Universitat de les Illes Balears

TRABAJO FIN DE GRADO

ESTUDIO COMPUTACIONAL DE LAS INTERACCIONES DE ENLACE DE CALCÓGENO EN COMPLEJOS PROTEÍNA-LIGANDO

Jose Antonio Fernández Riveras

Grado de Química

Facultad de Ciencias

Año Académico 2020-21

ESTUDIO COMPUTACIONAL DE LAS INTERACCIONES DE ENLACE DE CALCÓGENO EN COMPLEJOS PROTEÍNA-LIGANDO

Jose Antonio Fernández Riveras

Trabajo de Fin de Grado

Facultad de Ciencias

Universidad de las Illes Balears

Año Académico 2020-21

Palabras clave del trabajo:

Enlace de calcógeno, cálculos DFT, Análisis del Protein Data Bank

Nombre Tutor/Tutora del Trabajo Antonio Bauzá

Nombre Tutor/Tutora (si procede)

	Autor		Tutor	
Se autoriza la Universidad a incluir este trabajo en el Repositorio Institucional para su consulta en acceso abierto y difusión en línea, con fines exclusivamente académicos y de investigación	Sí	No	Sí	No
	☒	☐	☒	☐

Resumen

En este trabajo se ha estudiado de forma teórica la capacidad de ligandos que presentan átomos de Se y Te en su estructura para experimentar interacciones de enlace de calcógeno con proteínas. Tras una búsqueda inicial en la base de datos Protein Data Bank (PDB), se procedió al análisis de los complejos proteína-ligando seleccionados mediante cálculos teóricos al nivel de teoría B3LYP/6-311+G**. Para ello se llevó a cabo la creación de modelos teóricos de cada complejo y se analizó la fortaleza y direccionalidad de los enlaces de calcógeno. Finalmente, la interacción se caracterizó combinando las metodologías de “Atoms in Molecules” (AIM), Natural Bonding Orbital (NBO) y Noncovalent Interactions plot (NCIplot).

Resum

En aquest treball s’ha estudiat de forma teòrica la capacitat de lligands que presenten àtoms de Se i Te a la seva estructura per experimentar interaccions d’enllaç de calcògen amb proteïnes. Desprès de una recerca inicial a la base de dades Protein Data Bank (PDB), es va dur a terme l’anàlisi dels complexes proteïna-lligand seleccionats mitjançant càlculs teòrics al nivell de teoria B3LYP/6-311+G**. Per tal d’aconseguir-ho, es varen crear models teòrics de cada complex i es varen analitzar la fortalesa i direccionalitat de la interacció d’enllaç de calcògen. Finalment, la interacció es va caracteritzar combinant les metodologies de “Atoms in Molecules” (AIM), Natural Bonding Orbital (NBO) i Noncovalent Interactions plot (NCIplot).

Abstract

In this work, the ability of ligands bearing Se and Te atoms in their structure to establish chalcogen bonding interactions with protein residues has been theoretically studied. After a preliminary search in the Protein Data Bank (PDB), the selected protein-ligand complexes were analyzed by means of theoretical calculations at the B3LYP/6-311+G** level of theory. To achieve this, a theoretical model of each protein-ligand system was built and the strength and directionality of the chalcogen bonds evaluated. Finally, the interaction was further characterized by combining “Atoms in Molecules” (AIM), Natural Bonding Orbital (NBO) and Noncovalent Interactions plot (NCIplot) methodologies.

Índice

1. Introducción. 1

2. Objetivos y plan de trabajo 5

2.1. Objetivos 5

2.2. Plan de trabajo 5

3. Metodología. 6

4. Discusión de resultados. 12

5. Conclusiones 21

6. Publicación 22

1. Introducción.

La química supramolecular se define como “la química más allá de la molécula”. Está basada en el estudio de las interacciones intermoleculares de carácter no covalente, cubriendo de esta manera aspectos estáticos, como el reconocimiento molecular y aspectos dinámicos como los fenómenos de extracción y transporte selectivo a través de membranas.[1] Se trata de un campo de la ciencia altamente interdisciplinario que cubre características químicas, físicas y biológicas de las especies químicas, mantenidas juntas y organizadas por medio de interacciones intermoleculares. Sus raíces se extienden a la química orgánica y los procedimientos sintéticos para la construcción de receptores, en química de coordinación y complejos de iones metal-ligando, en química física y los estudios experimentales y teóricos de las interacciones, y los procesos biológicos que comienzan con unión y reconocimiento de sustrato[2].

El enlace de hidrógeno es fuerza no covalente omnipresente en el campo de la Química Supramolecular, el cual se define como la interacción atractiva entre un átomo de hidrógeno de una molécula o un fragmento molecular X-H en el que X es un átomo más electronegativo que H, y un átomo o grupo de átomos de la misma molécula, o diferente, en la que haya evidencia de interacción.[3] Un enlace de hidrógeno típico se puede representar como X-H···Y-Z. X-H representa el grupo donante de enlace de hidrógeno (por ejemplo, un grupo OH, NH o FH). El aceptor puede ser un átomo neutro con pares de electrones solitarios o un anión, o un fragmento o molécula Y-Z, donde Y está unido a Z.

En biología, las interacciones de enlace de hidrógeno juegan un papel fundamental en procesos clave para la vida, como son la formación y estabilización de la doble hélice de ADN[4], el plegamiento de proteínas[5], o la estabilización de complejos proteína-ligando[6]. Dentro del campo de las proteínas, se define una jerarquía de cuatro niveles de estructura según su complejidad.[7] En primer lugar, la estructura primaria corresponde a la secuencia de aminoácidos unidos por enlace peptídico. En segundo lugar, la estructura secundaria hace referencia a disposiciones estables de la estructura primaria. Por otro lado, la estructura

[1] T. T. Cebada, Nanoquímica y nanotecnología. Nuevos materiales, polímeros y máquinas moleculares. Ed. Universidad Autónoma de Madrid. Fundación General, España, **2002**.
[2] J. -M. Lehn, *Angew. Chem. Int. Ed.,* **1990**, *29*, 1304-1319.
[3] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legnon, B. Mennucci, D. J. Nesbitt, *Pure Appl. Chem.*, **2011**, *83*, 1637-1641.
[4] P. L. Privalov, C. Crane-Robinson, *Eur. Biophys. J.*, **2020**, *49*, 315-321.
[5] G. D. Rose, R. Wolfenden, *Structure*, **1993**, *22*, 381–415.
[6] A. -C. C. Carlsson, M. R. Scholfield, R. K. Rowe, M. C. Ford, A. T. Alexander, R. A. Mehl, P. S. Ho, *Biochemistry,* **2018**, *57*, 4135-4147.
[7] N. D. Lehninger, M. M. Cox, Principios de Bioquímica. 4ª ed. Omega, España, **2005**.

terciaria hace referencia al plegamiento tridimensional del polipéptido. Finalmente, si una proteína posee dos o más subunidades polipeptídicas, se denomina estructura cuaternaria. El enlace de hidrógeno es clave para la formación de la estructuras secundaria y terciaria, tal y como demuestran diversos estudios tanto teóricos[8] como experimentales[9].

Una manera de conocer la topografía de la proteína es mediante la técnica de difracción de rayos X, en la cual se generan patrones de difracción a partir del cristal proteico para posteriormente calcular el mapa tridimensional de densidad electrónica de la proteína. Posteriormente, se asigna la localización de los átomos en la estructura en base a la densidad electrónica, siendo ésta un indicador fiable de la posición de los núcleos atómicos.[3] La utilización de dicha técnica permite la obtención de gran cantidad de información estructural sobre sistemas biológicos, lo cual ha facilitado la investigación sobre las interacciones existentes en macromoléculas biológicas más allá del enlace de hidrógeno, tanto a nivel intramolecular como a nivel intermolecular.

En concreto, durante la década pasada emergió un nuevo grupo de interacciones no covalentes basadas en los elementos pertenecientes al bloque p de la tabla periódica, denominadas interacciones de tipo σ-hole.[10] Esta familia de fuerzas no covalentes se compone de las interacciones tetrel bond (enlace tetrel)[11], pnictogen bond (enlace de pnictógeno),[12] chalcogen bond (enlace de calcógeno, ChB)[13] y halogen bond (enlace de halógeno).[14] Entre ellas, el enlace de calcógeno se ha estudiado y caracterizado tanto de forma teórica como experimental en diversas áreas de la Química Supramolecular, como son como la ciencia de materiales, catálisis, síntesis o la química medicinal.[15]

Para comprender la naturaleza de dicha interacción y aplicarla como herramienta supramolecular en nuevos contextos físicos y químicos, es necesario comprender su fundamento teórico. Dicha interacción se describe como una fuerza no covalente de carácter atractivo entre una región electrofílica asociada con un átomo de calcógeno en una molécula y una región nucleófila situada en otra molécula, o en la misma. Desde un punto de vista orbitálico, ésta se caracteriza por la donación desde un orbital enlazante/par solitario correspondiente a la especia rica de electrones a un orbital antienlazante (σ*) correspondiente al enlace Ch-X (donde X representa cualquier átomo o fragmento molecular). Un enlace de

[8] S. J. Grabowski, J. Leszczynski, Unrevealing the Nature of Hydrogen Bonds: π-Electron Delocalization Shapes H-Bond Features. Intramolecular and Intermolecular Resonance-Assisted Hydrogen Bonds. Hydrogen Bonding—New Insights, 487–512.
[9] R. P. Sheridan, R. H. Lee, N. Peters, L. C. Allen, *Biopolymers*, **1979**, *18*, 2451–2458.
[10] A. Bauzá, T. J. Mooibroek, A. Frontera, *ChempPhysChem,* **2015**, *16*, 2485-2517.
[11] A. Bauzá, T. J. Mooibroek, A. Frontera, *Angew. Chem. Int. Ed.*, **2013**, *52*, 12317-12321.
[12] S. Scheiner, *Acc. Chem. Res.*, **2013**, *46*, 288-288.
[13] W. Wang, B. Ji, Y. Zhang, *J. Phys. Chem. A*, **2009**, *113*, 8132–8135.
[14] G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati, G. Terraneo, *Chem. Rev.*, **2016**, *116*, 2478-2601.
[15] D. J. Pascoe, K. B. Ling, S. L. Cockroft, J. Am. Chem. Soc., **2017**, *139*, 15160-15167.

calcógeno típico (ChB) se denota mediante la fórmula R-Ch···A, donde Ch es el donante del ChB, siendo cualquier átomo de calcógeno (O, S, Se y Te) presentando una o varias regiones electrofílicas (déficit de densidad electrónica), R es el resto de la entidad molecular R-Ch que contiene el donante de ChB y A es el aceptor de ChB y es típicamente una región nucleófila (átomo con pares de electrones solitarios o un anión).[16]

Este se basa en la aparición de zonas de potencial electrostático positivo a lo largo de la prolongación de los enlaces covalentes X–Ch (Ch = O, S, Se y Te, ver Figura 1a). Como se puede observar, las cuatro molécules expuestas (OF_2, SF_2, SeF_2 y TeF_2) presentan dos σ-holes diferenciados, uno por cada enlace covalente Ch–F, aumentando el tamaño del σ-hole y el valor del potencial electrostático al pasar del O al Te, así como lo hace su polarizabilidad. Además, la presencia de sustituyentes electronegativos enlazados al átomo de Ch también aumenta el valor del potencial en el σ-hole (ver Figura 1b), aumentando la fortaleza de la interacción de enlace de calcógeno.

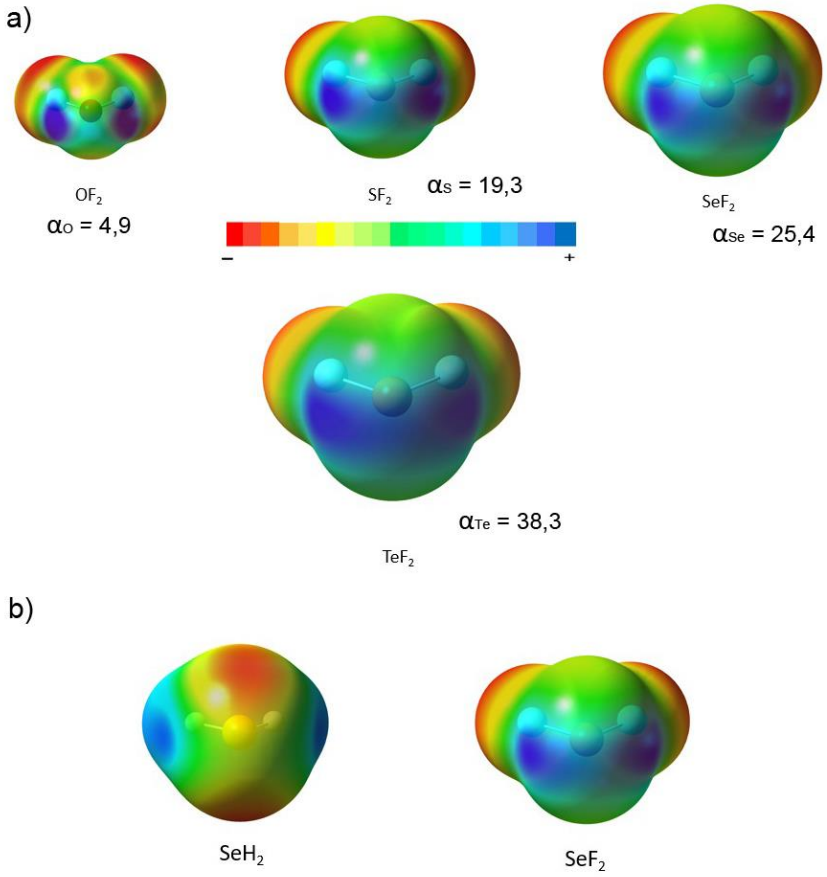

Figura 1. Mapas de potencial electrostático de las moléculas OF_2, SF_2, SeF_2 y TeF_2. Los valores de polarizabilidad para O, S, Se y Te (α) también se hallan indicados.[10] Los valores de energía en los puntos seleccionados de la superficie (0.001 u.a) se muestran en kcal/mol.

[16] C. Aakeroy, D. Bryce, G. Desiraju, A. Frontera, A. Legon, F. Nicotra, K. Rissanen, S. Scheiner, G. Terraneo, P. Metrangolo, G. Resnati, *Pure Appl. Chem.*, **2019**, *91*, 1889-1892.

En biología, los enlaces de calcógeno que se han estudiado implican mayoritariamente al S,[17] debido a su elevada abundancia con respecto a átomos de calcógeno más pesados como el Se o el Te, los cuáles se incorporan en proteínas[18] y ácidos nucleicos[19] para determinar su estructura mediante difracción de rayos X MAD (Multiwavelength Anomalous Diffraction) o SAD (Single-wavelength Anomalous Diffraction). El principal propósito dichas técnicas es la determinación de la posición de átomos pesados,[20] y se diferencian principalmente en la utilización de una única onda (SAD) o varias (MAD) para llevar a cabo el proceso de difracción cristalográfico. A pesar de las bondades dichas técnicas para la determinación estructural de macromoléculas biológicas, poco se ha ahondado en las implicaciones moleculares (a nivel de reconocimiento e interacción) resultantes de la incorporación de Se y Te en el sistema. Por ende, en el presente trabajo se han estudiado mediante cálculos teóricos al nivel de teoría B3LYP/6-311+G* las interacciones de enlace de calcógeno en complejos proteína-ligando que presenten átomos de Se y Te en su estructura. Para ello, se llevó a cabo una búsqueda de estructuras inicial en la base de datos cristalográfica Protein Data Bank (PDB)[21], de la cual se seleccionaron un set de estructuras para su posterior análisis teórico. Éste consistió en la creación de un modelo teórico del ligando de Se/Te y del aminoácido responsable del paper de donador de electrones. Sobre dichos modelos se evaluó la fortaleza de la interacción, así como su direccionalidad. Finalmente, los complejos de enlace de calcógeno se caracterizaron mediante el uso de las técnicas “Atoms in Molecules” (AIM)[22], Natural Bonding Orbital (NBO)[23] y Noncovalent Interactions plot (NCIplot)[24].

[17] K. Kříž, J. Fanfrlík, M. Lepšík, *ChemPhysChem*, **2018**, *19*, 2540-2548.
[18] J. Błaszczyk, *Biochem. Anal. Biochem.*, **2014**, *3*, 1000155.
[19] L. Lin, J. Sheng, Z. Huang, *Chem. Soc. Rev.*, **2011**, *40*, 4591-4602.
[20] F. Hernández Cano, C. Foces-Foces, M. Matinez-Ripoll, Cristalografía, Nuevas Tendencias, Madrid, **1995**.
[21] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne *Nuc. Acids Res.*, **2000**, *28*, 235-242.
[22] F. Cortes-Guzman, R. Bader. *Coord. Chem. Rev.*, **2005**, *249*, 633-662.
[23] F. Weinhold, C. R. Landis, Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective; Cambridge University Press: Cambridge, UK, **2005**.
[24] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. -P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.*, **2011**, *7*, 625-632.

2. Objetivos y plan de trabajo

2.1. Objetivos

Los objetivos de este Trabajo de Fin de Grado son:

- Buscar evidencias que demuestren la presencia de enlaces de calcógeno en sistemas proteína-ligando.
- Analizar su fuerza y direccionalidad y caracterizarlas mediante diversos análisis computacionales.

2.2. Plan de trabajo

- Inspección de la base de datos PDB sobre proteínas que contengan ligandos de selenio y teluro en su estructura.
- Obtención de mapas de potencial electrostático (MEP) del ligando de selenio/teluro.
- Obtención de modelos teóricos del complejo proteína-ligando.
- Cálculo de energía de interacción del complejo proteína-ligando.
- Análisis orbitálico (NBO), análisis de puntos críticos (AIM) y mapeado de interacciones no covalentes (NCIplot).

3. Metodología.

Búsqueda en la base de datos Protein Data Bank

Inicialmente, se realizó una búsqueda en el PDB a través del portal web EMBL-EBI[25],utilizando como único criterio de búsqueda ligandos que contuvieran Se y Te en su estructura. Un posterior análisis de la búsqueda descartó aquellas estructuras que presentasen carga negativa neta (lo cual imposibilitaba el hecho que se comportasen como especies aceptoras de electrones) y los segundos debido al gran número de estructuras presentes, lo cual dificultaba en gran medida la obtención de resultados en un tiempo razonable. El siguiente paso fue asignar cada proteína a su ligando, para lo cual se consideraron todas las estructuras de rayos X con una resolución inferior a 2.5Å.

Esto dio lugar a la obtención de 18 ligandos (15 estructuras), de los cuáles se seleccionaron 16 ligandos. Una vez seleccionados, las estructuras se analizaron mediante el programa Discovery Studio[26], lo cual consistió en la visualización de las interacciones entre el átomo de Se/Te y los residuos vecinos de la proteína, para así poder determinar la presencia o no de interacciones de enlace de calcógeno. Esto se realizó atendiendo a los siguientes criterios geométricos:

- Distancia (d) Ch···A (A = O, C) equivalente a la suma de radios de van der Waals + 0.5Å.
- Ángulo C–Ch···A (θ) entre 145 y 180º.

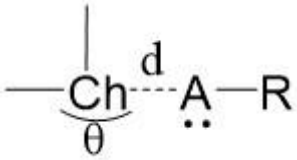

Figura 2. Esquema básico de interacción de enlace de calcógeno (θ= ángulo C–Ch···A en grados; d=distancia en Å; R= Csp^2 o Csp^3).

[25] F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter R. D. Finn, R. Lopez. *Nuc. Acids Res.,* **2019**, *47*, W636-W641.
[26] BIOVIA, Dassault Systèmes, Discovery Studio, R2 Client, San Diego: Dassault Systèmes, 2017.

De esta manera, los candidatos que cumplen estas condiciones se recogen en la Tabla 1:

Tabla 1. Candidatos para cálculos teóricos.

PDBID	IDLigando	IDAminoácido	Distancia (Å)	Ángulo C–Se···A (º)
3USP	05L610	ASP401	3.562	153.4
5FSB	6YR303	TYR154	3.871	148.5
1O9V	SNG1177	SER117	3.554	150.7
5EO8	TFU405	ASP193	3.851	134.3[a]
1EL7	MTD801	GLY344	3.072	169.6

[a]Esta estructura se consideró a pesar de la pobre direccionalidad de la interacción para así poder tener un segundo dador de electrones cargado en el estudio.

Cálculos teóricos sobre modelos proteína-ligando.

Una vez determinada la lista de complejos proteína-ligando, se llevó a cabo la creación de los modelos teóricos. Para ello se consideró únicamente el ligando de Se/Te por un lado y por otro, el aminoácido implicado en la interacción de enlace de calcógeno. Además, se modificaron los extremos de la estructura del aminoácido incorporando grupos metilo (ver Figura 3).

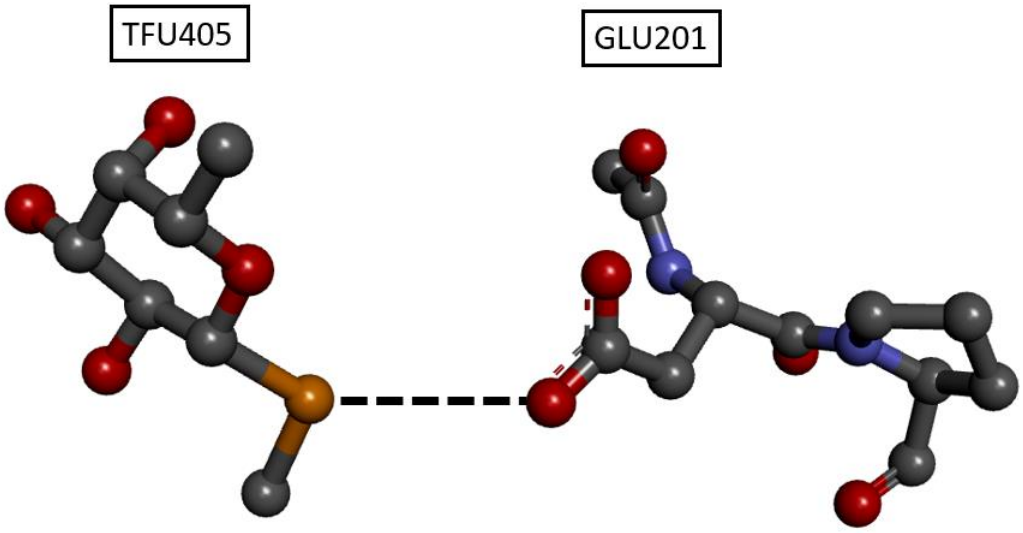

Figura 3. Modelo teórico utilizado para evaluar la interacción de enlace de calcógeno en la estructura 5EO8.

Por otro lado, también es necesario comentar que en el modelo de la estructura 5FSB se realizador dos modificaciones en el ligando de Se para poder evaluar únicamente la

interacción de enlace de calcógeno (ver Figura 4), correspondientes con la sustitución de un grupo alcohol y un grupo éter por hidrógenos.

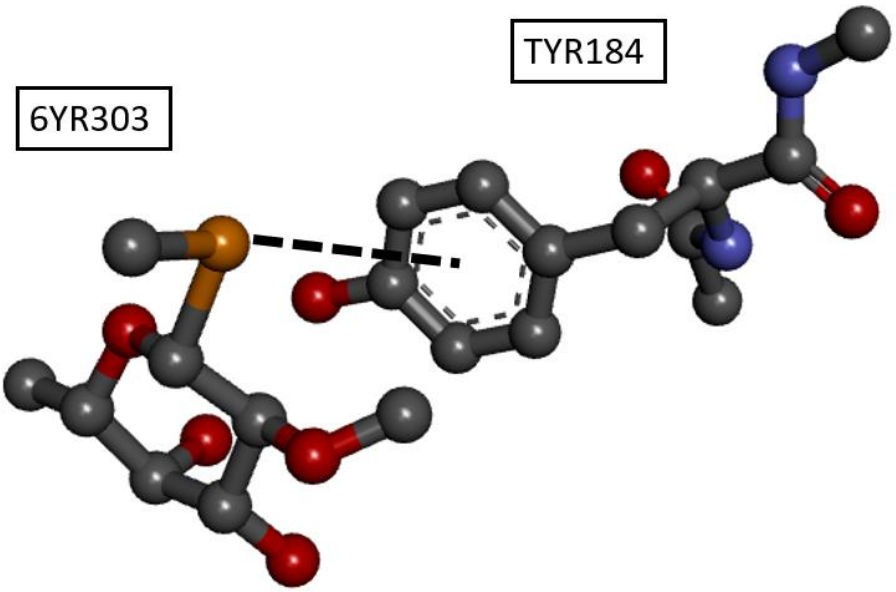

Figura 4. Modelo teórico utilizado para evaluar la interacción de enlace de calcógeno en la estructura 5FSB.

Los cálculos teóricos se realizaron mediante el programa Gaussian 09.[27] Este programa es un recurso muy útil para analizar la estructura electrónica de moléculas. El programa trabaja procesando un fichero de entrada (input) que se trata de un archivo de texto. Al terminar los cálculos, el programa devuelve un fichero de salida (output) con los resultados obtenidos. Para facilitar la generación del archivo de entrada, se trabaja con el programa de visualización molecular Gaussview.[12]

De esta manera, el archivo de entrada de Gaussian09W se divide en:

- Link0, donde están especificados parámetros como la ruta del fichero, memoria a utilizar, el número de procesadores. La información empieza con el símbolo %.
- El tipo de trabajo que se va a ejecutar, método teórico y bases a utilizar. Se especifican con palabras clave. Empieza con un signo #.
- Título que describirá el trabajo.

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

- Especificaciones, donde está la información de la carga y multiplicidad de la molécula y seguidamente las coordenadas de los átomos.

Con esto, el fichero de entrada tiene el siguiente aspecto (Figura 5):

```
%nprocshared=5
%mem=1024MB
%chk=5eo8_TFU405_GLU201_mod.chk
# opt b3lyp/6-31+g geom=connectivity

Title Card Required

-1 1
 C                -1   25.72900000   26.53100000    2.77900000
 C                -1   25.70400000   25.60100000    1.58200000
 O                -1   26.09900000   24.44300000    1.69200000
```

Figura 5. Ejemplo de fichero de input de Gaussian09W.

Antes de la obtención de las energías de interacción de enlace de calcógeno, se realizó la optimización de los átomos de H sobre los modelos teóricos creados. Para ello se utilizó método B3LYP[28] con base 6-31G*[29]. En el caso del átomo de Te la base 6-31G* no estaba disponible, por lo que se optó por especificar la base LanL2DZ[30] para éste y para el resto de los átomos, 6-31G*. A la hora de especificar que la optimización ha de ser únicamente para los átomos de hidrógeno, en el archivo de entrada se puso un "-1" entre el símbolo del átomo y las coordenadas a todos los átomos que no fueran hidrógeno, y un 0 a todos los hidrógenos, entre el símbolo y las coordenadas, resultando algo similar a la siguiente figura:

```
C                -1  130.94500000    1.42800000  150.79600000
O                -1  129.27100000   -3.02700000  148.89100000
O                -1  128.13400000    0.37400000  149.71400000
O                -1  126.71900000   -1.68000000  148.52300000
H                 0  134.79288126   -3.63847636  148.44950596
H                 0  132.21149509   -2.72463855  144.64598475
H                 0  129.55254367   -3.68439887  144.77475631
```

Figura 6. Ejemplo de optimización de átomos de hidrógeno.

Para la obtención de la energía de interacción (al nivel B3LYP/6-311+G**),[29] se calculó la energía total del complejo ($E_{complejo}$), la energía de ligando ($E_{ligando}$) y la energía del aminoácido ($E_{a.a.}$) y se obtendrá por diferencia:

$$E_{int} = E_{complejo} - E_{ligando} - E_{a.a} \quad (1)$$

[28] K. Kim, K. D. Jordan, *J. Phys. Chem. A,* **1994**, *98*, 10089–10094.
[29] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, L. A. Curtis, *J. Comp. Chem.*, **2001**, *22*, 976-984.
[30] S. Chiodo, N. Russo, E. Sicilia, J. Chem. Phys., **2006,** 125, 104107.

Las diferentes energías recogidas en la ecuación 1 se obtuvieron realizando cálculos puntuales de energía (single point) sobre las geometrías que contenían los átomos de H optimizados. En este sentido, el método B3LYP se ha utilizado de manera exitosa para el estudio de interacciones σ-hole con especies neutras y cargadas dadoras de electrones.[31]

El cálculo de las superficies de potencial electrostático (MEP) se realizó al nivel B3LYP/6-311+G** para cada uno de los ligandos seleccionados de Se/Te mediante el programa Gaussian09.[27]

Seguidamente se realizó en análisis de puntos críticos AIM,[22] que se trata de un valor de una propiedad física de un sistema completo (generalmente moléculas) obtenido mediante la suma del valor de esta propiedad para todos los átomos del sistema. Suponiendo un sistema perfecto, este estudio permite observar cómo los cambios en átomos y grupos atómicos afectan a las propiedades moleculares. Para poder realizar los análisis AIM se utilizó el programa AIMall.[32]

Posteriormente se llevó a cabo el análisis NCIplot,[24] identificando la presencia y la naturaleza favorable o desfavorable de las interacciones atractivas no covalentes presentes. Este análisis se basa en un esquema de color compuesto por una escala rojo-amarillo-verde-azul, siendo rojo para interacciones repulsivas y azul para interacciones fuertes de carácter atractivo. Las interacciones repulsivas y atractivas de carácter débil se colorean en amarillo y verde respectivamente.

Finalmente se realizaron los análisis NBO,[23] el cual es una técnica útil para el estudio de las interacciones orbitales responsables de la formación de enlaces covalentes/no covalentes.[33]. Esto nos dará información de los tipos de orbitales dadores y aceptores de electrones implicados en las interacciones entre el ligando y el aminoácido. Para ello se utilizó también el programa Gaussian09.

Problemas encontrados y soluciones propuestas

Entre los problemas encontrados a la hora de realizar este estudio, se han tenido los siguientes problemas:

- Errores en cálculos de NBO: Lnk1e in C:\G09W\l607.exe
- Errores de optimización en ligandos que poseen Te en su estructura cuando se pretendían optimizar los átomos de hidrógeno.

[31] A. Bauzá, I. Alkorta, A. Frontera, J. Elguero, *J. Chem. Theory Comput.*, **2013**, *9*, 5201-5210.
[32] AIMAll (Version 19.10.12), Todd A. Keith, TK Gristmill Software, Overland Park KS, USA, 2019.
[33] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.,* **1988**, *88*, 899–926.

El error de cálculo NBO no se pudo solventar, por lo que se descartó la estructura. El error de optimización de ligando de Te se pudo solventar añadiendo la keyword GENECP para poder aplicar dos bases diferentes de manera específica.

4. Discusión de resultados.

Estructuras seleccionadas de Se

Entre las estructuras recogidas en la Tabla 1, 3USP, 5FSB, 1O9V y 5EO8 contienen Se en la estructura del ligando. En esta sección se discutirá en profundidad los resultados correspondientes a las estructuras 1O9V, 3USP y 5FSB. Los resultados energéticos sobre la estructura 5EO8 se recogen en la Tabla 2 (ver abajo).

Análisis MEP

Comenzamos el estudio computacional realizando un análisis molecular de la superficie de potencial electrostático (MEP) de ligandos de selenio presentes en las estructuras 1O9V, 3USP y 5FSB. Como podemos ver en la figura 7, en todos los casos se obtuvo un valor de potencial electrostático positivo sobre el átomo de Se, más concretamente en la prolongación del enlace C-Se, denotando la presencia de σ-hole, lo que los hace adecuados para establecer interacciones favorables con especies ricas en electrones. De esta manera, la estructura 1O9V presenta un valor de +10.66 kcal/mol, la estructura 3USP un valor de +6.26 kcal/mol y, finalmente, la estructura 5FSB, un valor de +13.80 kcal/mol, presentando la primera un valor más positivo de σ-hole que las otras dos. Además, los átomos de hidrógeno de los grupos OH presentes en los tres compuestos exhiben valores positivos en el MEP, lo que confiere versatilidad a estos restos para interactuar con grupos ricos en electrones mediante una combinación enlace de hidrógeno/calcógeno.

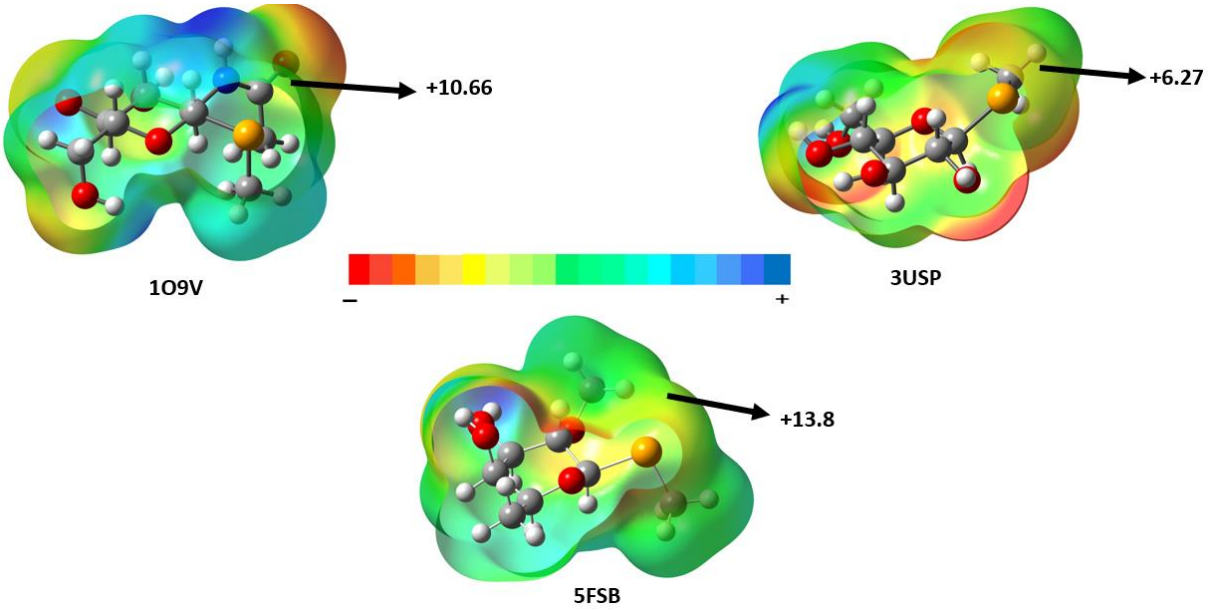

Figura 7. Superficies de potencial electrostático (MEP) correspondientes a los ligandos presentes en las estructuras 1O9V, 3USP y 5FSB (SNG1177, 05L610 Y 6YR303 respectivamente). Los valores de energía en los puntos seleccionados de la superficie (0.001 u.a) se muestran en kcal/mol.

Estructura 1O9V

Concretamente, 1O9V involucra a una proteína de tipo lectina, las cuáles se hallan de manera ubicua en la naturaleza (presentes en plantas, virus y animales) y juegan un papel importante en procesos de reconocimiento molecular de carbohidratos[34]. La estructura 1O9V (Figura 8a) representa un complejo proteína-ligando entre la lectina F17a-G y N-acetil-D-glucosamina (GlcNAc). La enzima F17a-G juega un papel fundamental en procesos de adhesión bacteriana a las microvellosidades del intestino, provocando diarreas o septicemia en rumiantes[35]. Debido a las ya conocidas propiedades inhibidoras de GlcNAc[36], Buts y sus colaboradores[37] determinaron la estructura cristalina de la adhesina F17-G en presencia y ausencia de GlcNAc y evaluaron las afinidades de monosacáridos sobre dos variantes de F17a-G.

Durante su estudio, los autores utilizaron derivados de Me-Se (ver Figura 8b) para así poder elucidar la estructura cristalina de la lectina F17a-G mediante la técnica MAD. Curiosamente, este selenocompuesto exhibió afinidad superior en comparación con un monosacárido ordinario. Una explicación plausible estaría relacionada con la estabilización adicional del ligando de selenio a través de un enlace de calcógeno entre el grupo Se-metilo y el grupo OH de SER117.

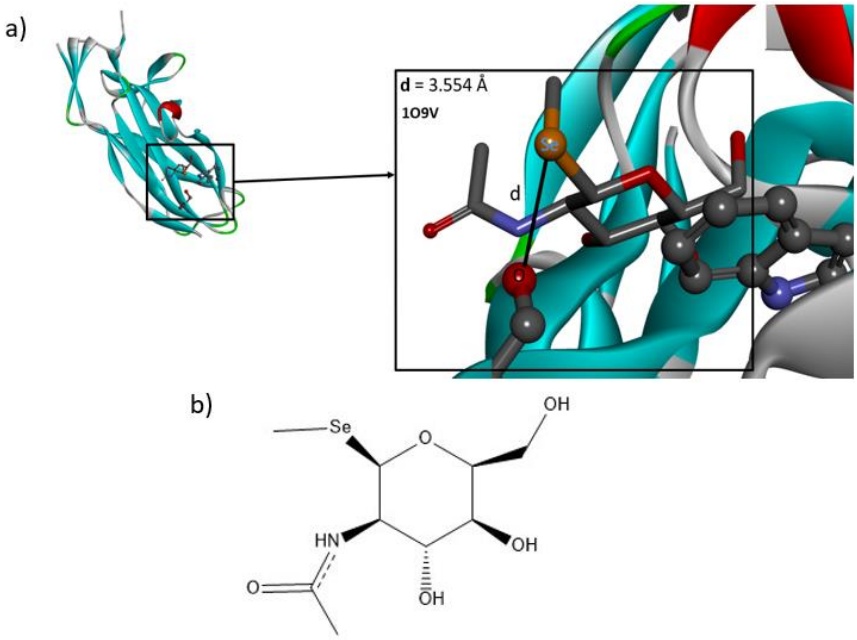

Figura 8 (a) Vista parcial de la estructura de rayos-X 1O9V. La interacción de enlace de calcógeno se halla magnificada en la parte derecha de la imagen. (b) Compuesto de Se utilizado como ligando.

[34] H. Lis, N. Sharon, Prot. Nuc. Acids, **1981**, 371-447.
[35] P. F. Lintermans, P. Pohl, A. Bertels, G. Charlier, J. Vandekerckhove, J. Van Damme, J. Schoup, C. Schlicker, T. Korhonen, H. De Greve, *Am. J. Vet. Res.*, **1988**, *49*, 1794-1799.
[36] Y. Bertin, J. -P. Girardeau, A. Darfeuille-Michaud, M. Contrepois, *Infect. Immun.*, **1996**, *64*, 332-342.
[37] L. Buts, J. Bouckaert, E. De Genst, R. Loris, S. Oscarson, M. Lahmann, H. De Greve, *Mol. Microbiol.*, **2004**, *49*, 705–715.

Estructura 3USP

El complejo proteína-ligando correspondiente a la estructura 3USP[38] involucra un neurotransmisor de sodio simporte (NSS). La función biológica de un NSS consiste en la eliminación de glicina, ácido γ-aminobutírico, serotonina, norepinefrina y dopamina liberada sinápticamente, para completar el proceso de transmisión en el sistema nervioso central[39]. El mal funcionamiento de los NSSs se relaciona con varias enfermedades del sistema nervioso, como depresión, epilepsia, intolerancia ortostática, ansiedad y Parkinson[40]. En el estudio de Wang y sus colaboradores[37] se llevó a cabo la determinación estructural de LeuT (un NSS procariótico de *Aquifex aeolicus*[41]). Para lograrlo, los autores cristalizaron LeuT en un entorno similar a una membrana lipídica para que, de esta manera, pudieran estudiar sus propiedades de conformación y unión al sustrato en unas condiciones biológicas fiables. Mediante la utilización del compuesto n-heptilo seleno-β-d-glucósido, se pudieron mapear los distintos lugares de unión del sustrato. Interesantemente, solo se identificó un único lugar de unión para dicho compuesto de Se, en el cual se estableció una interacción de enlace de calcógeno entre el grupo Se del ligando y un grupo carboxilato perteneciente al residuo ASP401, lo cual contribuyo a la unión y estabilización del ligando.

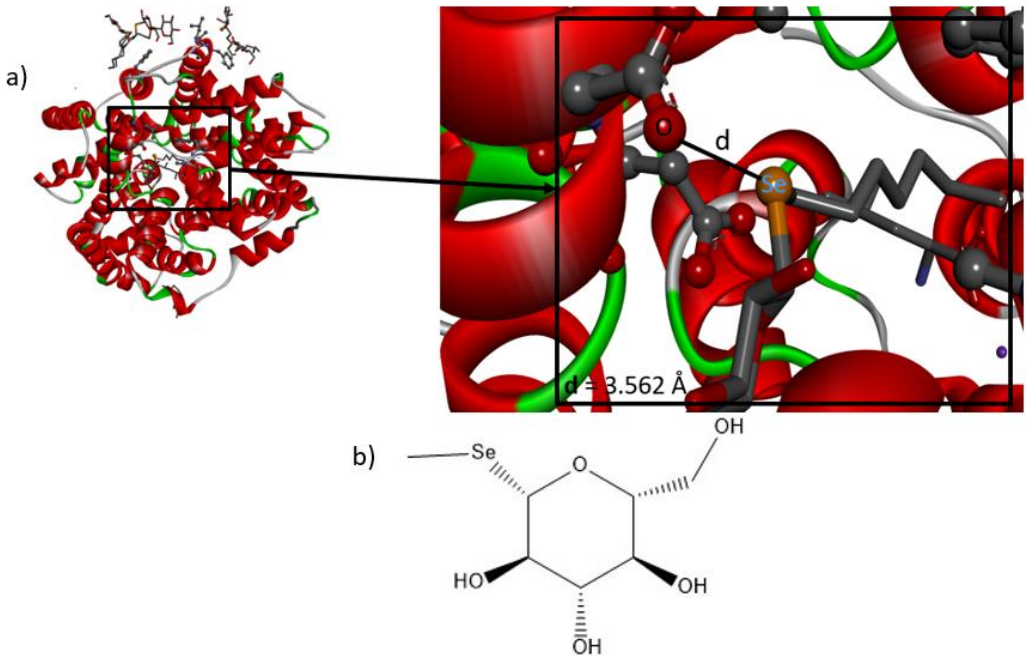

Figura 9. (a) Vista parcial de la estructura de rayos-X 3USP. La interacción de enlace de calcógeno se halla magnificada en la parte derecha de la imagen. (b) Compuesto de Se utilizado como ligando

[38] H. Wang, J. Elferich, E. Gouaux, *Nat. Struct. Mol. Biol.*, **2012**, *19*, 212-219.
[39] F. Ashcroft, D. Gadsby, C. Miller, *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **2009**, *364*, 145–147.
[40] M. Hahn, R. Blakely, *Pharmacogenomics J.*, **2002**, *2*, 217-235.
[41] A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, E. Gouaux, *Nature*, **2005**, *437*, 215-223.

Estructura 5FSB

Finalmente, la estructura 5FSB[42], involucra a las familias de proteínas tectoninas, implicadas en los mecanismos de defensa antimicrobianos. En el estudio, Sommer y colaboradores,[41] estudiaron el reconocimiento de glicanos o-metilados por las tectoninas, analizado mediante una combinación de experimentos estructurales y biofísicos de la proteína Lb-Tect2 complejada con dos ligandos diferentes. Los autores utilizaron 2-o-metil-seleno-L-fucopiranosido (Figura 10a) para resolver la estructura de rayos-X de la proteína, utilizando para ello la técnica MAD. Como se puede observar en la Figura 10a, en dicho sistema se establece una interacción calcógeno–π entre el ligando de Se y el anillo aromático perteneciente al residuo TYR184, hecho que pasó desapercibido para los autores originales. La interacción de tipo calcógeno–π (Ch-π) simplemente se diferencia en el hecho de que el donador de electrones implica un sistema π, ya sea aislado (por ejemplo, eteno) o conjugado (benceno).[10]

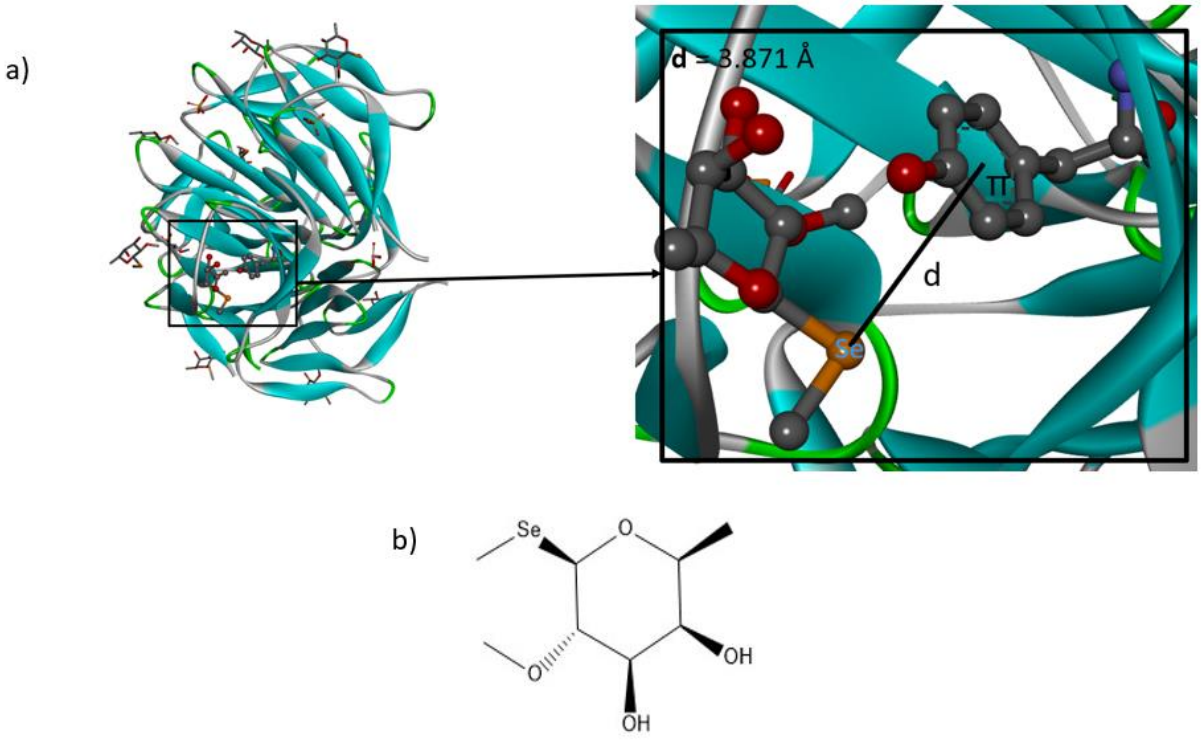

Figura 10 (a) Vista parcial de la estructura de rayos-X 5FSB. La interacción de enlace de calcógeno se halla magnificada en la parte derecha de la imagen. (b) Compuesto de Se utilizado como ligando.

[42] R. Sommer, O. N. Makshakova, T. Wohlschlager, S. Hutin, M. Marsh, A. Titz, M. Künzler, A. Varrot, *Structure*, **2018**, *26*, 391-402.

Estructura seleccionada de Te.

Análisis MEP

Comenzamos el estudio computacional realizando un análisis molecular de la superficie de potencial electrostático (MEP) del ligando de Te en 1EL7. Como podemos ver en la figura 11, se obtiene un valor de potencial electrostático positivo sobre el átomo de Te, más concretamente en la prolongación del enlace C-Te, denotando la presencia de σ-hole, lo que lo hace adecuado para establecer interacciones favorables con especies ricas en electrones.

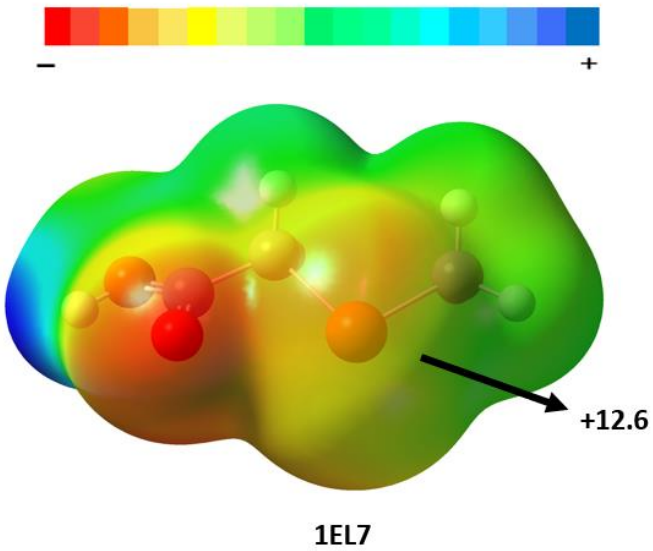

Figura 11. Superficies de potencial electrostático (MEP) correspondientes al ligando de la estructura 1EL7 (MTD511). El valore de energía en el puntos seleccionado de la superficie (0.001 u.a) se muestra en kcal/mol.

Estructura 1EL7

El complejo proteína-ligando 1EL7[43] corresponde a una sarcosina oxidasa monomérica. Se trata de una flavoenzima bacteriana cuya función principal es catalizar la desmetilación oxidativa de sarcosina (N-metilglicina), produciendo glicina, formaldehído y peróxido de hidrógeno como productos de reacción. La sarcosina es un metabolito común del suelo que actúa como única fuente de carbono y energía para muchos microorganismos. Estudios han demostrado la presencia de una molécula de flavín adenín dinucleótido (FAD) unido covalentemente al residuo CYS315 cerca del extremo COOH terminal de la proteína. El sustrato carboxilato es esencial para la unión, ya que aminas simples, como la etilmetilamina, no perturba la absorción del espectro de MSOX (Monomeric Sarcosine Oxidase) ni inhibe su actividad. Se reemplazó el grupo amino por $CH_3XCH_2CO_2^-$ (X= CH_2, O, S, Se o Te) y se observó que la constante de disociación cuando X=Te era tan solo dos veces mayor que la estimada para la sarcosina. De esta manera es de esperar que aparezca una interacción tipo ChB entre el ligando y el aminoácido GLY344 (Figura 12a).

[43] P. Trickey, M. A. Wagner, M. S. Jorns, F. S. Mathews, *Structure*, **1999**, *7*, 331-45.

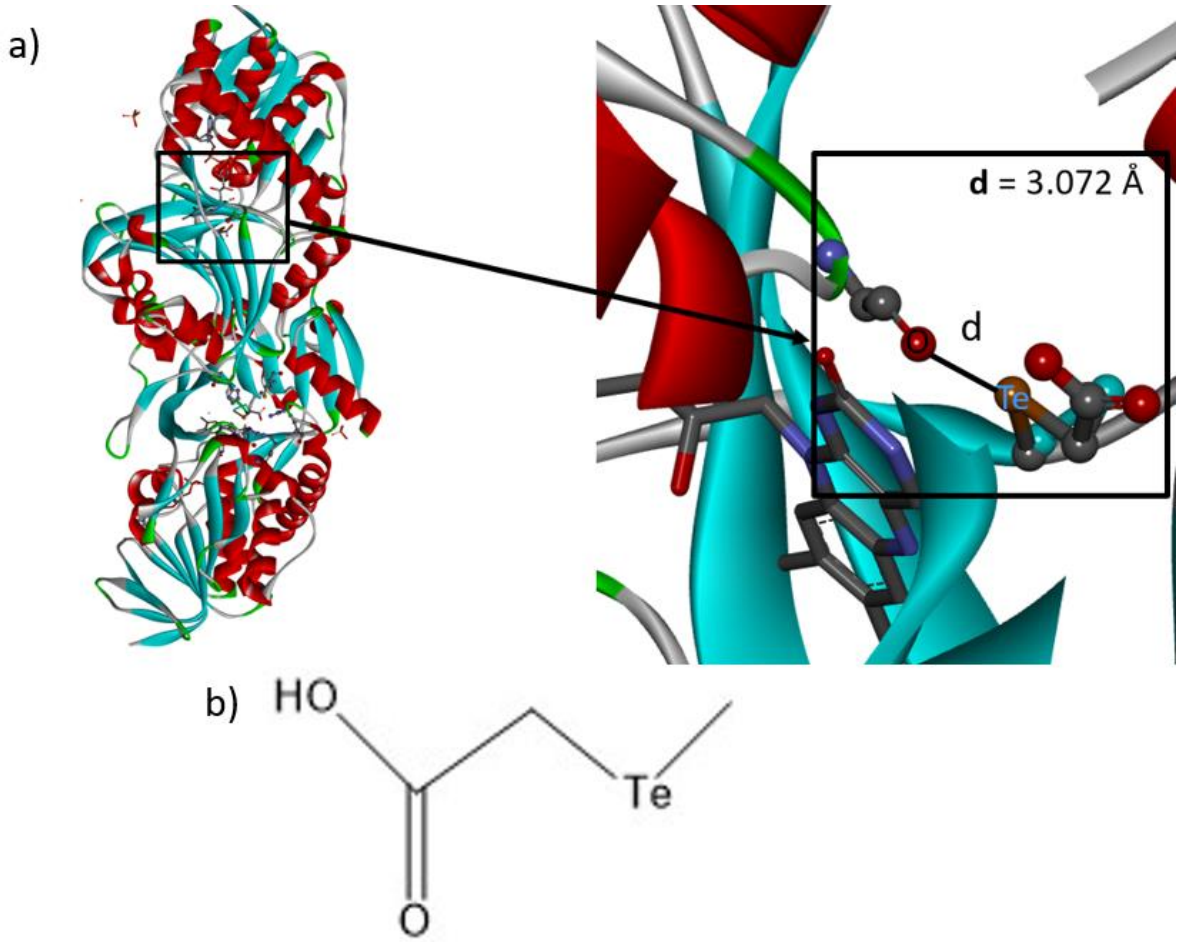

Figura 12. (a) Vista parcial de la estructura de rayos-X 1EL7. La interacción de enlace de calcógeno se halla magnificada en la parte derecha de la imagen. (b) Compuesto de Se utilizado como ligando.

Estudio energético

Las energías de interacción, distancias Ch···A (A = O, C) y los ángulos de C-Ch····A (A= O, C) se ven recogidos en la tabla 2. Como se puede observar, en la mayoría de ellos aparece una interacción energética favorable entre –0.9 y –8.2 kcal/mol. Además, el complejo 5FSB, que involucra TYR como donante de electrones obtuvo la energía de enlace más favorable del estudio (–8.2 kcal/mol). Por otro lado, el complejo 5EO8 obtuvo la energía de interacción más débil del estudio (–0.9 kcal/mol). También cabe resaltar que en los complejos estudiados también participan interacciones adicionales, como el enlace de hidrógeno o las interacciones CH–HC y CH–π, las cuáles también contribuyen a la estabilidad del complejo proteína-ligando. Dichas interacciones serán analizadas con más detalle en los análisis AIM y NCIplot (ver abajo). Finalmente, los ángulos C-Ch···A oscilan entre 150.7 y 169.6 grados, por lo que

exhiben cierta direccionalidad, como se ha observado en los estudios cristalográficos previos[44].

Tabla 2. Lista de estructuras seleccionadas (PDBID), concretando el código del ligando y del aminoácido implicados en la interacción de enlace de calcógeno. El valor de la distancia (d, en Å), el ángulo (θ) C–Ch···A, en º y la energía de interacción (ΔE) también se muestran.

PDB ID	ID ligando	ID Aminoácido	d	θ	ΔE
3USP	05L610	ASP401	3.562	153.4	-7.5
5FSB	6YR303	TYR154	3.871	148.5	-8.2
1O9V	SNG1177	SER117	3.554	150.7	-4.6
5EO8	TFU405	ASP193	3.851	134.3[a]	-0.9
1EL7	MTD801	GLY344	3.072	169.6	-1.1

Análisis AIM y NCIplot

Las interacciones de enlace de calcógeno presentes en los complejos de Se/Te correspondientes a las estructuras 1O9V, 3USP, 5FSB y 1EL7 han sido caracterizadas mediante un análisis de AIM y un NCIplot.

La distribución de los puntos críticos de enlace (BCPs) intermoleculares para los cuatro complejos proteína-ligando se muestran en la figura 13. En todos los casos el ChB se caracteriza por la presencia de BCP que conecta tanto al átomo de Se/Te del ligando con átomos de O pertenecientes a residuos de los aminoácidos. Además, interacciones de tipo enlace de hidrógeno (HB), CH···HC y CH-π también participan en el mecanismo de reconocimiento entre ligando y aminoácido. Por ejemplo, en 1O9V dos BCPs conectan i) dos grupos CH y ii) un resto OH del aminoácido SER117 con un grupo CH del ligando de Se, formando así una interacción CH···HC y un HB.

[44] M. Fourmigué, A. Dhaka, *Coord. Chem. Rev.*, **2020**, *403*, 213084.

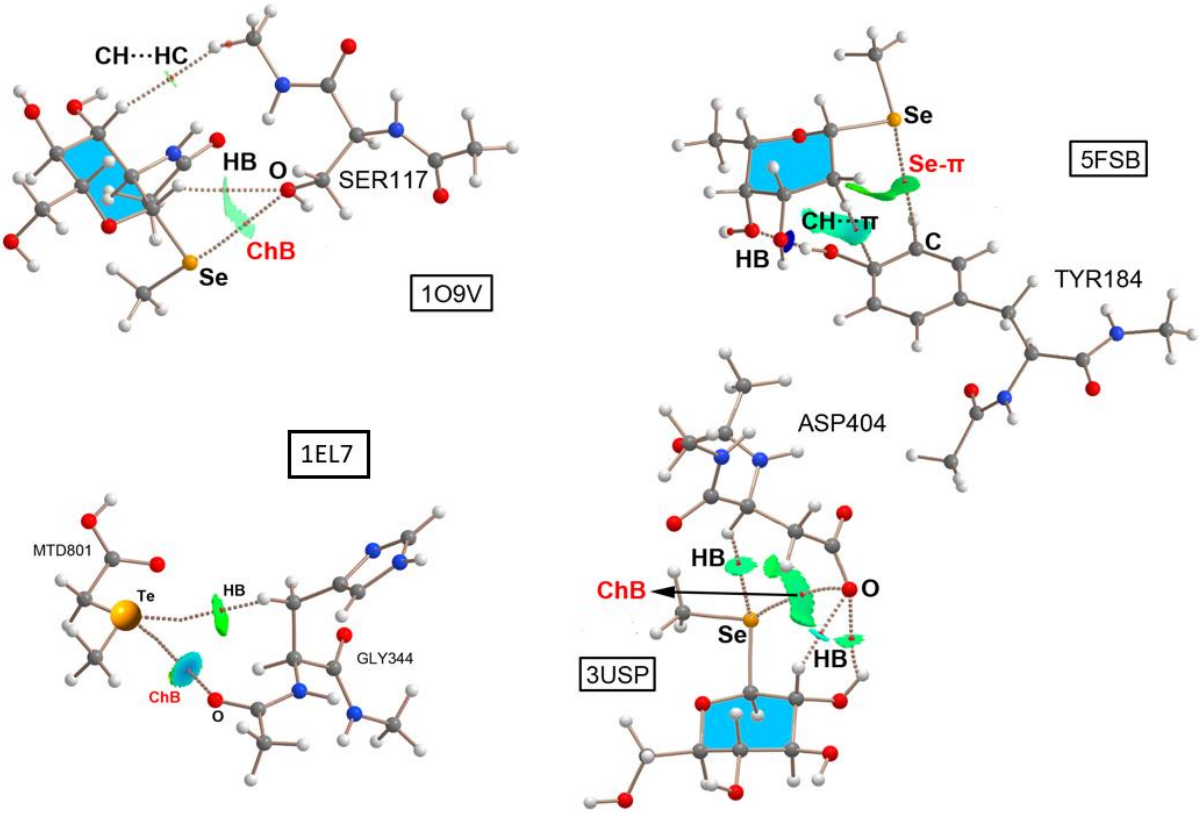

Figura 13. Distribución de puntos críticos (en rojo) en los complejos 19OV, 5FSB, 1EL7 y 3USP. Los BCP correspondientes a interacciones HB, CH–HC y CH–π también se hallan indicados. También se muestran las superficies NCIplot entre los ligandos y los aminoácidos.

En la estructura 3USP se establecen HBs auxiliares que implica al grupo carboxilato de ASP404 con los grupos OH y CH del ligando. El átomo de Se también se ve involucrado en un HB, conectándose a un grupo CH del aminoácido. Por otro lado, en 5FSB, participan dos interacciones auxiliares en la unión del ligando de Se al aminoácido. Aparece un HB entre el grupo OH de TYR184 y un OH del ligando y una interacción CH-π que involucra un CH del ligando de Se al sistema aromático de TYR184.

Finalmente, en el caso de 1EL7 encontramos un HB auxiliar que implica a un grupo CH del aminoácido con el Te y, el grupo carbonilo de la amida interacciona con el Te mediante enlace de calcógeno.

Para finalizar, en la figura 13 también podemos ver una isosuperfície verdosa ubicada entre el átomo de Se y Te y la zona rica en electrones de átomos de O o C, lo que confirma la existencia y naturaleza favorable de los enlaces ChB y Ch-π. Además, también se confirma la presencia de las interacciones auxiliares que también están implicadas en el reconocimiento de los ligandos de Se y Te.

Análisis NBO

Finalmente, se ha estudiado las implicaciones de las contribuciones orbitálicas en las interacciones de enlace de calcógeno descritas. Para ello, se han realizado cálculos NBO centrándonos en el análisis de perturbación de segundo orden que es útil para evaluar las interacciones dador-aceptor (Tabla 3). Este análisis revela una contribución que proviene de una donación de un par de electrones solitario (LP) perteneciente al aminoácido a un orbital antienlazante (BD*) C-Ch en los complejos proteína-ligando correspondientes a las estructuras 1O9V, 5EO8 y 1EL7, comprendido entre 0.1 y 2.4 kcal/mol. Por otro lado, en las estructuras 3USP y 5FSB se observó una interacción orbitálica entre un orbital enlazante (BD) de tipo π correspondiente a un enlace C–O y C–C y un orbital antienlazante (BD*) C–Se, respectivamente.

Tabla 2. Resultados del análisis NBO. LP, BD y BD* corresponden a par solitario, orbital enlazante y orbital antienlazante, respectivamente.

PDBID	Donador	Aceptor	$E^{(2)}$ (kcal/mol)
3USP	BDC–O	BD*Se–C	0.2
5FSB	BDC–C	BD*Se–C	0.1
1O9V	LP O	BD*Se–C	0.2
5EO8	LP O	BD*Se–C	0.1
1EL7	LP O	BD*Te-C	2.4

5. Conclusiones

Los resultados obtenidos en este trabajo enfatizan la capacidad de ligandos de selenio para formar enlaces de calcógeno en aminoácidos como ASP, ASN, SER, TYR, con fuerzas de interacción entre -0.91 y -8.2 kcal/mol. Los enlaces ChBs y Ch-π así como las interacciones adicionales HB, CH···HC y CH-π descritos con anterioridad se caracterizaron adicionalmente utilizando metodología AIM. La presencia y naturaleza favorable de las interacciones débiles se confirman mediante índice de visualización NCIplot. Por otro lado, el análisis NBO revela contribuciones orbitálicas de los LP de átomos de O, o del sistema π de un doble enlace C=O o C=C del aminoácido con un orbital antienlazante C-Se del ligando, lo que ayuda a la formación de complejos no covalentes y demuestra la naturaleza σ-hole de la interacción.

Estos resultados son útiles para ampliar los conocimientos sobre los ChB biológicos, así como para demostrar que, para la determinación mediante difracción de rayos X, la adición de ligandos con selenio en su estructura, puede generar interacciones que estabilicen aún más la proteína, por lo que se debe tener en cuenta a la hora de conocer la estructura real de la misma. Por tanto, es importante estudiar todas las posibles interacciones que puedan existir entre diferentes agregados moleculares. Así se podrá comprobar si, al añadir nuevas moléculas a la macroestructura, ésta se ve estabilizada o desestabilizada, dando lugar a nuevas aplicaciones dentro del campo de la Química Biológica.

6. Publicación

Please do not adjust margins

Journal Name

ARTICLE

Selenium Chalcogen bonds are involved in protein-carbohydrate recognition: A combined PDB and theoretical study.

Received 00th January 20xx,

Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Jose A. Fernández Riveras,[a] Antonio Frontera[a] and Antonio Bauzá[a,*]

In this manuscript the ability of Selenium carbohydrates to undergo Chalcogen bonding (ChB) interactions with protein residues has been studied at the RI-MP2/def2-TZVP level of theory. An inspection of the Protein Data Bank (PDB) revealed Se···A (A = O, C and S) intermolecular contacts involving Se-pyranose ligands and ASP, TYR, SER and MET residues. Theoretical models were built to analyse the strength and directionality of the interaction together with "Atoms in Molecules" (AIM), Natural Bonding Orbital (NBO) and Non Covalent Interactions plot (NCIplot) analyses, which further assisted in the characterization of the ChBs described herein. We expect the results from this study will be useful to expand the current knowledge regarding biological ChBs as well as to increase the visibility of the interaction among the carbohydrate chemistry community.

Introduction

The myriad of potential applications involving noncovalent interactions (NCIs) makes them a crucial tool for scientitts devoted to the field of supramolecular chemistry.[1] From the beginning of the century, NCIs have earned a general recognition that has moved the field far beyond classical and ubiquitous forces such as hydrogen bonding and π–π stacking interactions. Concretely, the σ-hole crown of interactions,[2] composed by halogen, chalcogen, pnictogen and tetrel bonds, has become a reliable and robust resource for driving molecular recognition and self-assembly processes. In this regard, the theoretical analysis[3] of NCIs is fundamental for achieving a full exploitation in several research areas (e.g. materials science, catalysis, biochemistry, etc).[4] Particularly, Chalcogen bonds (ChBs) have been contextualized in areas such as catalysis,[5] materials science,[6] main group chemistry reactivity[7] and molecular recognition[8], among others. In biology, although the interaction has been studied in the context of protein-ligand chemistry,[9] most studies focus on S as σ-hole donor, due to its higher abundacy compared to heavier chalcogens such as Se and Te. In this regard, the selenium labelling of sugars for the phasing of carbohydrate binding proteins was firstly proposed by Loris and coworkers to solve the structure of the F17-G fibrial adhesin.[10] Nowadays, the use of Se to determine the X-ray crystal structure of proteins[11] and nucleic acids[12] through the multi/single-wavelength anomalous diffraction (MAD/SAD) phasing techniques has left some unanswered questions regarding the noncovalent interactions established between these Se-derivatives and their surroundings. Bearing this in mind, our curiosity was directed to protein-ligand complexes involving Se-carbohydrate ligands. Although the number is scarce in the Protein Data Bank (PDB), some of them are involved in the formation of protein-ligand complexes estabilized by ChB interactions. Hence, their investigation might be useful for scientists working in the field of carbohydrate recognition and protein-ligand crystallography as well as to further expand the current biological picture of Se. In this manuscript our aim is to investigate by means of *ab initio* calculations the ability Selenium carbohydrates to undergo ChBs using several Se-pyranose ligands as σ-hole donors and SER, ASP, TYR and MET residues as electron rich species (see Fig. 1). Calculations at the RI-MP2/def2-TZVP level of theory indicate that the interaction is favourable and ranges from weak to moderately strong nature. Furthermore, AIM (Atoms in Molecules), NBO (Natural Bonding Orbital) and NCIplot (Non Covalent Interactions plot) analyses further characterized the noncovalent forces described herein and gave further insights on the orbital interactions accounting for the stabilization of the complexes studied.

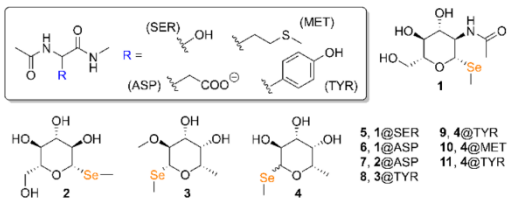

Fig. 1 Compounds **1** to **4**, aminoacids used as electron donor moieties and complexes **5** to **11** used in this study.

[a] *Department of Chemistry, Universitat de les Illes Balears, Crta. de Valldemossa km 7.5, 07122 Palma (Baleares), Spain. Fax: +34 971 173426*

Electronic Supplementary Information (ESI) available: Cartesian coordinates of all complexes. See DOI: 10.1039/x0xx00000x

Please do not adjust margins