran en diferents zones. En canvi si les regions són massa grans, es podria esperar més d'una conformació en una sola regió.

Després del procés de cerca exhaustiva es distingeixen tres tipus de solucions anomenades  $\text{best}_n$ ,  $\text{dens}_n$  i  $\text{aver}_n$  on  $n$  són el número de regions que ha combinat. El fitxer  $\text{best}$  et dona la millor solució trobada amb les condicions que li has indicat i amb el seu corresponen RMSD. El fitxer  $\text{aver}$  en canvi el que fa és fer una mitjana de totes les solucions trobades i també del RMSD. Per tant podríem dir que és el mètode més robust ja que no considera una única solució possible. Finalment el fitxer  $\text{dens}$

serveix per poder representar els resultats en el diagrama de Ramachandran ja que et diu el pes que té cada angle  $\phi$  i  $\psi$ .

### 3.6 Criteris d'ètica i sostenibilitat

En aquest cas la química computacional és sostenible perquè no genera residus ja que no es necessiten reactius químics per realitzar l'estudi. El nostre treball en qüestió ajuda a identificar compostos comparant les dades computacionals simulades amb les experimentals. Així doncs amb un anàlisi de NMR en pots tenir prou per identificar el compost químic de manera que no cal fer més proves de caracterització del compost i abarateixes les proves d'identificació en el sentit de treball al laboratori i reactius químics.

## 4. Resultats i discussió

### 4.1 Validació del mètode de càlcul

S'intenta demostrar les diferències entre una base especialitzada en càlculs de J però d'una mida petita i una base estàndard del Gaussian. Per escollir una de les diferents bases estàndards s'han fet diverses comparacions entre les següents: D95(d,p), D95(d,p)u, cc-pVDZ, cc-pVTZ, cc-pVDZu i cc-pVTZu i ens ajudarem dels càlculs ja realitzats en la referència 7, ja que s'utilitzaran les seves estructures representatives<sup>7</sup> (-92,110) que té un pes del 65% i (-56.5,-48.5) que té un pes del 35%. Només es calculen les constants d'acoblament entre la  $^1J(N_{11},C_{13})$  i la  $^2J(C_{13},N_{21})$  ja que són les que presentaven un error més elevat i les que tenen un marge més gran de millora.

Els resultats obtinguts amb aquesta base estàndard es comparen amb l'efecte del solvent i amb dues bases especialitzades en el càlcul de J però que tindran una mida raonable. Aquestes dues bases són la 6-31G-J<sup>11</sup> i la ccJ-pVDZ<sup>12</sup>. S'han triat aquestes dues bases pel fet de ser d'una mida petita i perquè en els seus corresponents articles donaven uns resultats bastant acurats comparables a d'altres bases més grans. També es veurà si efectivament hem pogut millorar l'article en el qual ens basem<sup>7</sup> a on s'utilitza com a funcional el B3LYP i com a base la D95(d,p).

A les dues taules següents es pot observar quina combinació de funcional i base general donen un resultat més similar a l'experimental. L'error relatiu s'expressa en tant per cent i en valor absolut respecte el valor experimental  $^1J(N_{11}-C_{13})=-11.34$  Hz i  $^2J(C_{13}-N_{21})=-8.45$  Hz). A la taula 3 s'introdueix una comparació entre diferents bases utilitzant el funcional B3LYP. S'agafa com a referència la base cc-pVTZu ja que és la que té més funcions de base, mida més gran i és la que s'esperaria els millors resultats i es comparen totes les altres bases envers ella. A la taula 4 es realitza una comparació entre diferents funcionals que s'han utilitzat.

Taula 3. Comparació dels valors de J per diferents bases amb el funcional B3LYP. Els valors experimentals són:  $^1J(N_{11},C_{13})=-11.34$  i  $^2J(C_{13},N_{21})=-8.45$ .

J(Hz)	D95(d,p)	D95(d,p)u	cc-pVDZ	cc-pVTZ	cc-pVDZu	cc-pVTZu
	(-92,110)					
$^1J(N_{11},C_{13})$	-12.14	-11.34	-21.77	-10.44	-12.00	-11.54
$^2J(C_{13},N_{21})$	-11.66	-10.35	-9.33	-9.28	-10.67	-10.66
	(-56.5,48.5)					
$^1J(N_{11},C_{13})$	-8.43	-8.06	-18.67	-6.82	-8.60	-8.26
$^2J(C_{13},N_{21})$	-9.58	-8.52	-8.38	-7.40	-8.88	-8.80
	$\langle J \rangle$					
$^1J(N_{11},C_{13})$	-10.81	-10.15	-20.63	-9.13	-10.77	-10.35
$^2J(C_{13},N_{21})$	-10.90	-9.68	-8.96	-8.59	-10.00	-9.98
Error relatiu $^1J$	4.70	10.46	81.90	19.47	5.00	8.71
Error relatiu $^2J$	28.98	14.53	6.09	1.68	18.40	18.11
Suma errors	33.69	24.99	87.99	21.15	23.40	26.80

Es pot observar que la base modesta en termes de dimensió D95(d,p)u té uns errors comparables a la base de referència i en diferents casos combinada amb alguns funcionals és encara millor. Així doncs es compara la base D95(d,p)u amb diferents funcionals i entre parèntesis s'introdueix el valor de referència (la base cc-pVTZu)

Taula 4: comparació valors de J per diferents funcionals amb la base D95(d,p)u i cc-pVTZu (valors en parèntesis).

J(Hz)	B3LYP	tHCTH	OPBE	MPwPW91	MPW1WK
	(-92,110)				
$^1J(N_{11},C_{13})$	-11.34 (-11.54)	-9.77 (-10.25)	-7.36 (-7.34)	-7.97 (-8.04)	-12.90 (-13.13)
$^2J(C_{13},N_{21})$	-10.35 (-10.66)	-10.84 (-11.22)	-10.71 (-11.04)	-11.03 (-11.37)	-9.40 (-9.72)
	(-56.5,48.5)				
$^1J(N_{11},C_{13})$	-8.06 (-8.26)	-6.39 (-6.86)	-4.14 (-4.14)	-4.46 (-4.53)	-10.00 (-10.25)
$^2J(C_{13},N_{21})$	-8.52 (-8.80)	-8.97 (-9.32)	-8.64 (-8.93)	-9.14 (-9.46)	-7.56 (-7.82)
	$\langle J \rangle$				
$^1J(N_{11},C_{13})$	-10.15 (-10.35)	-8.59 (-9.07)	-6.23 (-6.22)	-6.74 (-9.84)	-11.88 (-12.13)
$^2J(C_{13},N_{21})$	-9.68 (-9.98)	-10.18(-10.56)	-9.98 (-10.30)	-10.37 (-10.70)	-8.75 (-9.06)
Error relatiu $^1J$	10.46 (8.71)	24.27 (20.05)	45.04 (8.71)	40.58 (39.95)	4.80 (6.93)
Error relatiu $^2J$	14.53 (18.11)	20.53 (24.94)	18.13 (45.14)	22.68 (26.92)	3.60 (7.17)
Suma errors	24.99 (26.80)	44.80 (46.05)	63.17 (21.93)	63.26 (66.57)	8.40 (14.10)

Els resultats amb el funcional MPW1WK són molt bons i combinat amb la base D95(d,p)u és el millor (sempre s'utilitza l'error per tal de poder-ho determinar). Pot simular els valors de J de manera similar a com ho fa la base estàndard cc-pVTZu amb el guany de cost computacional que això representa. Així doncs, es faran tots els càlculs de J de les 625 estructures optimitzades amb el funcional MPW1WK. Els càlculs en tot l'espai del diagrama de Ramachandran de constants d'acoblament spin-spin es realitzaran amb tres bases diferents: D95(d,p)u, 6-31G-J, ccJ-pVDZ. També es tornarà a utilitzar la base D95(d,p)u per incloure l'efecte del solvent i veure quines diferències s'observen.

## 4.2 Obtenció de quadrícules de J

**Constant d'acoblament  $^3J$ .** De la figura 6 a la 11 s'observen els resultats dels càlculs de constant d'acoblament spin-spin més rellevants realitzats en funció dels angles diedres  $\phi$  i  $\Psi$  en el format dels diagrames de Ramachandran. En aquestes figures, l'angle  $\Psi$  és el que es troba representat a l'eix vertical mentre que l'angle  $\phi$  es troba a l'eix horitzontal. La llegenda relaciona els colors de les figures amb els valors de les constants (expressades amb Hertz). Si les J són funció d'un únic angle diedre ( $\phi$  o  $\Psi$ , però no els dos) com prediu l'equació de Karplus, els diagrames es veuran com a franges horitzontals o verticals. Tanmateix, si les J estan influenciades per els dos angles la figura serà molt més complexa.

Només tres de les dotze J que s'utilitzen tenen un comportament pur de l'equació de Karplus i es veuen com a franges. Aquestes tres figures descriuen la  $^3J(C_6-H_{14})$  (Fig. 6), la  $^3J(C_6-C_{15})$  i la  $^3J(C_6-C_{16})$  que són únicament funció de  $\phi$ . Només es representa la  $^3J(C_6-H_{14})$  amb la nostre millor combinació de funcional i base com es veurà en el següent apartat (MPW1WK/D95(d,p)u) i la compararem amb la combinació que es va fer servir en l'article<sup>7</sup> (B3LYP/D95(d,p)). La figura a) és la situada a l'esquerre mentre que la b) es troba a la dreta.

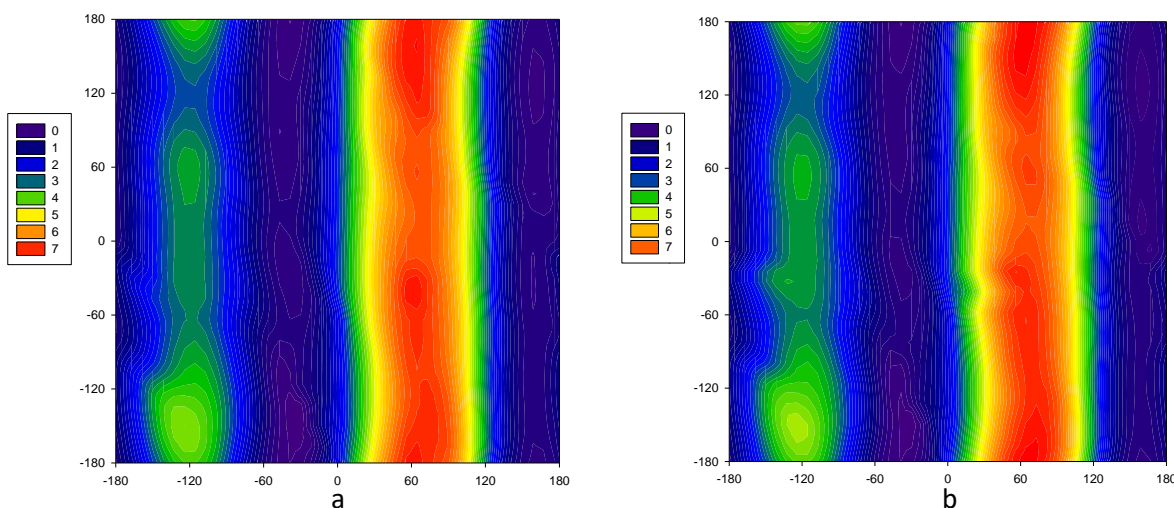


Figura 6:  $^3J(C_6-H_{14})$  en funció de  $\phi$  i  $\Psi$ . a) MPW1WK/D95(d,p)u i b) B3LYP/D95(d,p).

Es pot observar com la forma dels dos diagrames de Ramachandran és similar i només varia substancialment en la zona on  $\phi=-120^\circ$  i  $\Psi=-60^\circ$ . Com es veurà més endavant no és una constant d'acoblament conflictiva ni determinant a l'hora de determinar l'estructura però és un clar exemple d'una J que sí segueix l'equació de Karplus. Tot seguit es representa la  ${}^3J(H_{12}-H_{14})$  (Fig. 7) amb la mateixa combinació de funcionals i base que s'acaba d'explicar. Es tracta de la clàssica constant d'acoblament a tres enllaços que involucra dos hidrògens i que és útil per veure que no segueix l'equació de Karplus ja que exhibeix una completa dependència tant de  $\phi$  com de  $\Psi$ .

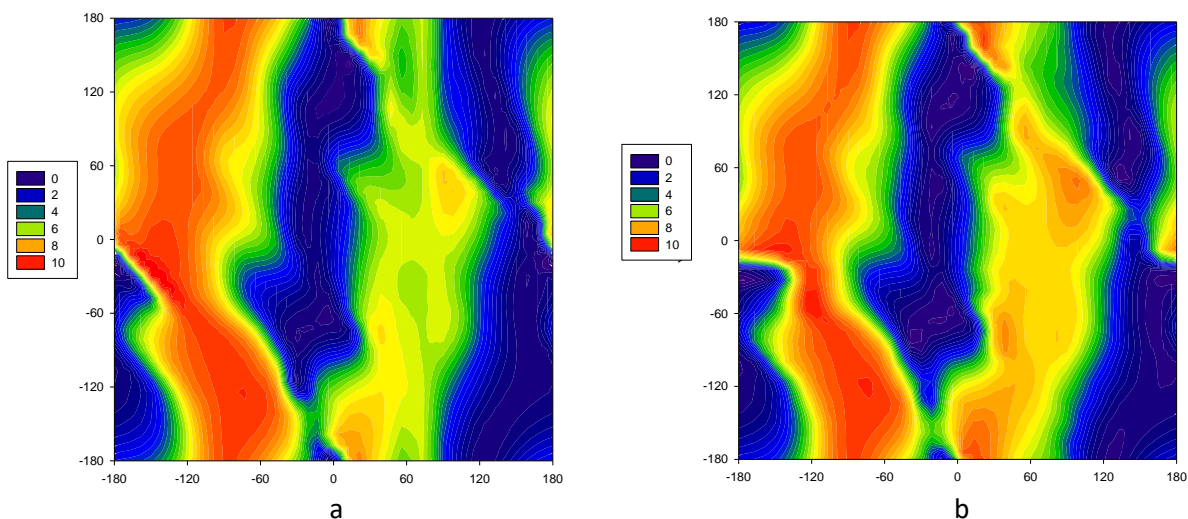


Figura 7:  ${}^3J(H_{12}-H_{14})$  en funció de  $\phi$  i  $\Psi$ . a) MPW1WK/D95(d,p)u i b) B3LYP/D95(d,p).

Altres cop s'observa una forma molt similar però amb diferències a  $\phi=-120^\circ$  i  $\Psi=-60^\circ$  i  $\phi=45^\circ$  i  $\Psi=140^\circ$ . La constant d'acoblament que sí que presenta bastanta conflictivitat és la  ${}^3J(C_{13}-H_{22})$ . Aquesta és la J més elevada quan s'utilitza el nostre funcional MPW1WK. Per aquesta raó representarem totes les combinacions possibles que hem fet tant de funcional com de base.

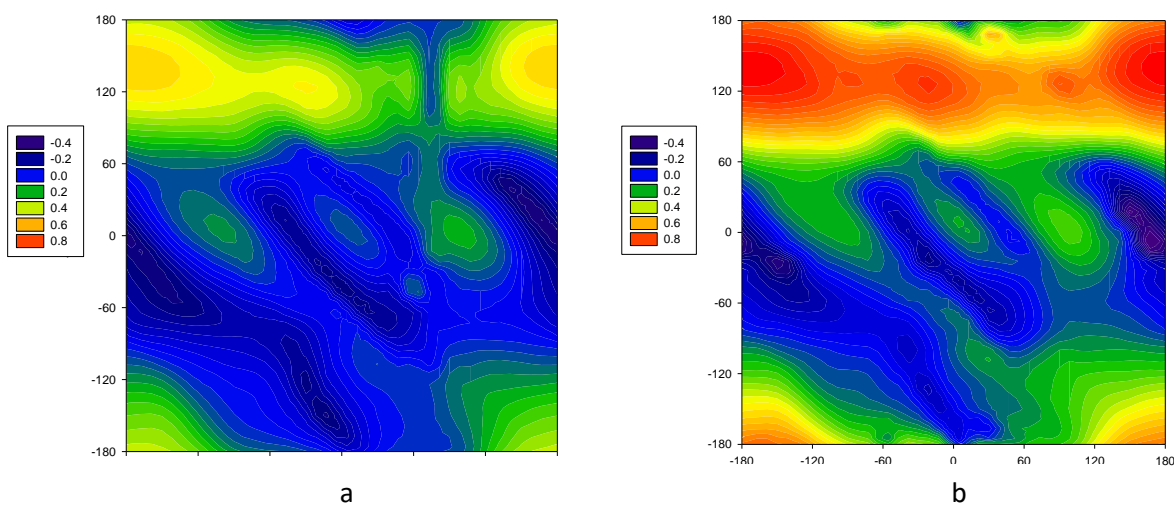


Figura 8:  ${}^3J(C_{13}-H_{22})$  en funció de  $\phi$  i  $\Psi$ . a) MPW1WK/D95(d,p)u i b) B3LYP/D95(d,p).

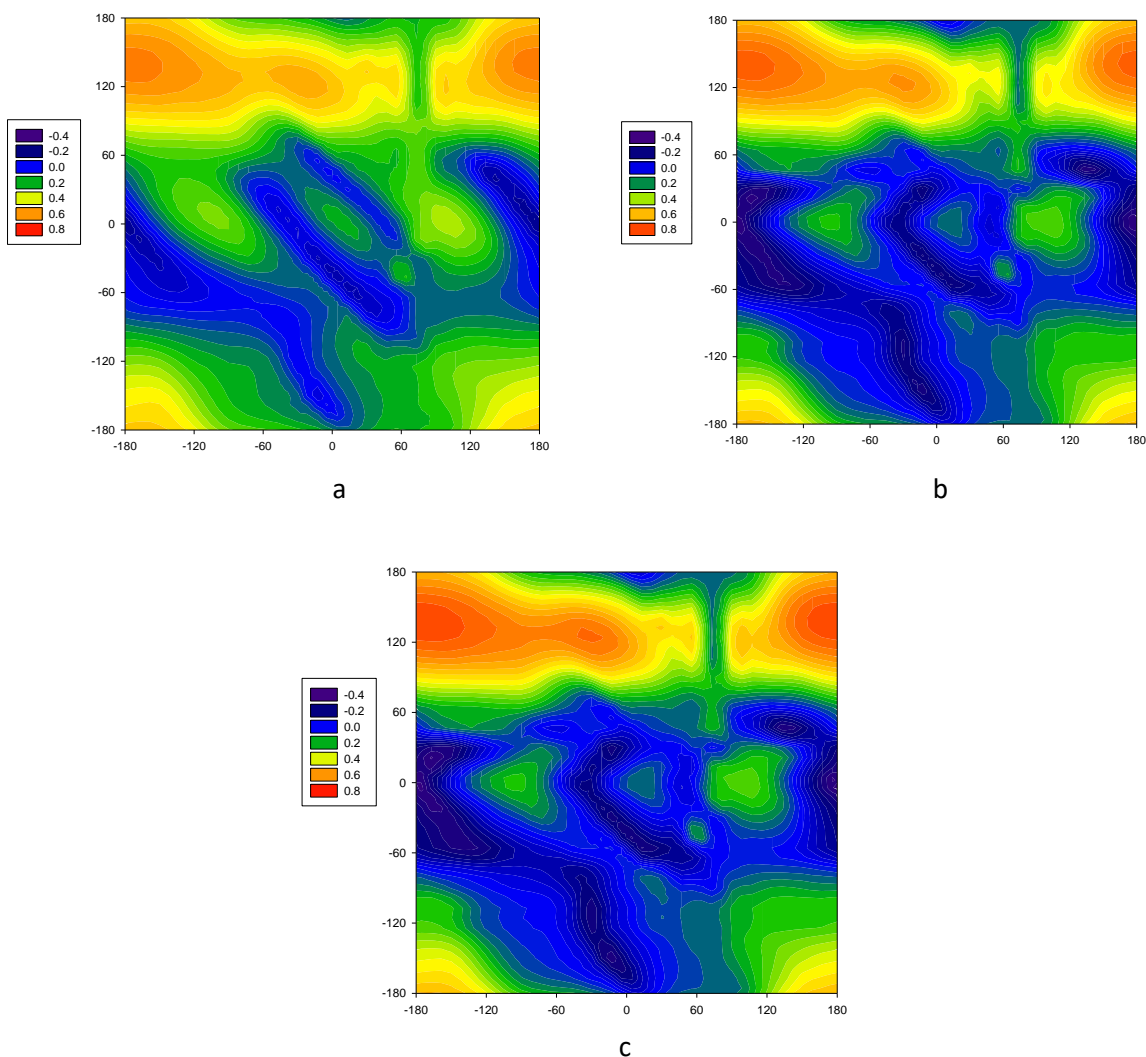


Figura 9:  ${}^3J(C_{13}-H_{22})$  en funció de  $\phi$  i  $\psi$ . a)MPW1WK/D95(d,p)u amb solvent, b)MPW1WK /6-31G-J i c)MPW1WK /ccJ-pVDZ.

Tal i com s'indica a la taula 6 el valor experimental de la constant d'acoblament spin-spin representada és de 0.65 Hz. Si ens fixem en la figura 8a) el valor màxim és de 0.5 Hz i per tant no és possible que combinant diferents regions el valor interpolat i ponderat sigui pròxim al 0.65 Hz. El mateix passa amb la inclusió de solvent de la figura 9a). Tot i l'error d'aquesta constant d'acoblament individual amb la combinació de funcional i base MPW1WK/D95(d,p)u, veurem en el següent apartat que les altres  $J$  s'ajusten molt bé al valor experimental i és la combinació que presenta unes estructures teòriques més semblants a les experimentals i per tant un dels objectius futurs pot ser la de millorar aquesta constant d'acoblament sense disminuir la precisió amb les altres  $J$ . Utilitzant com a bases la 6-31G-J i la ccJ-pVDZ ens podem aproximar més el valor de 0.65 Hz ja que s'observa que els valors són més elevats i més pròxims a l'experimental. Tot i així quan s'aplica l'algorisme i s'observa el RMSD és massa elevat. En conclusió, la nova metodologia que s'utilitza empitjora els resultats d'aquesta  $J$  individual respecte a l'utilitzada en la referència 7. Les grans diferències de forma



entre totes les figures són a la zona central del diagrama de Ramachandran ( $\phi$  entre  $-50^\circ$  i  $+50^\circ$  i  $\psi$  entre  $-70^\circ$  i  $+50^\circ$ ).

**Constant d'acoblament  $^2J$ .** La figura 10 ensenya la dependència dels valors de la constant d'acoblament geminal  $^2J(C_{13}-N_{21})$  en funció dels dos angles,  $\phi$  i  $\psi$ , que és la que presenta un error més gran en l'article<sup>7</sup>. Es tracta d'una constant d'acoblament que involucra àtoms de la cadena principal de la molècula. Altre cop es representarà la nostra millor combinació de funcional i base (MPW1WK/D95(d,p)u) i la combinació de l'article (B3LYP/D95(d,p)) per tal de comparar-los i veure si efectivament s'aconsegueix millorar aquesta constant.

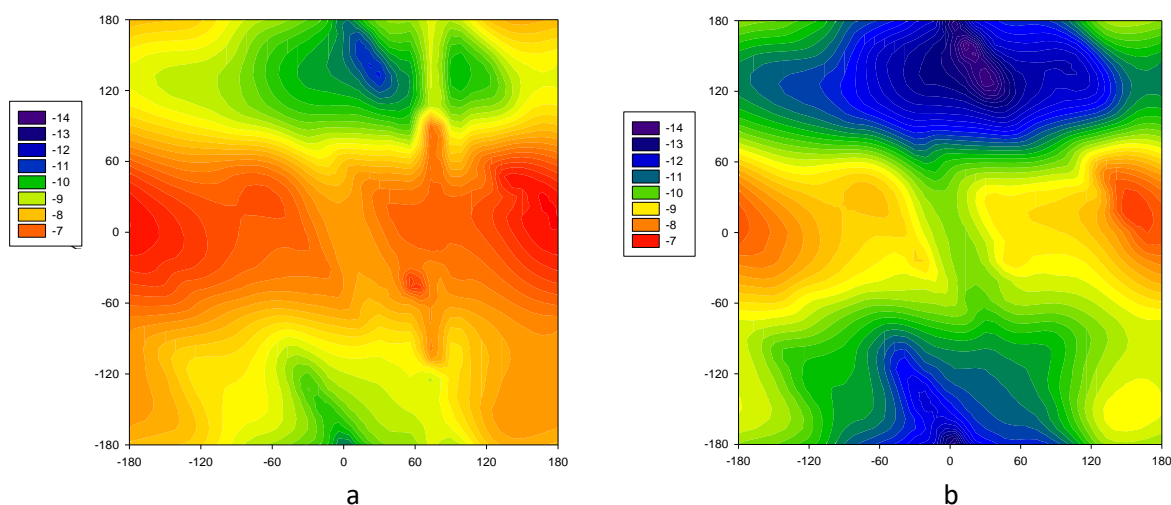


Figura 10:  $^2J(C_{13}-N_{21})$  en funció de  $\phi$  i  $\psi$ . a) MPW1WK/D95(d,p)u i b) B3LYP/D95(d,p).

El valor experimental de la constant d'acoblament geminal és de 8.45 Hz i anteriorment en l'article<sup>7</sup> quan s'interpolaven les diferents zones amb la combinació de funcional i base que s'utilitzava s'obtenia una J calculada de 10.89 Hz, per tant hi havia un error molt elevat. Es pot observar que amb els valors de J de la figura 10a) serà més fàcil trobar una solució que interpoli constants de diferents regions i el seu valor tingui un error més petit que l'esmentat anteriorment. Aquesta és la J que més ha millorat canviant el funcional de B3LYP a MPW1WK. A la figura 10b) hi ha zones que s'obté una J amb el valor -14 Hz mentre que la figura 10a) està molt més limitat entre -7Hz a -10 Hz.

La forma de les figures en tot el diagrama de Ramachandran torna a ser molt semblant amb una petita diferència a  $\phi +90^\circ$  i  $\psi$  entre  $+150^\circ$  i  $+50^\circ$ . El més important és el canvi de valors entre una figura i altra.

**Constant d'acoblament  $^1J$ .** La última constant d'acoblament representada és la  $^1J(N_{11},C_{13})$ . Com s'ha fet anteriorment es representa a la figura 11a) les noves dades obtingudes mentre que a la figura 11b) s'observen les dades les quals tenen capacitat de millora. Els valors de les J en els dos diagrames també depenen tant de  $\phi$  com de  $\psi$ .

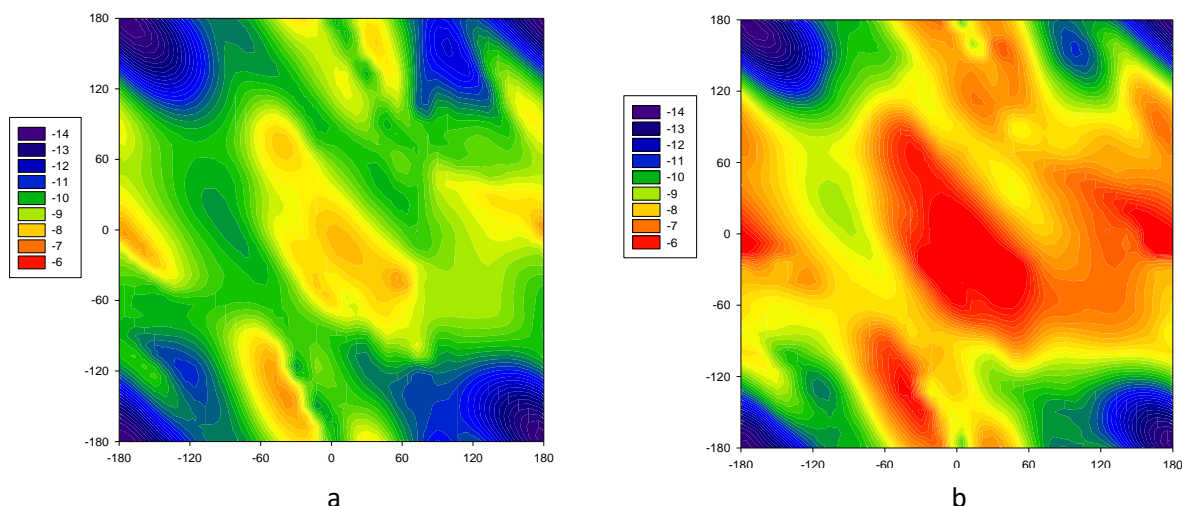


Figura 11:  $^1J(N_{11}-C_{13})$  en funció de  $\phi$  i  $\psi$ . a) MPW1WK/D95(d,p)u i b) B3LYP/D95(d,p).

El valor experimental de  $^1J(N_{11},C_{13})$  és de -11.34 Hz. Quan s'executava el programa amb les dades de la figura 11b) donava una J calculada de 9.14 Hz amb un gran error. En aquest cas podem tornar a veure com els valors de les noves dades són molt més pròxims al valor experimental ja que el seu valor oscil·la entre -8 Hz i -13 Hz mentre que a la figura 11a) és de -5 Hz a -13 Hz. Serà més fàcil interpolar valors de diferents regions amb les dades noves i que el resultat s'aproximi a -11.34 Hz que no interpolar valors de diferents regions amb les J de la figura 11a). En el següent apartat veurem que efectivament és així.

### 4.3 Estructura secundària del pèptid Acetil(Ala)<sub>3</sub>NH<sub>2</sub>

En aquest punt les quadrícules de J s'han recalculat amb el funcional MPW1WK i les bases D95(d,p)u (amb efecte de solvent i sense), 6-31G-J i ccJ-pVDZ per tot l'espai conformacional de l'aminoàcid central de l'Ala<sub>3</sub>. Els resultats que s'obtenen del programa de cerca de les estructures ens indiquen que les dues primeres combinacions donen resultats molt similars (és a dir, el pes de les diferents estructures és similar i les estructures trobades estan situades a angles  $\phi$  i  $\psi$  pràcticament anàlegs en els dos diagrames) però la inclusió de solvent fa que l'error augmenti molt considerant el RMSD com aspecte determinant. El mateix passa amb els resultats obtinguts amb les bases 6-31G-J i ccJ-pVDZ. En ambdós casos s'obtenen estructures

secundàries a  $\phi$  i  $\Psi$  molt similars, però la base 6-31G-J és molt més acurada si observem l'error. Per tant el que es farà serà comparar els resultats obtinguts amb les bases D95(d,p)u i la 6-31G-J.

En primer lloc es considera l'espai de Ramachandran sencer dividit amb quatre quadrants. Els resultats obtinguts combinant fins a quatre estructures diferents (una de cada regió) indiquen que només els quadrants assenyalats com 1 i 2 a les figures 12 i 13 amb  $\phi < 0$  contenen estructures amb factors de pes importants. Tot i això s'observen diferents importants entre els càlculs fent servir la base D95(d,p)u i la 6-31G-J. En aquest últim cas s'obtenen estructures que no es corresponen a les conformacions esperades. En canvi amb la base D95(d,p)u s'obté una important contribució d'una estructura hèlix (a la regió 2) i una contribució menor d'estructures de tipus PpII a la regió 1. Tanmateix per tal de descartar que aquest resultat sigui un efecte d'haver triat aquestes regions en una segona etapa es consideren unes altres regions del diagrama de Ramachandran que englobin les estructures secundàries habituals.

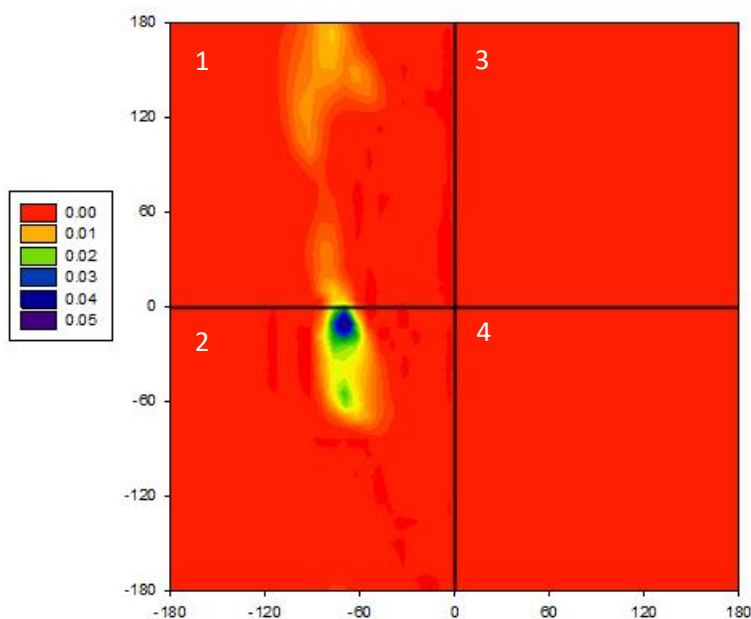


Figura 12. Distribució de probabilitat fent servir quatre quadrants que cobreixen tot el diagrama de tipus Ramachandran amb les  $\langle J \rangle$  calculades amb la base D95(d,p)u.

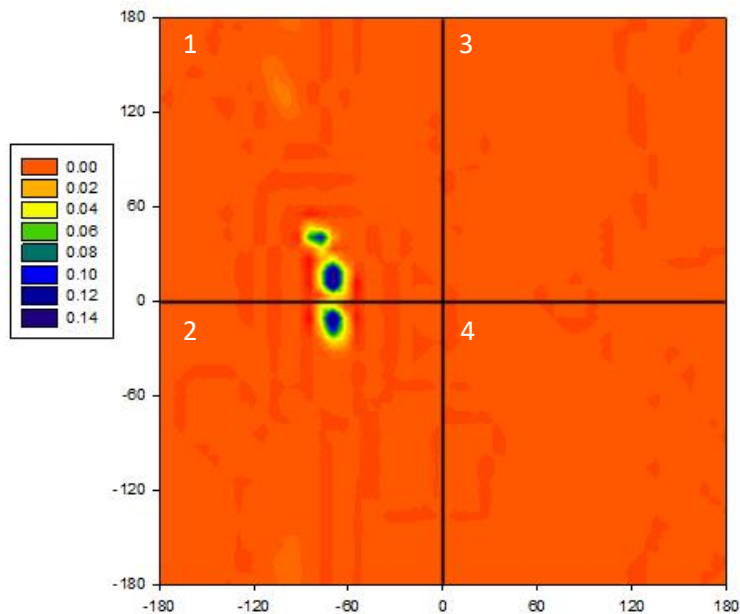


Figura 13. Distribució probabilitat fent servir quatre quadrants que cobreixen tot el diagrama de tipus Ramachandran amb les  $\langle J \rangle$  calculades amb la base 6-31G-J.

Així doncs, es divideix una part del diagrama de Ramachandran amb cinc parts d'igual àrea de manera que cada part hagi de contenir una estructura i així no hi hagi solapament. S'introdueix una regió on  $\phi > 0$  (regió 5) per tenir en compte una hipotètica contribució d'una estructura hèlix alfa de tipus *left-handed*. La forma d'aquestes regions es pot observar a les figures 14 i 15.

Un cop definides les noves regions, es pretén determinar com afecten tots els paràmetres que aplica l'algorisme a l'hora de fer la cerca exhaustiva que determina l'estructura secundària del pèptid d'Ala<sub>3</sub>. Es realitzen diferents proves per veure quina resolució en angles i pesos és la millor i després s'introdueix un error màxim per cada J individual de manera que podem veure com influeix en el número de solucions i en el RMSD. Com s'ha comentat anteriorment s'observa que hi ha dues J que empitjoren bastant el resultat en termes d'error que són la  $^3J(C_{13}-H_{22})$  i la  $^1J(N_{11}-C_{13})$ . Quan es baixen els dos errors individuals d'aquestes dos J al nivell de 0.35 Hz (valor permès per totes les altres 10 J) no s'observen solucions possibles amb un RMSD inferior a 0.5. Per tant es va tenir que augmentar el valors dels dos errors permesos per tal de trobar possibles solucions encara que el corresponent RMSD augmentés. L'error utilitzat per la  $^3J(C_{13}-H_{22})$  és de 0.4 Hz mentre que en el cas de la  $^1J(N_{11}-C_{13})$  és de 0.8 Hz. Així, s'explora la dependència de les resolucions dels angles  $\delta\phi$  i  $\delta\psi$  (es proven cada 5 i 10 graus), la resolució dels pesos  $\delta w$  (0.02 i 0.05) i el  $RMSD_{m\grave{a}x}$  que s'accepta (0.40).

Els resultats obtinguts es poden observar a la taula 5 on n és el nombre de zones que desitgem combinar i Q són el número de solucions que compleixen totes les condicions abans esmentades.

Taula 5: resum dels resultats dels quatre ajustos comparant les dues millors bases.

Ajust	$\delta(\phi, \psi)$	$\delta w$	RMSD <sub>màx</sub>	N	Q	RSMD
MPW1WK/D95(d,p)u						
1	10	0.05	0.40	No solució possible		
2	5	0.05	0.40	3	12	0.3242
3	10	0.02	0.40	4	10	0.3332
4	5	0.02	0.40	3	46	0.3208
MPW1WK/6-31G-J						
	10	0.05	0.40	2	34	0.2489
1	10	0.05	0.40	3	14058	0.2327
	10	0.05	0.40	4	1805839	0.2281
2	5	0.05	0.40	2	26	0.2601
	5	0.05	0.40	3	5242	0.2403
	5	0.05	0.40	4	388265	0.2358
3	10	0.02	0.40	2	285	0.2460
	10	0.02	0.40	3	257270	0.2402
4	5	0.02	0.40	2	487	0.2348
	5	0.02	0.40	3	699992	0.2372

En el cas del mètode MPW1WK/D95(d,p)u el resultat amb un RSMD es troba per la prova 4 a on es troben fins a 46 combinacions diferents possibles involucrant tres regions diferents. En el cas del mètode MPW1WK/6-31G-J s'obtenen RMSD més baixos que indiquen un millor ajust a les dades experimentals i moltes més combinacions possibles d'estructures de fins a quatre regions diferents. Com que en el procés d'ajust s'han obtingut diferents combinacions d'estructures que compleixen els criteris el que es pot fer és determinar la distribució de probabilitat dels angles diedres per cada regió . Això representar les figures 14 i 15 pels dos mètodes emprats.

A la figura 14 s'observen tres zones clarament diferenciades i importants com a resultat del procés d'ajust a les J experimentals. Aquestes corresponen a contribucions d'estructura lamina  $\beta$  (regió 1), PpII (regió 2) i alfa hèlix (regió 4). Contràriament, amb la base 6-31G-J (figura 15) s'observa bé l'alfa hèlix però les estructures on  $\psi > 90$  no tenen cap pes i, per contra, apareix una nova estructura a la zona 3 situada a la zona superior de l'alfa hèlix que no es correspon amb cap conformació esperable i per tant no podem considerar que sigui un resultat bo. D'aquesta manera la combinació de la base 6-31G-J i funcional MPW1WK no dona resultats correctes com tampoc la combinació del funcional MPW1WK amb la base ccJ-pVDZ que s'ha descartat degut al seu gran error i no es discuteixen els seus resultats.

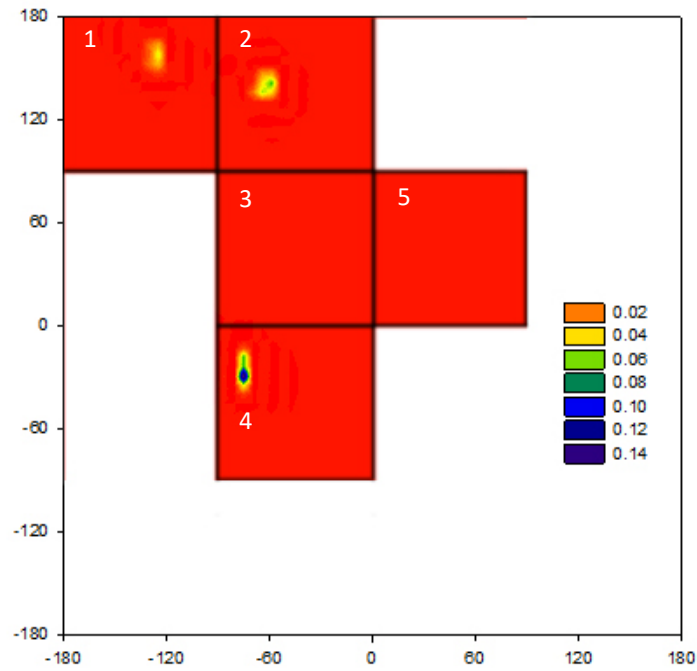


Figura 14: distribució de probabilitat fent servir cinc regions que cobreixen les zones a on les estructures tenen pes amb la base D95(d,p)u.

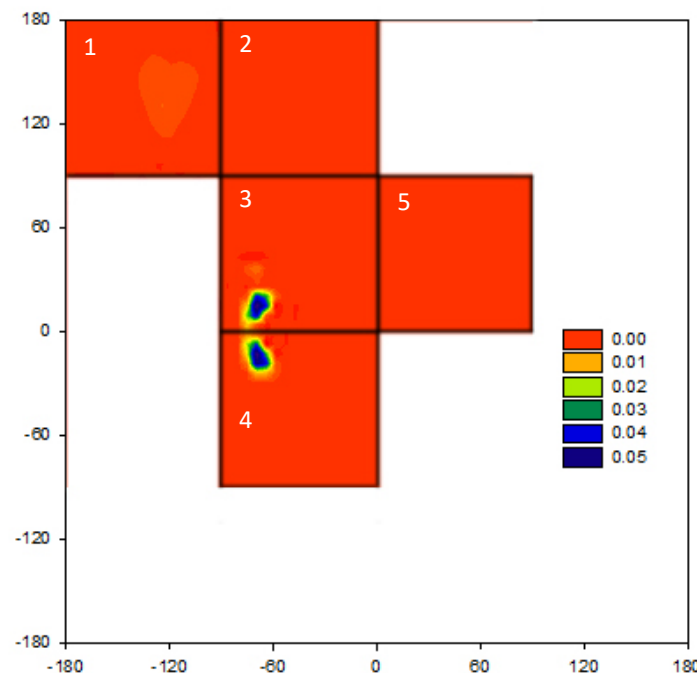


Figura 15: distribució de probabilitat fent servir cinc regions que cobreixen les zones a on les estructures tenen pes amb la base 6-31G-J.

A partir de la distribució de probabilitats es poden calcular les estructures promig de cada regió i el seu pes. Els resultats obtinguts es troben a la taula 6. En el cas del mètode MPW1WK/D95(d,p)u, a la zona on  $\phi=-130^\circ$  i  $\Psi=160^\circ$  tenim una estructura de làmina  $\beta$  amb un pes de 0.24. La zona situada a  $\phi=-60^\circ$  i  $\Psi=140^\circ$  es correspon a l'estructura de Pp II que pondera 0.35. S'ha de recordar que el pes de PpII ha sigut

molt qüestionat entre els diferents articles <sup>3-6</sup> i s'observa que el pes calculat, en aquest treball, d'aquesta estructura dista bastant dels 0.84 o 0.92 que s'indiquen en els articles <sup>3,4,6</sup> però està en bastanta concordança amb el 0.3 que s'afirma a la referència 5. Tot i així el nostre objectiu era aconseguir aquestes estructures i a partir d'elles determinar-ne el pes mentre que en els seus articles <sup>3-6</sup> ja van suposar que aquestes estructures eren les principals i van partir directament d'aquestes. D'aquesta manera el mètode utilitzat ha sigut diferent. Finalment quan  $\phi = -75^\circ$  i  $\Psi = -35^\circ$  l'estructura és d'alfa hèlix amb un pes de 0.41, és a dir, és la conformació amb un pes més elevat.

Taula 6. Valors mitjans de  $\phi$  i  $\Psi$  i les corresponents desviacions estàndards ( $\sigma$ ) per la millor aproximació en termes de RMSD (veure taula 5). Les J de la taula 6 s'han calculat amb aquests pesos de ponderació.

Regió	Pes	$\phi$	$\sigma(\phi)$	$\Psi$	$\sigma(\Psi)$
MPW1WK/D95(d,p)u					
1	0.238	-127.0	3.6	157.5	5.3
2	0.349	-62.3	3.3	139.8	3.9
3	0.000	0.0	0.0	0.0	0.0
4	0.413	-75.0	0.0	-26.7	5.9
5	0.000	0.0	0.0	0.0	0.0
MPW1WK/6-31G-J					
1	0.196	-123.3	15.1	141.3	17.3
2	0.039	-58.9	24.5	129.4	23.8
3	0.368	-69.3	9.1	16.4	12.8
4	0.372	-67.7	8.7	-17.0	10.3
5	0.025	38.8	29.9	38.3	31.6

Finalment, els valors teòrics per cada J es calculen com la mitjana ponderada entre els valor de la J per les estructures mitjanades ponderades amb els pesos corresponents. Els resultats obtinguts amb la millor combinació possible de cada base (en termes de RMSD) es mostren a la taula 7 juntament amb les dades experimentals amb les que s'ha fet l'ajust i els resultats obtinguts a la referència 7. Els ajustos d'aquest treball utilitzen totes les J experimentals, mentre que l'ajust fet a la referència 7 no inclou les dues últimes J ja que donaven un gran error. De fet si s'inclouen totes les 12 J en el càlcul de RMSD es veu que el valor obtingut és molt més gran que els ajustos fets amb aquest treball. D'aquesta manera doncs, els resultats es milloren bastant, sobretot pel que fa a la <sup>2</sup>J (13,21) i la <sup>1</sup>J (11,13). La base 6-31G-J dóna un RSMD menor que no pas la base D95(d,p)u amb el nou funcional, tot i així, quan mirem la figura 15 s'observa que les zones representatives que es troben no es corresponen amb cap conformació esperable com ja s'ha comentat anteriorment. Així doncs el baix error es pot correspondre a que matemàticament sigui possible, un ajust vàlid de les 12 J experimentals.

Comparant els resultats obtinguts amb els dos mètodes, la zona 1 i 4 tenen pesos similars i la diferència rau en la regió on es troba el pes restant. Mentre que amb la base D95(d,p)u la zona 2 és la important, amb la base 6-31G-J aquest pes es troba a la 4 a on mirant els diferents resultats experimentals que hi ha publicats i els estudis que s'han fet això no hauria de ser així i no és possible. És per això que encara que el seu RMSD és inferior a la base D95(d,p)u, es descarta ja que la solució que s'obté no s'acorda amb les estructures esperades.

.Taula 6: comparació de les J experimentals i calculades. a)MPW1WK/D95(d,p)u, b)MPW1WK/6-31G-J i c)B3LYP/D95(d,p). J calculades a partir de la ponderació de la taula 7.

J	Exp. (Hz)	Calc (Hz) <sup>a</sup>	Calc (Hz) <sup>b</sup>	Calc (Hz) <sup>c</sup>
<sup>3</sup> J 12,14	5.68±0.03 <sup>d</sup>	5.54	5.72	5.65
<sup>3</sup> J 6,14	1.84±0.07 <sup>d</sup>	1.52	1.95	1.75
<sup>3</sup> J 12,16	1.13±0.00 <sup>d</sup>	0.94	1.22	1.09
<sup>3</sup> J 12,15	2.39±0.03 <sup>d</sup>	2.07	2.76	2.34
<sup>3</sup> J 6,16	0.25±0.01 <sup>d</sup>	0.54	0.37	0.37
<sup>2</sup> J 13,22	0.65±0.06 <sup>d</sup>	0.26	0.27	0.49
<sup>2</sup> J 11,16	0±0.1 <sup>e</sup>	0.17	0.09	0.08
<sup>h2</sup> J 6,22	0±0.1 <sup>e</sup>	0.03	0.26	0.07
<sup>h3</sup> J 6,21	0±0.2 <sup>e</sup>	0.03	0.02	0.03
<sup>3</sup> J 6,15	2.80±0.1 <sup>e</sup>	2.48	2.63	2.82
<sup>2</sup> J 13,21	8.45±0.02 <sup>d</sup>	8.48	8.63	10.89
<sup>1</sup> J 11,13	11.34±0.03 <sup>d</sup>	10.69	11.88	9.14
RMSD		0.320	0.228	0.977

<sup>d</sup> Informació suplementària de la referència 3.

<sup>e</sup> Informació de la referència 7.

Així doncs, la combinació del funcional MPW1WK i la base D95(d,p)u en els càlculs de J d'aquest treball és la millor ja que reproduïx perfectament les tres estructures representatives com són l'alfa hèlix, làmina  $\beta$  i l'hèlix de PpII, donant un error d'ajust lleugerament superior (0.3) però acceptable.



## 5. Conclusions

We propose a method for the determination of secondary structure of peptides based on the fit of coupling constants to their values calculated using DFT (Density Functional Theory). We don't assume any model structures and we use the entire Ramachandran space. This method has been applied for the structure determination of acetyl(Ala)<sub>3</sub>NH<sub>2</sub> with reasonable success. We expect it to be generally useful for larger structures in the future.

Most J coupling constants of the central amino acid of Ala<sub>3</sub> peptide model depend on the two dihedral angles  $\phi$  and  $\Psi$ . The three exceptions,  $^3J(\text{C}_6\text{-H}_{14})$ ,  $^3J(\text{C}_6\text{-C}_{15})$  and  $^3J(\text{C}_6\text{-C}_{16})$  are a function of only of  $\phi$ . For this reason the Karplus equation cannot always apply. On the other hand, for  $^1J$  and  $^2J$  constants, Karplus equation can not be applied and only DFT methods are of use.

We use different functionals and basis sets to compute the J values over the whole Ramachandran space. The choose of functional is shown to be essential. The effect of uncontracted basis it's also important. The smaller and specialized in the J's calculation basis sets such as 6-31G-J and ccJ-pVDZ don't give better results than the standards basis set D95(d,p) uncontracted used in this work. The solvent effect it is worst because when we interpolate the J's of the different zones the consequence error is higher than without solvent. We discarded the solvent in future calculations. The final result with the functional MPW1WK and the basis set D95(d,p)u is the best. We obtain three clearly different zones that correspond to alpha helix,  $\beta$ -pleated sheet and the polyproline helix structures. Even so, the error is close to 0.3 and we cannot considerer this error a perfect result; we are working on the calculations of J's with the MPW1WK functional and the aug-cc-pVTZ-J big basis set with the aim of reduce further the error. Nevertheless,It is clear that with the functional MPW1WK the results improve compared to the B3LYP.

In this work we demonstrate that structural determination of peptides can be made with good accuracy by comparing the calculated and experimental J's.

## 6. Referències

- 1) Hore, P.J. *Nuclear Magnetic Resonance* (1a ed.). 1995. Oxford: Oxford University.
- 2) M. J. Minch, *Concepts in Magnetic Resonance*, 1994, **6**, 41.
- 3) J. r. Graf, P.H. Nguyen, G.Stock i H. Schwalbe, *J. Am. Chem. Soc.*, 2007, **129**, 1179.
- 4) R. Schweitzer-Stenner, *J. Phys. Chem. B*, 2009, **113**, 2922
- 5) Z. Shi, C.A. Olson, N. R Kallenbach et al., *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 9190.
- 6) J.A. Vila, H. A. Baldoni, D.R. Ripoll et al., *Biophysical Journal*, 2004, **86**, 731.
- 7) P. Salvador, N. Juranic i J. Dannenberg, *J. Phys. Chem*, submitted al 2015.
- 8) M. J. Frisch, et al., *GAUSSIAN 09 (Revision A.2.)*, Gaussian, Inc., Wallingford, CT, 2009.
- 9) N. F. Ramsey, *Physical Review*, 1953, **91**, 303.
- 10) J Vaara, J. Jokisaari, R. E. Wasylishen and D. L. Bryce, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2002, **41**, 233.
- 11) H. Kjaer i S.P.A Sauer, *J. Chemi Theory and Comput*, 2011, **7**, 4070.
- 12) U. Benedikt, A.A.Auer i F. Jensen, *J. Phys. Chem*, 2008, **129**, 064111.
- 13) P. Salvador, M. I.–H. Tsai i J.J. Dannenberg, *J. Phys. Chem*, 2011, **13**, 17484.