



Universitat Autònoma de Barcelona

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  [http://cat.creativecommons.org/?page\\_id=184](http://cat.creativecommons.org/?page_id=184)

**ADVERTENCIA.** El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

**WARNING.** The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>





# **Estudios Teóricos y Computacionales para la Síntesis Enzimática del Enlace Glicosídico**

**María Fernanda Mendoza Muñoz**

**Tesis de Doctorado**

Programa de Doctorado en Química

Supervisores

**Laura Masgrau i Fontanet**

**Josep M. Lluch i López**

Departament de Química

Facultat de Ciències

2016

Memoria presentada para aspirar al grado de Doctora por

***María Fernanda Mendoza Muñoz***

Visto y aprobado por:



Laura Masgrau Fontanet

Directora



Josep Maria Lluch

Director

Bellaterra, 19 de octubre 2016.



***“...la tarea no consiste tanto en ver lo que nadie ha visto aún, sino en pensar lo que nadie ha pensado aún sobre lo que todo el mundo ve...”***

Arthur Schopenhauer

## Tabla de Contenidos

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introducción</b>   | <b>1</b>  |
| 1.1      | Los Glicanos y su Relevancia                                | 1         |
| 1.2      | Visión General de las Glicosil Hidrolasas (GHs)             | 3         |
| 1.2.1    | Clasificación   | 3         |
| 1.2.2    | Topología 3D y Plegamientos                                 | 5         |
| 1.2.3    | Mecanismos de Reacción de las GHs                           | 8         |
| 1.3      | Visión General de las Glicosiltransferasas (GTs)            | 11        |
| 1.3.1    | Clasificación   | 13        |
| 1.3.2    | Topología 3D y Plegamientos                                 | 15        |
| 1.3.3    | Mecanismos Catalíticos de las GTs                           | 19        |
| 1.4      | Estado del Debate en el Momento de Iniciar la Tesis         | 28        |
| <b>2</b> | <b>Objetivos</b>  | <b>34</b> |
| 2.1      | Objetivos Generales   | 35        |
| 2.2      | Objetivos Específicos                                       | 35        |
| <b>3</b> | <b>Fundamentos y Métodos Teóricos</b>                       | <b>36</b> |
| 3.1      | Mecánica Molecular  | 36        |
| 3.1.1    | Función de Energía Potencial                                | 37        |
| 3.1.2    | Análisis de los Términos de la Función de Energía Potencial | 38        |
| 3.2      | Métodos de la Química Cuántica                              | 40        |
| 3.2.1    | El Método Variacional                                       | 41        |
| 3.2.2    | El Método Hartree-Fock                                      | 42        |
| 3.3      | La Teoría de los Funcionales de la Densidad                 | 45        |
| 3.3.1    | El Método Kohn-Sham   | 47        |
| 3.3.2    | Aproximación de la Densidad Local o LDA                     | 49        |
| 3.3.3    | Aproximación GGA y meta GGA                                 | 50        |
| 3.3.4    | Hiper GGA o Métodos Híbridos                                | 51        |
| 3.4      | Métodos Semiempíricos                                       | 52        |
| 3.4.1    | Métodos Semiempíricos Hartree-Fock                          | 52        |
| 3.4.2    | Métodos Semiempíricos DFT                                   | 53        |
| 3.5      | Aproximación QM/MM  | 54        |
| 3.5.1    | Expresión de Energía QM/MM                                  | 55        |
| 3.5.2    | Esquemas de Interacción Electroestática                     | 56        |
| 3.5.3    | Tratamiento de Átomos Frontera                              | 57        |

|        |   |     |
|--------|---|-----|
| 3.5.4  | Otras Interacciones QM/MM no Enlazantes.....  | 58  |
| 3.6    | Dinámica Molecular .....  | 59  |
| 3.6.1  | Algoritmo de la MD .....  | 59  |
| 3.6.2  | Restricciones y Paso de Integración.....  | 60  |
| 3.6.3  | Condiciones Periódicas de Contorno y Radios de Corte .....                                      | 60  |
| 3.6.4  | Condiciones Experimentales y Colectivos Termodinámicos .....                                    | 61  |
| 3.6.5  | Potencial de Fuerza Media.....  | 61  |
| 3.6.6  | Método Umbrella Sampling (US).....  | 62  |
| 3.7    | Metodología de Docking .....  | 63  |
| 3.7.1  | Métodos de Búsqueda y Flexibilidad Molecular .....  | 63  |
| 3.7.2  | Funciones de Puntuación.....  | 64  |
| 4      | Glucosil-3-fosfoglicerato Sintasa: GpgS.....  | 65  |
| 4.1    | Aspectos Generales .....  | 65  |
| 4.1.1  | Datos Cristalográficos de GpgS: Conformaciones de PGA y Posibles Aguas en el Sitio Activo.....  | 68  |
| 4.2    | Modelos y Metodología .....   | 72  |
| 4.2.1  | Construcción del Modelo Inicial .....   | 72  |
| 4.2.2  | Minimizaciones y Simulaciones de Dinámica Molecular: Obtención del Modelo 0 .....               | 75  |
| 4.2.3  | Simulaciones QM/MM del Mecanismo de Ataque Frontal .....  | 79  |
| 4.2.4  | Cálculos QM/MM para la Hidrólisis de UDP-Glc por la $W_2$ .....                                 | 81  |
| 4.2.5  | Refinamiento del Modelo: Obtención del Modelo 1.....  | 81  |
| 4.2.6  | Modelo 2: Modelo sin Aguas en el Centro Activo.....   | 83  |
| 4.2.7  | Simulaciones Adicionales del Mecanismo de Reacción: Híbrido 1.A e Híbrido 1.B.....              | 84  |
| 4.2.8  | Cálculo de Contribuciones Electroestáticas .....  | 84  |
| 4.2.9  | Cálculo de Cargas Naturales e Interacciones de Orbitales Naturales .....                        | 85  |
| 4.2.10 | Simulación de Umbrella Sampling para el Mecanismo de Ataque Frontal                             | 85  |
| 4.3    | Resultados y Discusión .....  | 87  |
| 4.3.1  | Modelo 1 o Complejo Catalítico: Mecanismo de Ataque Frontal.....                                | 90  |
| 4.3.2  | Modelo 2 o Complejo Catalítico sin Aguas en el Centro Activo: Mecanismo de Ataque Frontal ..... | 102 |
| 4.3.3  | Híbrido 1.A: Mecanismo de Ataque Frontal.....   | 108 |
| 4.3.4  | Híbrido 1.B: Mecanismo de Ataque Frontal.....   | 112 |
| 4.3.5  | Modelo 1, Híbrido 1.A y 1.B: Mecanismo de Ataque del agua $W_2$ o Hidrólisis de UDP-Glc .....   | 117 |



|       |   |     |
|-------|---|-----|
| 4.4   | Conclusiones.....   | 119 |
| 5     | $\alpha$ 1,4-N-Acetilhexosaminiltransferasa: EXTL2-GalNAc.....  | 121 |
| 5.1   | Aspectos Generales.....   | 121 |
| 5.1.1 | EXTL2 y Estructuras Cristalográficas.....   | 124 |
| 5.2   | Modelos y Metodología.....  | 125 |
| 5.2.1 | Construcción del Modelo Inicial para la Transferencia de GalNAc.....                                    | 125 |
| 5.2.2 | Minimizaciones y Dinámica Molecular.....  | 127 |
| 5.2.3 | Simulaciones QM/MM.....   | 128 |
| 5.2.4 | Cálculo de Contribuciones Electrostáticas, Cargas Naturales e Interacción de Orbitales Naturales.....   | 132 |
| 5.2.5 | Cálculo de Energía Libre para el Mecanismo de Ataque Frontal.....                                       | 133 |
| 5.3   | Resultados y Discusión.....   | 134 |
| 5.3.1 | Mecanismo de Ataque Frontal.....  | 138 |
| 5.3.2 | Mecanismo de Ataque del Asp246: Formación del CGE.....  | 147 |
| 5.4   | Conclusiones.....   | 156 |
| 6     | $\alpha$ 1,4-N-Acetilhexoaminiltransferasa: EXTL2-GlcNAc.....   | 159 |
| 6.1   | Aspectos Generales.....   | 159 |
| 6.2   | Modelos y Metodología.....  | 160 |
| 6.2.1 | Construcción de los Modelos.....  | 160 |
| 6.2.2 | Cálculos de Docking.....  | 161 |
| 6.2.3 | Simulaciones de Dinámica Molecular.....   | 165 |
| 6.2.4 | Simulaciones QM/MM.....   | 166 |
| 6.2.5 | Cálculo de Contribuciones Electrostáticas, Cargas Naturales e Interacciones de Orbitales Naturales..... | 169 |
| 6.2.6 | Mutantes <i>in silico</i> .....   | 169 |
| 6.2.7 | Simulación de Umbrella Sampling.....  | 169 |
| 6.3   | Resultados y Discusión.....   | 170 |
| 6.3.1 | Mecanismo de Ataque Frontal.....  | 173 |
| 6.3.2 | Mecanismo de formación del CGE: Ataque Nucleofílico de Asp246.....                                      | 190 |
| 6.3.3 | Mutaciones <i>in silico</i> y Mecanismo de Reacción.....  | 196 |
| 6.3.4 | Simulaciones de Dinámica Molecular de las Enzimas Mutantes.....   | 204 |
| 6.4   | Conclusiones.....   | 211 |
| 7     | Comparaciones Adicionales y Discusión General.....  | 213 |
| 7.1   | Introducción.....   | 213 |
| 7.2   | Metodología y Modelos.....  | 215 |
| 7.3   | Resultados y Discusión.....   | 215 |

|                                       |     |
|---------------------------------------|-----|
| <b>7.4 Conclusiones</b> .....         | 224 |
| <b>8 Conclusiones Generales</b> ..... | 225 |
| REFERENCIAS.....                      | 228 |



# 1 Introducción

## 1.1 Los Glicanos y su Relevancia

Se les llama glicanos, carbohidratos, sacáridos o simplemente azúcares a todas las moléculas formadas por diferentes combinaciones de monosacáridos. Estas unidades de azúcar pueden combinarse para formar cadenas de distintas longitudes que van desde disacáridos, pasando por oligosacáridos (2-10 monómeros), hasta polisacáridos que exhiben una compleja estructura tridimensional. A diferencia de lo que ocurre en la síntesis del ADN o de las proteínas, no existe o no se ha identificado una plantilla en base a la cual los azúcares se ensamblen pero sí se sabe que la síntesis del enlace glicosídico es influenciada por factores como el metabolismo celular, el tipo de célula, el estado de desarrollo, y la disponibilidad de nutrientes.<sup>1</sup> Todos estos factores generan una gran diversidad de glicanos y esta misma característica los convierte en especies difíciles de estudiar y manipular en el laboratorio. Los análisis de la base de datos SWISS-PROT han indicado que en los humanos, más de la mitad de las proteínas están glicosidadas<sup>2</sup> y los datos de secuenciación genómica han indicado que aproximadamente el 1% de cada genoma, desde Eubacterias hasta Archaea y Eucariotas, se relaciona con la codificación de enzimas que procesan azúcares.<sup>3</sup> Adicionalmente, la glicosilación tiene implicaciones en diferentes procesos y podría haber sido una fuerza impulsora de la evolución de las formas y funciones biológicas<sup>4</sup> ya que la ingente diversidad estructural de los glicanos puede ser un factor clave en la especiación y evolución de diferentes especies y poblaciones.<sup>5,6</sup>

Los glicanos no sólo son capaces de auto ensamblarse sino que también pueden combinarse con los otros tres grupos de moléculas más importantes para la vida en la tierra, ADN, proteínas y lípidos. Así forman un subgrupo de moléculas que se conocen como glicoconjugados; entre estos destacan las glicoproteínas, proteoglicanos, peptidoglicanos, glicolípidos, nucleótidos, etc. Los glicoconjugados llevan a cabo directamente una gran variedad de funciones como por ejemplo: almacenamiento energético, mantenimiento de la integridad estructural de la célula, almacenamiento y transferencia de la información, reconocimiento de la información, interacciones intercelulares, regulación celular, respuesta inmune, virulencia y defensa química. (Figura 1) Además, la glicosilación de productos naturales es crucial y provee numerosas y nuevas posibilidades para la explotación de moléculas bioactivas.<sup>7</sup>

Por otra parte, los azúcares son el producto químico de la fotosíntesis y los principales nutrientes de las plantas; algunos de estos azúcares forman parte de las paredes celulares de las plantas que están hechas de lignina, proteínas estructurales y otros glicanos. Esto tiene implicaciones en el desarrollo de biocombustibles pues una generación a gran escala de éstos a partir de la biomasa, requerirá de un completo entendimiento de la formación de las paredes celulares. La glicociencia podría también ser clave en la búsqueda de alternativas al petróleo para la manufacturación de materiales poliméricos basados en





sus secuencias aminoacídicas.<sup>13</sup> Este hallazgo inspiró la creación de un nuevo esquema de clasificación de GHs, complementario al de la IUB y cuyo principio subyace en la relación existente entre secuencia y plegamiento. Por lo anterior, es esperable que sin más información que las estructuras primarias se puedan formar familias que deberían mostrar plegamientos similares. De este modo, un total de 301 secuencias de GHs fueron comparadas usando métodos de alineación especialmente diseñados para tratar con grupos de secuencias de baja identidad (menor al 25%),<sup>14,15</sup> definiéndose 35 familias cuando (i) al menos dos secuencias mostraron una similitud significativa de aminoácidos y (ii) cuando una baja similitud de éstos fue encontrada con otra familia. 18 de estas familias son mono-específicas (contienen sólo un código EC) y las otras 17 son poli-específicas conteniendo al menos 2 códigos EC.<sup>12</sup> Los miembros de una familia de GHs comparten el mismo tipo de plegamiento aunque curiosamente, muchas secuencias que fueron agrupadas en familias diferentes muestran no sólo el mismo tipo de plegamiento sino también aparatos catalíticos similares, aun actuando sobre sustratos distintos. Todo esto, secuencias y especificidades no relacionadas junto a aparatos catalíticos y plegamientos similares, sugiere que tales secuencias podrían haber descendido desde un ancestro común, evolución divergente, pero las similitudes de sus secuencias están en el límite de lo detectable.<sup>16</sup>

Además de poder inferir información evolutiva, otra ventaja de esta clasificación es que una proteína, traslación génica o un dominio catalítico puede ser asignado a una familia incluso desconociendo su actividad enzimática.<sup>17</sup> Adicionalmente, otra mejoría clave es que permite adentrarse en el mecanismo catalítico de la enzima, dado que los residuos del sitio activo son esencialmente conservados dentro de una familia de GHs. El explosivo aumento de las estructuras tridimensionales permitió observar que los plegamientos suelen ser más conservados que las secuencias y así, diferentes familias de GHs que presentan plegamientos y maquinarias catalíticas semejantes fueron reagrupados en ‘clanes’ dentro la base de datos CAZy. De los 15 clanes, definidos desde la A hasta la N, el más grande es el clan GH-A, que comprende las familias 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113 y 128 (Tabla 1). Éste incluye una variedad de GHs que retienen la configuración del carbono anomérico y que actúan sobre  $\beta$ -D- o  $\alpha$ -L-glicósidos. Los residuos catalíticos de los miembros de este clan, se localizan en la misma hebra de un barril estructural  $(\alpha/\beta)_8$ .<sup>16,18</sup>

**Tabla 1.** Clanes de GHs, función, topología tridimensional y familias agrupadas por secuencia.

| Clan | Función             | Topología 3D                     | Familias  |
|------|---------------------|----------------------------------|---|
| A    | $\beta$ -Retención  | $(\beta/\alpha)_8$               | 1 2 5 10 17 26 30 35 39 42 50 51 53 59 72 79 86 113 128 |
| B    | $\beta$ -Retención  | Barril $\beta$ tipo remolino     | 7 16  |
| C    | $\beta$ -Retención  | Barril $\beta$ tipo remolino     | 11 12   |
| D    | $\alpha$ -Retención | $(\beta/\alpha)_8$               | 27 31 36  |
| E    | $\alpha$ -Retención | 6 Hélice enrollada $\beta$       | 33 34 83 93   |
| F    | Inversión           | 5 Hélice enrollada $\beta$       | 43 62   |
| G    | $\alpha$ -Inversión | $(\alpha/\alpha)_6$              | 37 63   |
| H    | $\alpha$ -Inversión | $(\beta/\alpha)_8$               | 13 70 77  |
| I    | $\beta$ -Inversión  | $\alpha+\beta$ (Núcleo Lisozima) | 24 46 80  |
| J    | $\alpha$ -Retención | 5 Propelas $\beta$               | 32 68   |
| K    | $\beta$ -Retención  | $(\beta/\alpha)_8$               | 18 20 85  |
| L    | $\alpha$ -Inversión | $(\alpha/\alpha)_6$              | 15 65 125   |
| M    | $\beta$ -Inversión  | $(\alpha/\alpha)_6$              | 8 48  |
| N    | $\alpha$ -Inversión | Hélice $\beta$                   | 28 49   |

Cabe mencionar que las familias 3, 4, 6, 9, 14, 19, 21-23, 25, 29, 38, 40, 41, 43-45, 47, 52, 54-58, 60, 61, 64, 66, 67, 69, 71, 73-76, 78, 81, 82 y 84 no comparten similitudes estructurales con ninguna otra familia GH conocida y por ende, quedan fuera de los clanes.<sup>8</sup>

A continuación, se describirán los tipos de plegamientos de las GHs y después los mecanismos de reacción. Vale la pena notar que a pesar del inmenso número de familias GH conocidas hasta la fecha, sólo una de ellas no comparte el mismo tipo de mecanismo catalítico, la familia GH4. El resto de las familias trabajan a través de un mecanismo de catálisis ácido/base, reteniendo o invirtiendo la configuración del carbono anomérico.

## 1.2.2 Topología 3D y Plegamientos

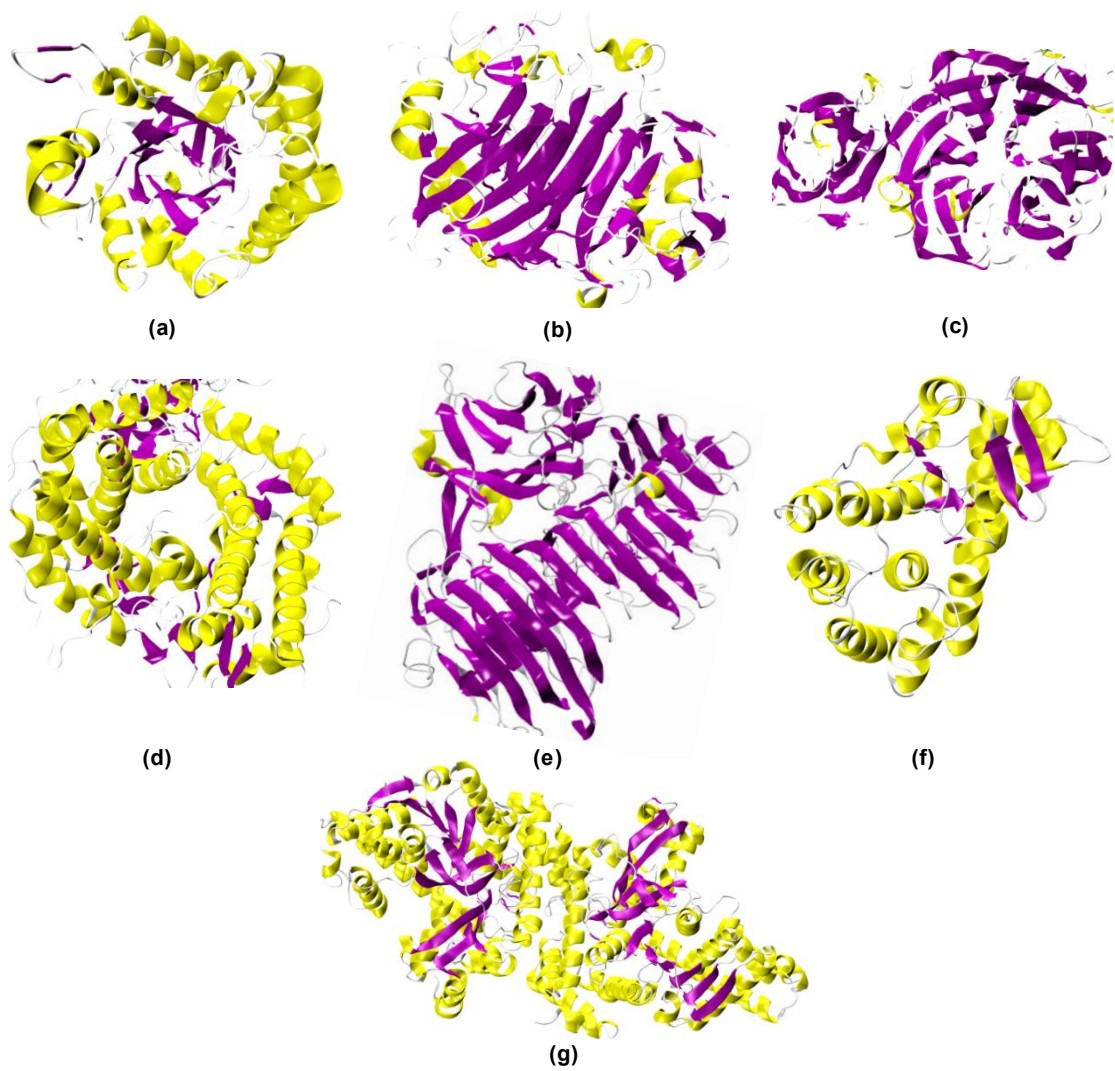
Se ha aprendido mucho desde que se elucidó la primera estructura tridimensional de la glucosidasa lisozima a través de cristalografía de rayos X.<sup>19</sup> En la década de los '90, el número de estructuras 3D se incrementó explosivamente revelando que diferentes plegamientos pueden catalizar el mismo tipo de reacción.

A continuación, se describirán brevemente los tipos de plegamientos mostrados por las glucosidasas, indicando la región ocupada por el sitio activo en cada caso:

- i. Barril ( $\beta/\alpha$ )<sub>8</sub>: consiste en 8 láminas  $\beta$  paralelas que forman el barril interior y la misma cantidad de hélices  $\alpha$  flanqueando en el exterior. Las familias 14, 29, 31, 67 y 84 a lo largo de los clanes A, D, H y K muestran esta topología. En el clan A, el nucleófilo y el residuo ácido-base están ubicados en la lámina 4 y 7 del barril, respectivamente. Aunque en los otros clanes y familias, la posición de estos residuos catalíticos difiere. (Figura 3a, código pdb: 2WHJ)
- ii. Barril  $\beta$  tipo remolino (del inglés  $\beta$  jelly-roll): se puede describir como un barril formado por 4 pares de láminas  $\beta$  antiparalelas, en donde el sitio activo queda localizado en una grieta superficial del barril. Los clanes B y C, muestran este plegamiento. (Figura 3b, código pdb: 4V20)
- iii. Hélice enrollada  $\beta$  (del inglés  $\beta$  propeller): se conforma sólo de láminas  $\beta$  y se caracteriza por presentar desde 4 a 8 'astas' arregladas de manera simétrica en torno a un eje central imaginario. Típicamente, a cada 'asta' la conforman 4 láminas  $\beta$  antiparalelas colocadas de forma tal que la primera y la última hebra son casi perpendiculares. Este tipo de estructura fue primeramente observada en los clanes F y J y el sitio activo se encuentra en una grieta a lo largo de la cara de la propela. Un sexto plegamiento de este tipo es usado por las sialidasas y neuramididasas del clan E, mientras que la familia 74 utiliza un séptimo plegamiento formado por 2 dominios propelas  $\beta$  con en el sitio activo ubicado en una ranura entre los dos dominios. (Figura 3c, código pdb: 4BBW)



- iv. Barriles  $(\alpha/\alpha)_n$ : para describir esta arquitectura usaremos los barriles tipo  $(\alpha/\alpha)_6$ , el cual es utilizado por los clanes G, L, M y las familias de GH9, GH88 y GH105. La estructura contiene 6 hélices  $\alpha$  que forman los 'peldaños' internos del barril, los cuales son rodeados por otras 6 hélices  $\alpha$ . A pesar de las similitudes estructurales entre estos clanes y familias, la cavidad del sitio activo difiere bastante entre uno y otro, al igual que las posiciones de los residuos catalíticos. (Figuras 3d, código pdb: 2VN4 )
- v. Hélice  $\beta$ : esta es una arquitectura inusual que se genera cuando láminas  $\beta$  que hacen giros, acaban por formar hélices paralelas de láminas  $\beta$  o hélices  $\beta$ . Las variaciones en el número de giros de las láminas, el número de láminas formadas y la arquitectura del sitio activo permiten diferenciar entre los clanes y familias que muestran este estilo de plegamiento, que corresponden al clan N y la familia GH82 y GH90. (Figura 3e, código pdb: 1OGM)
- vi. Núcleo  $\alpha+\beta$  (Núcleo Lisozima): formado por hélices  $\alpha$  y hojas  $\beta$  las cuales conforman la cavidad de unión de los sustratos y la ranura catalítica. Es importante decir que en este caso, la similitud de secuencia y plegamiento no es significativa y la conforman sólo algunos miembros de las familias 19, 22 y 23, y el clan I. (Figura 3f, código pdb: 2D05)
- vii. Plegamiento tipo-deshidrogenasa: los miembros de la familia GH4 muestran un arreglo espacial de varios residuos claves del sitio activo parecido al de las deshidrogenasas lactato/malato. Además, comparten un dominio de Rossmann para su cofactor Dinucleótido Nicotinamida Adenina (DNA), a pesar de que la similitud de secuencias es menor al 15%. (Figura 3g, código pdb:1OBB)



**Figura 3.** Representación de la estructura tridimensional de diferentes GHs. Cada plegamiento es representativo de un clan y es conocido como: (a) barril  $(\beta/\alpha)_8$ , (b) barril  $\beta$  tipo remolino, (c) 6 hélices enrolladas  $\beta$ , (d) barriles  $(\alpha/\alpha)_7$ , (e) hélices  $\beta$ , (f)  $\alpha+\beta$  (núcleo lisozima) y (g) tipo-deshidrogenasa.

Con esto queda reflejada la variedad estructural de las GHs.

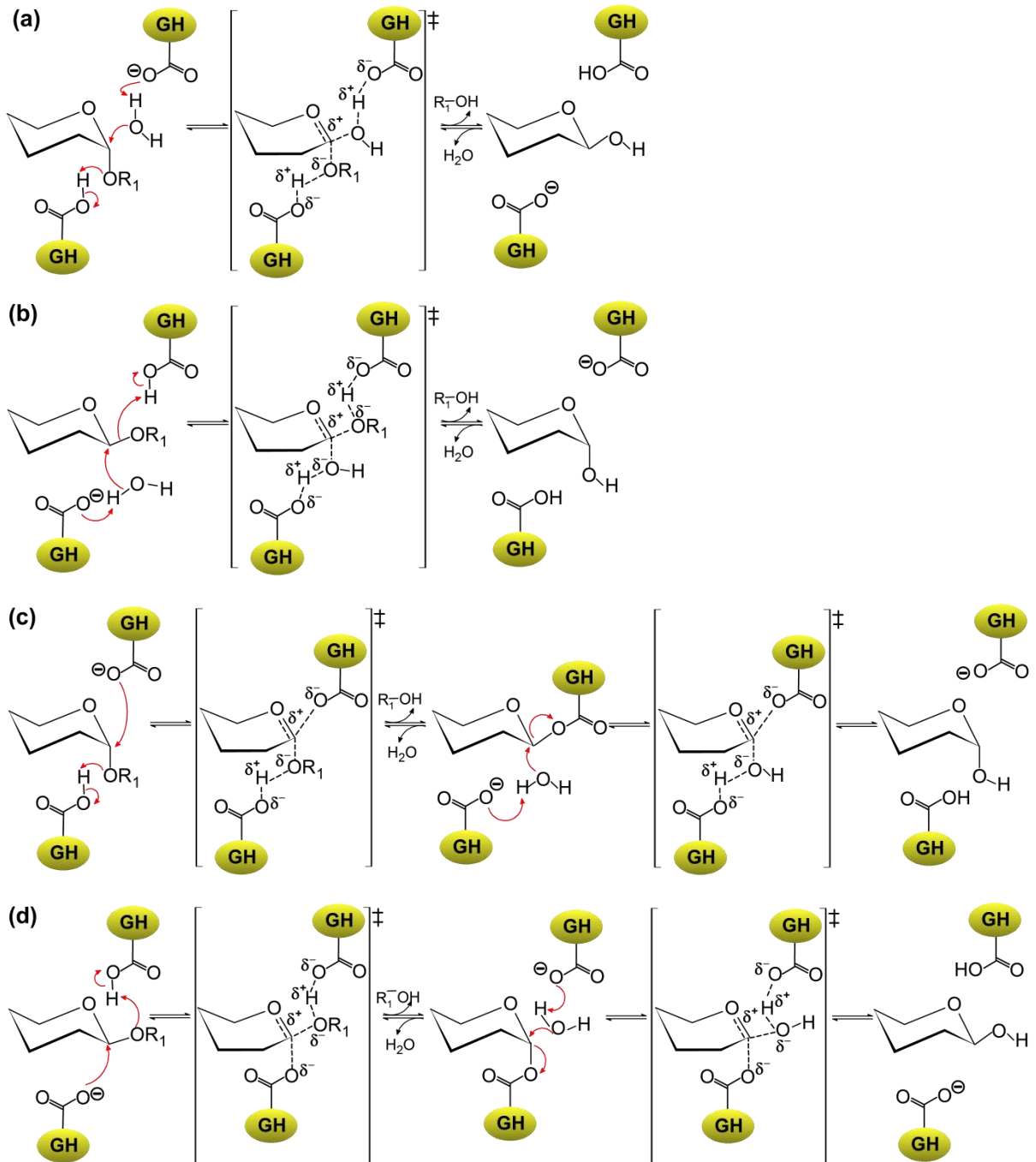
### 1.2.3 Mecanismos de Reacción de las GHs

La base conceptual de como las GHs llevan a cabo la ruptura del enlace glicosídico en una forma estereoespecífica fue explicada por primera vez por Koshland en 1953.<sup>20</sup> (Figura 4) Estas enzimas pueden generar dos tipos de productos estereoquímicos: uno en el que la configuración del carbono anomérico es invertida (GH-inv) (Figura 4a y 4b) y otro en el que es retenida (GH-ret) (Figura 4(c,d)). Koshland propuso que las GHs-inv catalizan el ataque del agua a través de un mecanismo de desplazamiento en una única etapa, mientras que las GHs-ret lo hacen por medio de un mecanismo en dos etapas que involucra un intermedio de reacción. Ambos proceden vía un estado de transición o TS (del inglés Transition State) con alto carácter de oxocarbenio. La estereoquímica del producto es dictada por la disposición espacial de los grupos catalíticos en el sitio activo de la enzima, específicamente, la disposición de dos aminoácidos conservados que llevan a cabo la catálisis ácida y básica necesaria para romper la unión glicosídica. Estos residuos suelen ser Asp o Glu y en el caso de las enzimas que retienen la configuración (GHs-ret), se encuentran aproximadamente a 5.0 Å el uno del otro, mientras que en las enzimas que invierten la configuración (GHs-inv), se encuentran aproximadamente a 9-10 Å, debido a que acomodan una molécula de agua que actúa como nucleófilo entre la base y el carbono anomérico.<sup>21</sup>

De manera general, en el mecanismo de inversión es necesaria la acción de un residuo ácido que protona al grupo saliente o aglucona, facilitando la salida de éste durante el rompimiento del enlace glicosídico. Simultáneamente, la molécula de agua es activada como nucleófilo por el residuo básico. De esta manera, la reacción ocurre a través de una única etapa de desplazamiento vía un estado de transición oxocarbenio (Figura 4a y 4b).

En el mecanismo de retención, la transformación procedería a través de un mecanismo de doble desplazamiento en dos pasos, cada uno de los cuales ocurre vía un estado de transición oxocarbenio. En la primera etapa se forma un intermedio covalente glicosil-enzima (CGE) que requiere de un residuo catalítico bien posicionado que actúe como nucleófilo y desplace a la aglucona. Para facilitar la salida de la aglucona, un residuo ácido/base se encarga de protonarlo y así se completa la primera etapa (Figura 4d y 4e). Luego, el CGE se hidroliza mientras la molécula de agua es deprotonada por el residuo ácido/base consiguiendo transformarla en un buen nucleófilo, (Figura 4c y 4d). Aunque este mecanismo es ampliamente consensuado para estas enzimas, no es el único que históricamente ha sido propuesto en la literatura pues ya en 1967, Phillips propuso un mecanismo de hidrólisis basado en la estructura cristalográfica de la lisozima de clara de huevo de gallina, la primera enzima para la que fue resuelta una estructura por cristalografía de rayos X.<sup>22</sup> En el mecanismo original de Phillips se sugirió que el azúcar se unía a la enzima en una conformación distorsionada con respecto a la silla para evitar interacciones estéricas con algunos residuos; además se planteó un mecanismo tipo  $S_N1$  con un intermedio oxocarbenio en el que el azúcar estaría adoptando una conformación de media silla semejante a la que sería necesaria para unirse a la enzima en primer lugar. De esta manera, la ruptura del enlace glicosídico sería promovida por un residuo que protona el oxígeno glicosídico (Glu35) mientras un residuo cargado negativamente (Asp52) estabiliza el intermedio oxocarbenio.<sup>22</sup> Aunque este mecanismo fue posteriormente sustituido por el de Koshland, que es el aceptado para la mayoría de GH-ret, en algunos libros de texto aún se

puede encontrar descrito para la lisozima de clara de huevo de gallina. Actualmente, diversas técnicas experimentales como cristalografía y espectroscopia de masas,<sup>23-25</sup> así como estudios teóricos del mecanismo, apuntan claramente a que el doble desplazamiento es el mecanismo utilizado por la lisozima.<sup>26,27</sup>



**Figura 4.** Mecanismos de Koshland para la hidrólisis de glicanos catalizada por GHs ( $R_1$  representa un azúcar). Según el producto estereoquímico obtenido, cada mecanismo es ejecutado por: (a)  $\alpha$ -GH-inv, (b)  $\beta$ -GH-inv, (c)  $\alpha$ -GH-ret y (d)  $\beta$ -GH-ret.

A pesar de que la propuesta mecanística de Phillips resultó no ser correcta, su comentario acerca de la distorsión del anillo en el complejo de Michaelis es de gran trascendencia y hoy en día la conformación del sustrato es un tópico infaltable en todas las discusiones que se refieren al mecanismo catalítico de las GHs, sin importar si se trata de GH-inv o GH-ret. Experimentos de efecto isotópico<sup>28</sup> han indicado que ambos mecanismos, inversión y retención, involucran un TS tipo oxocarbenio; esta especie se caracteriza por una hibridación  $sp^2$  y el desarrollo de una carga positiva en el carbono anomérico, la que puede ser parcialmente estabilizada por donación electrónica del oxígeno del heterociclo. La estructura que adopta el azúcar para satisfacer las necesidades electrónicas de esta especie es distorsionada y no corresponde a la silla  ${}^4C_1$ . Las conformaciones distorsionadas que puede adoptar el azúcar para formar esta especie oxocarbenio incluyen 2 botes ( $B_{2,5}$ ,  ${}^{2,5}B$ ), 2 medias sillas ( ${}^4H_3$ ,  ${}^3H_4$ ) y 4 sobres ( ${}^3E$ ,  $E_3$ ,  ${}^4E$ ,  $E_4$ ); por consiguiente se esperaría que los TS asociados a reacciones catalizadas por GHs adopten alguna de estas conformaciones.<sup>29</sup> Adicionalmente, estudios de cristalografía de rayos X y de Resonancia Magnética Nuclear (RMN) han mostrado que las GHs albergan los sustratos en una conformación distorsionada, específicamente el monómero que se ubica en el sitio -1 (el monómero que se hidroliza y forma el CGE),<sup>30-32</sup> algo que ya había señalado Phillips en su estudio pionero con la lisozima de clara de huevo de gallina. Esta distorsión del azúcar en el complejo de Michaelis fue tempranamente asociada con la adopción de una conformación más cercana a la del TS de la reacción y luego se relacionó con la reducción de ciertas interacciones estéricas y con el adecuado posicionamiento (pseudo-axial) del oxígeno glicosídico para la protonación por parte del residuo ácido/base, facilitándose el ataque nucleofílico sobre el carbono anomérico.<sup>33</sup> Estas especiales características del azúcar en el sitio -1 han sido confirmadas para varios complejos de Michaelis de diferentes GHs a través de simulaciones QMMM.<sup>29,34</sup>

Finalmente, la sistemática distorsión del azúcar en el sitio -1 de los complejos de Michaelis de diversas GHs, y la distorsión del azúcar en el TS parece ser una propiedad intrínseca del sustrato más que algo inducido por la enzima<sup>29</sup> ha llevado a la búsqueda de las conformaciones que serían más adecuadas para la catálisis cuantificando las propiedades del azúcar que se unirá a la enzima pero fuera del ambiente enzimático (azúcar aislado). Así, se podrían predecir cuales conformaciones estarían presentes en una determinada familia de GH durante la reacción química. En este ámbito, el trabajo de Carme Rovira está proporcionando detalles moleculares muy valiosos a la comunidad científica.

### 1.3 Visión General de las Glicosiltransferasas (GTs)

Las enzimas que se estudian en esta tesis son glicosiltransferasas (GTs). Las GTs (EC 2.4.x.y) catalizan la formación de un enlace glicosídico de forma altamente estereo y regioespecífica. Utilizan un sustrato dador del monosacárido que se transferirá en el que el azúcar está activado, y un sustrato aceptor que puede ser un mono-, oligo- o polisacárido, una proteína, un lípido, fragmentos de ADN o una gran variedad de moléculas orgánicas. Esta función las convierte en actores fundamentales en la biosíntesis de polisacáridos y glicoconjugados, moléculas de tamaños variados que llevan a cabo un amplio abanico de funciones que van desde estructura y almacenamiento, hasta señalización altamente específica.<sup>35</sup> Análogamente a lo visto para las GHs, la reacción de transferencia puede darse con inversión o retención de la configuración del carbono anomérico del azúcar transferido, de modo que se distingue entre GTs-inv y GTs-ret. Estas enzimas están presentes tanto en organismos eucariotas como procariontes y generalmente muestran una alta especificidad por el sustrato dador que contiene el azúcar y por el sustrato aceptor. En los organismos eucariotas, las GTs se localizan normalmente en el aparato de Golgi; las GTs asociadas al aparato de Golgi son proteínas transmembrana de tipo II que poseen un gran dominio catalítico globular C-terminal que se orienta hacia la parte luminal del aparato de Golgi.<sup>36</sup>

A pesar de la gran diversidad de secuencias aminoacídicas que se conocen de GTs, las semejanzas estructurales son sorprendentes.<sup>37</sup> Todas las GTs son proteínas globulares para las que inicialmente fueron identificadas sólo 2 topologías tridimensionales globales; una de ellas es el plegamiento GT-A que posee un dominio N-terminal que se asocia al nucleótido-azúcar (sustrato dador) y está formado por varias hebras  $\beta$  flanqueadas por hélices  $\alpha$ , semejante a los dominios tipo-Rossmann. El dominio C-terminal consiste casi completamente en una mezcla de láminas  $\beta$  en donde se encuentra la cavidad en la que se alojará el sustrato aceptor. Cabe destacar que hasta antes del año 2004, todas las enzimas con plegamiento GT-A que se conocían mostraban un motivo de tres residuos Asp-X-Asp (DXD) o Glu-X-Asp (EXD), que participa en la unión del cofactor metálico que necesitan muchas de las enzimas GT-A para llevar a cabo la catálisis.<sup>38</sup> Contrariamente, en el plegamiento GT-B no aparece el motivo DXD ni el metal pero existen dos dominios tipo Rossmann en vez de uno, en el dominio N-terminal se encuentra el sitio de unión del aceptor y en el dominio C-terminal se alojará el sustrato dador. Más tarde, fue reportada la estructura cristalográfica de dos enzimas que mostraban plegamientos distintos al GT-A o GT-B. Una de ellas corresponde a la enzima bifuncional GT peptidoglicano,<sup>39</sup> más específicamente el dominio o módulo GT<sub>51</sub> (familia GT51)<sup>8</sup> al cual se le asignó un plegamiento tipo-lizosima y que usa un fosfolípido-azúcar como sustrato dador; la otra enzima corresponde al dominio C-terminal soluble de *Pyrococcus* STT3 perteneciente a la familia GT66, una de las 8 subunidades que forman una OST (OligoSacarilTransferasa) que usa un fosfolípido como sustrato dador de azúcar también.<sup>40</sup> Previamente y través de métodos bioinformáticos, se predijo que los miembros de la familia GT66 (además de otras que se indicarán más adelante) adoptarían un nuevo tipo de plegamiento, el GT-C.<sup>41</sup> Sin embargo, esta estructura debilitó la clasificación GT-C debido a que reveló que la porción de secuencia aminoacídica en la cual se basó el alineamiento correspondía a una región transmembrana que no contiene al lazo en el que supuestamente se encontraría el sitio



**Figura 6.** Complejos activados dadores de azúcar comúnmente conocidos como: (a) dadores Leloir o nucleótido mono y difosfato azúcar, respectivamente y (b) dadores no-Leloir que mantienen al menos un grupo fosfato enlazado al azúcar pero no contienen un nucleósido.

Típicamente, la mayoría de las enzimas con plegamiento GT-A catalizan la ruptura del enlace nucleótido-azúcar estabilizando la carga del nucleótido saliente con un metal divalente como  $Mg^{2+}$  o  $Mn^{2+}$ , el cofactor. Es importante destacar que el grupo saliente no es expulsado o desplazado del sitio activo luego de producirse la ruptura del enlace pues éste sigue interaccionando con el sustrato aceptor a través de un enlace de hidrógeno fuerte con el grupo hidroxilo nucleofílico. Esta interacción clave no sólo promueve la ruptura del enlace nucleótido-azúcar sino que también favorece el ataque nucleofílico necesario para completar la reacción de transferencia.<sup>42</sup> Por otra parte, también se conocen algunas GT-A que no trabajan con el cofactor metálico y lo reemplazan por aminoácidos cuyas cadenas laterales tienen una carga neta positiva o son polares. Sólo por mencionar un par de ejemplos, la sialiltransferasa de la familia GT42 (CstII) muestra a la Tyr156, a la Tyr162 y a la Asn31 interaccionando con el nucleósido-fosfato;<sup>43</sup> la  $\beta$ 1,6-N-acetilglucosaminiltransferasa núcleo 2 de tipo Leucocito (C2GnT-L) es una GT14 que posee una Arg378 y una Lys401 interaccionando con el nucleósido-difosfato.<sup>44</sup> Curiosamente, el uso de estos aminoácidos es reminiscente de la estrategia independiente de metal usada por las enzimas de plegamiento GT-B; la evidencia estructural proporcionada por CstII y C2GnT-L sugiere un proceso de convergencia de mecanismos catalíticos entre las enzimas GT-A y GT-B.<sup>45</sup> Posteriores análisis bioinformáticos han mostrado que estas dos superfamilias de enzimas carecen de semejanzas a un nivel local y global, lo que también sugiere que han evolucionado independientemente.<sup>41</sup>

### 1.3.1 Clasificación

La dificultad en la clasificación de las GTs, al igual que para las GHs, tiene su origen en la inmensa variedad estructural y funcional exhibida por estas enzimas. Por ejemplo, un grupo de sustituciones aminoacídicas puede cambiar la especificidad de una enzima e incluso alterar la especificidad por el enlace dependiendo de la naturaleza del sustrato involucrado.<sup>38,46</sup>

Una clasificación de las GTs que utilizan como complejo dador de azúcar nucleótidos monofosfato, comúnmente uridil difosfato (UDP en inglés), fue introducida ya hace bastante tiempo,<sup>35,47,48</sup> pero no fue hasta el año 2003 que se realizó un estudio completo de secuencias relacionadas con GTs, sobre 7200 hasta esa fecha, ampliando la clasificación para incluir las GTs que usan otros complejos dadores de azúcar como dolicol-difosfato, 1-fosfato, nucleótidos o un lípido-difosfato. (Figura 6a y 6b).<sup>3</sup> Las GTs que no utilizan dadores de tipo Leloir se llaman también GTs no-Leloir. Dentro de este último grupo se encuentran por ejemplo las fosforilasas; estas enzimas pueden catalizar la formación de un enlace glicosídico utilizando como dador un grupo fosfato o un pirofosfato-azúcar y también catalizar la ruptura de este enlace formando un fosfato-azúcar (reacción reversible). Este es el motivo por el que algunas de ellas están clasificadas como GHs (sucrosa fosforilasa de GH13 o celobiosa fosforilasa de GH94), mientras que otras quedan dentro de familias de GTs como la glicógeno fosforilasa o la almidón fosforilasa, familia GT35.<sup>49</sup> Cabe



mencionar que aquellas GTs que usan disacáridos, oligosacáridos o polisacáridos como dadores de azúcar, tales como ciclomaltoextrina glucanotransferasa (EC 2.4.1.19, GH13), dextransucrasas (EC 2.4.1.5, GH70), xiloglucano endotransglicosidasa (EC 2.4.1.207, GH12), son en realidad transglicosidasas y están relacionadas con las GHs en un nivel estructural, mecanístico y evolutivo por lo que han sido asignadas a familias de GHs.<sup>12,17,50</sup> Adicionalmente, si bien las denominadas fosfato-1-azúcar transferasas han sido clasificadas como GTs, en realidad catalizan la reacción entre un UDP-azúcar y un fosfato dolicol para formar UMP (Uridil MonoFosfato) y un azúcar-difosfatodolicol. Es decir, catalizan la sustitución en un átomo de fósforo y no hay transferencia de azúcar, su código es EC 2.7.8.x.<sup>3</sup>

Como ya se ha indicado, una de las clasificaciones existentes se basa en el tipo de resultado estereoquímico que genera la reacción, en donde la configuración del carbono anomérico puede resultar invertida o retenida. También se probó una clasificación basada en la especificidad hacia determinados sustratos,<sup>24</sup> sin embargo la utilidad de tal clasificación quedaba muchas veces limitada tanto para las GHs como para las GTs ya que no da información estructural y deja de funcionar bien para aquellas enzimas que pueden actuar sobre diferentes sustratos.<sup>35</sup> Así, tal como fue descrito previamente para las GHs, las GTs han sido clasificadas en familias bajo un criterio de semejanza en la secuencia aminoacídica y también usando algoritmos de reconocimiento de plegamientos tridimensionales (3D)<sup>51</sup> (Tabla 2).

**Tabla 2.** Función, topología tridimensional y familias agrupadas por secuencia con plegamiento conocido o no.

| Función   | Topología 3D     | Familias   |
|-----------|------------------|--|
| Inversión | GT-A y variantes | <b>2 7 12 13 14 16 17 21 25 29 31</b> 40 42 <b>43 49 54</b> 67 73 74 75 82 84 92           |
| Inversión | GT-B y variantes | <b>1 9 10 11 18</b> 19 <b>23 28 30 33</b> 37 38 <b>41 47 52 56 61 63 65 68 70 80 90 94</b> |
| Retención | GT-A y variantes | <b>6 8 15 24 27 32</b> 34 <b>44 45 55 60 62 64 69 71 77 78 81 88</b>                       |
| Retención | GT-B y variantes | <b>3 4 5 20 35 72 79 93</b>  |
| Inversión | Otras            | <b>51 66</b>   |
| Inversión | No conocida      | <b>22 26 39 48 50 53 57 58 59 76</b> 83 85 87  |
| Retención | No conocida      | 89   |

Las familias con caracteres en negrita tienen al menos un miembro con estructura 3D resuelta,<sup>52</sup> las familias para las que se ha predicho que pueden adoptar cualquiera de los dos plegamientos establecidos, GT-A o GT-B, se indican en cursiva.<sup>36</sup> Las familias en rojo comprenden secuencias humanas y las familias con plegamiento desconocido son proteínas integrales de membrana con excepción de la familia GT26.

Es interesante comentar que se esperaría que esta clasificación basada en secuencia permita predecir aspectos estructurales y mecanísticos (inversión o retención), tal como pasa con las GHs en las cuales una vez se ha establecido el mecanismo para uno de los miembros de una determinada familia, el mecanismo puede ser con seguridad extendido al resto de los miembros de ésta. Esto parece no aplicar para las GTs pues se han reportado similitudes distantes entre familias de GT-inv y GT-ret.<sup>41,51</sup> Tales resultados indican que el mecanismo catalítico no puede ser siempre predicho con seguridad basándose únicamente en comparaciones de secuencias.<sup>36</sup>

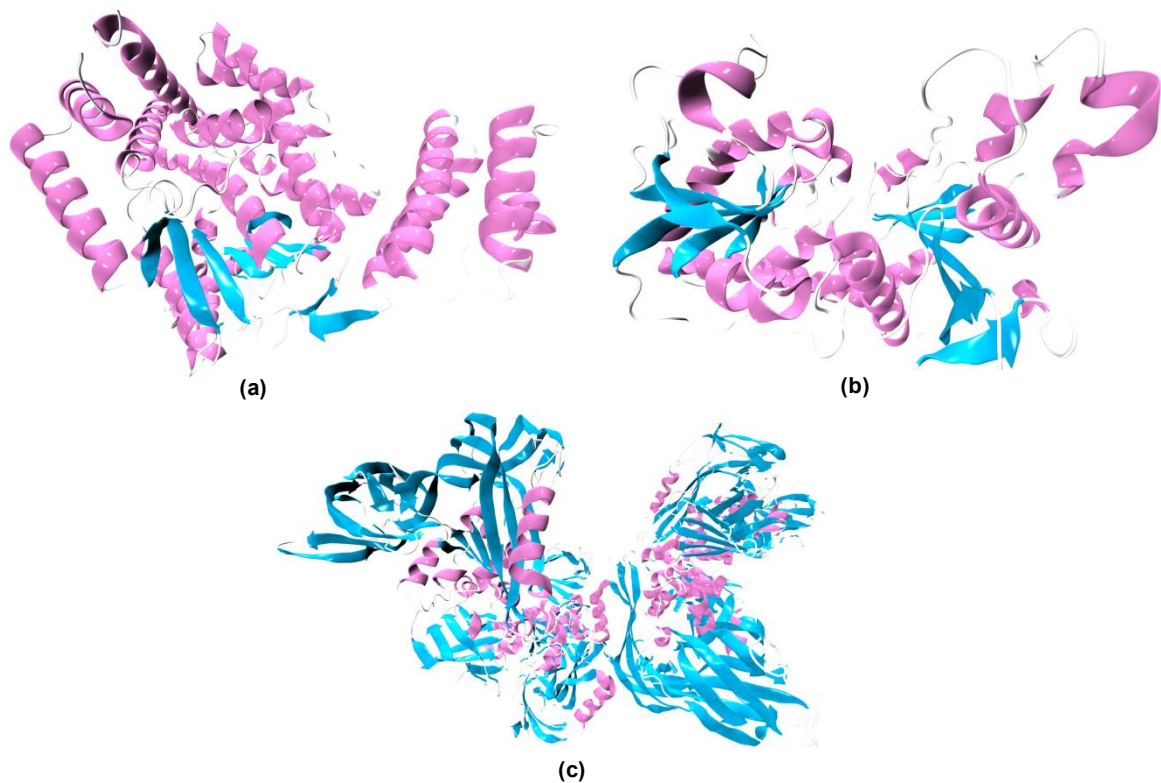
### 1.3.2 Topología 3D y Plegamientos

Para empezar, es importante comentar que la determinación de estructuras tridimensionales de las GTs ha sido un campo marcadamente obstaculizado por las dificultades para expresarlas en cantidades grandes, purificarlas y cristalizarlas. La primera estructura cristalográfica de una GT se publicó en 1994, pero en los últimos años ha habido un progreso considerable en el área. Así, por ejemplo, en 2006 ya se disponía de más de 100 estructuras correspondientes a 23 GTs distintas y cubriendo 17 familias; en 2012, el número de familias de GTs para las que se había resuelto la estructura tridimensional de al menos un representante ya se elevaba a 38.<sup>53</sup>

El número de plegamientos exhibidos por las GTs es bastante reducido habiéndose caracterizado inicialmente sólo dos tipos, el GT-A y el GT-B (Figura 7), a los que se cree pertenecen todas las GTs que utilizan nucleótidos monofosfatos dadores de azúcar. Además de esto, nuevos análisis de *threading* han predicho que muchas de las familias no caracterizadas de GTs deberían adoptar uno de estos dos plegamientos. Por un lado, esto podría reflejar la restricción que supone poseer un motivo de interacción/asociación para el nucleótido y por otra parte, apunta a la posible evolución desde un grupo reducido de secuencias progenitoras.<sup>45</sup> El plegamiento GT-A fue primeramente observado en la enzima *Bacillus subtilis* SpsA de la familia GT2 y consiste en un “sandwich”  $\alpha/\beta/\alpha$  cuya lámina  $\beta$  está formada por 7 u 8 hebras rodeadas por hélices  $\alpha$ , (Figura 7a) lo cual es reminiscente de plegamientos tipo Rossmann típicamente exhibidos por complejos proteína-nucleótido.<sup>36</sup> El plegamiento GT-A suele ser considerado como un único dominio plegado pero aun así se observa la presencia de un dominio N-terminal al cual se asocia el nucleótido y otro dominio C-terminal para la asociación del sustrato aceptor. El dominio del aceptor se aloja en una fisura rodeado por un lazo desordenado, mientras que el sustrato dador se aloja dentro de una fisura más profunda interaccionando directamente con varios residuos.<sup>54</sup> Frecuentemente, el motivo DXD es considerado como una característica distintiva de las GTs que muestran el plegamiento GT-A<sup>45</sup> y de hecho, en la mayoría de las GT-A está presente y participa en la unión del cofactor que comúnmente es un  $Mn^{2+}$ . Uno o dos de los aspartatos del motivo coordinan al catión divalente mientras que el resto del motivo interacciona con el sustrato dador. Sin embargo, estos no son motivos estrictamente conservados a lo largo de las enzimas que poseen el plegamiento GT-A, ya que existen enzimas con este mismo tipo de plegamiento que no tienen el motivo como por ejemplo, la GT-inv  $\beta$ 1,6-N-acetilglucosaminiltransferasa núcleo 2 de tipo Leucocito (C2GnT-L), una GT-A de la familia GT14;<sup>44</sup> o la GT-inv bifuncional sialiltransferasa de *Campylobacter jejuni* (CstII) de la familia GT42, la primera enzima independiente de metal para la que se reportó un plegamiento GT-A.<sup>43</sup> En un principio, diferentes análisis que involucraron elementos de la estructura secundaria y la topología de las láminas  $\beta$  de CstII,<sup>55</sup> condujeron a considerar que esta enzima podía representar un nuevo tipo de plegamiento el cual denominaron  $\alpha$ -2,3/8-sialiltransferasa CstII (SCOP fold); posteriormente tanto CstII como Cst-I, una GT-inv procedente de la misma bacteria y que pertenece también a la familia GT42,<sup>56</sup> se acabaron considerando simplemente variantes del plegamiento GT-A ya que sus diferencias se limitan a diferentes ordenamientos dentro de la lámina  $\beta$  central característica de este plegamiento.<sup>45,53</sup> De esta manera, aunque la gran mayoría de las GT-A poseen el motivo DXD, existen ejemplos de GT-A que carecen de la firma DXD.

El primer plegamiento GT-B conocido fue también la primera estructura cristalográfica conseguida para una GT dependiente de nucleósidos-difosfato y corresponde a la bacteriófaga T4  $\beta$ -glucosiltransferasa (BGT).<sup>57</sup> Este plegamiento resultó ser análogo al mostrado por la glucógeno fosforilasa y ha sido observado en varias otras GTs también. El plegamiento GT-B, está formado por dos dominios tipo Rossmann separados por una fisura profunda en la cual se localiza el sitio activo. (Figura 7b) El sustrato dador se aloja en la cima de la ranura interaccionando únicamente con residuos del dominio C-terminal y como la ranura está alineada con varios residuos cargados positivamente, el sustrato aceptor (una doble hebra de ADN) se puede alojar en esta misma cavidad.<sup>54</sup> Es importante añadir que también existen variantes de este plegamiento como por ejemplo, la enzima  $\alpha$ 1,6-fucosiltransferasa de mamífero, cuyo dominio catalítico es muy parecido al plegamiento GT-B.<sup>58</sup>

Además de los plegamientos ya mencionados, en el año 2003 fue sugerido un tercer tipo de plegamiento para las GTs, el denominado GT-C (Figura 7c). Mediante análisis iterativos de secuencias usando programas como BLAST,<sup>41</sup> se predijo que 8 familias de GTs (GT22, GT39, GT48, GT50, GT53, GT57, GT58 y GT59) podrían adoptar el plegamiento GT-C. Posteriormente, las familias GT66, GT83, GT85 y GT86 fueron también predichas para adoptar este plegamiento por la base de datos CAZy.<sup>45</sup> El plegamiento GT-C se relaciona con grandes proteínas integrales de membrana ubicadas en el retículo endoplasmático, exhibiendo de 8 a 13 hélices transmembrana y un sitio activo localizado en la región de un lazo de gran longitud. Las familias para las que se ha predicho adoptar este último tipo de plegamiento usan un lípido-fosfato como complejo activado dador de azúcar, con la excepción de los miembros de las familias GT48 y GT53 que usan UDP-glucosa y UDP-L-arabinosa, respectivamente. Como se mencionó anteriormente, en el año 2007 se determinó la estructura del dominio C-terminal soluble de *Pyrococcus furiosus oligosacaryltransferasa* STT3 asignada a la familia GT66.<sup>40</sup> Esta estructura reveló que la secuencia de aminoácidos en la cual se basó el alineamiento no incluye al lazo en el que se predijo que estaría el sitio activo, por consiguiente no ha podido ser demostrada ninguna actividad catalítica. De esto se deduce que la mayor similitud de estas enzimas podría no estar en el dominio catalítico sino en la región transmembrana; en otras palabras, no queda claro que la predicción de la topología GT-C pueda ser asignada con éxito a nuevas secuencias de GTs y lo único que indica de manera inequívoca, es la presencia de un gran componente trans-membrana.<sup>45</sup>



**Figura 7.** Representación de la estructura tridimensional de diferentes GTs. Cada plegamiento es representativo de una superfamilia y es conocido como: (a) GT-A código PDB 4dmw GT44, (b) GT-B código PDB 1jg7 GT63 y (c) GT-C código PDB 2zag.

En resumen, el conjunto de información estructural y bioinformática acumulada dan información básicamente de tres cosas:

- i. Superfamilia GT-A: es ampliada debido a que se encontró que las GTs dependientes de un azúcar nucleósido-difosfato y las nucleótido-transferasas comparten rasgos de su estructura tridimensional y forman una familia monofilética (descienden de una única secuencia o grupo de secuencias progenitoras que no comparten con ningún otro grupo).
- ii. Superfamilia GT-B: a pesar de que muestran dos dominios tipo Rossmann y plegamientos tridimensionales parecidos, no se encontraron semejanzas de secuencia con la superfamilia GT-A ni viceversa. Esto apunta a que estas enzimas han evolucionado independientemente (convergencia) o que su último ancestro común, si es que existió, es extremadamente antiguo.
- iii. Superfamilia GT-C: ésta relaciona 8 familias CAZy de GTs. La arquitectura predicha para el plegamiento GT-C es observada en grandes proteínas hidrofóbicas integrales de membrana, localizadas en el retículo endoplasmático sin embargo, la topología GT-C podría no ser asignada con éxito a nuevas secuencias de GTs.<sup>41,45,59-61</sup>

Futuros análisis de las familias clasificadas como GTs a las que aún no se les ha podido asignar ningún tipo de plegamiento por análisis de secuencia, deberían poder aportar nueva información del origen evolutivo de las GTs.

Varias GTs dependientes de nucleótidos-azúcar, que son las enzimas que se estudian en esta tesis, y que pueden mostrar un plegamiento GT-A o GT-B poseen al menos un lazo flexible que tiene un importante papel en la catálisis. Debido al alto grado de flexibilidad de este lazo, en la mayoría de los casos no es posible trazarlo ni en la apo ni en la holo enzima pero se encuentra siempre en la vecindad del sitio de unión del nucleótido-azúcar.<sup>38</sup> Un caso especial es la enzima  $\beta$ 1,4-galactosiltransferasa T1 ( $\beta$ 4Gal-T1, GT-A inversora),<sup>62-64</sup> en la que casi todos los residuos que forman los dos lazos flexibles han podido ser trazados en ambas formas, la apo y el complejo enzima-nucleótido-azúcar más metal. En la forma apo los residuos del sitio activo de ambos lazos están lejos de la cavidad activa mientras que en la holo los residuos se acercan al nucleótido formando interacciones polares, electrostáticas o hidrofóbicas (CH- $\pi$  stacking); esto da origen a dos tipos de conformaciones, la abierta y la cerrada, respectivamente. Este cambio conformacional inducido por la unión del sustrato ha sido observado para otras enzimas como por ejemplo: la  $\alpha$ 1,3-galactosiltransferasa ( $\alpha$ 3GT, GT-A retenedora),<sup>65,66</sup> la  $\alpha$ 1,4-galactosiltransferasa (LgtC, GT-A retenedora),<sup>67</sup> la enzima N-acetilglucosaminiltransferasa I (GnT I, GT-A inversora),<sup>68</sup> la enzima polipéptido N-Acetilgalactosaminiltransferasa-2 (ppGalNAcT-2, GT-A retenedora) trabajo en el cual se acuñó el término "conformación activa" para la conformación cerrada e "inactiva" para la abierta,<sup>69</sup> y también dos de las enzimas que se estudian en esta tesis,  $\alpha$ 1,4-N-Acetilhexosaminiltransferase que puede transferir GalNAc y GlcNAc (EXTL2, GT-A retenedora)<sup>70</sup> y la Glucosil-3-fosfoglicerato sintasa (GpgS, GT-A retenedora).<sup>71</sup> El reordenamiento de este lazo inducido por la unión del sustrato dador y ubicado en el extremo C-terminal de la secuencia, crea una especie de tapa sobre el sustrato dador facilitando la formación de interacciones entre el UDP y diferentes residuos; esta nueva conformación cerrada crea una cavidad que servirá para unir el sustrato aceptor.<sup>36,42,65</sup> Adicionalmente, posteriores estudios estructurales, calorimétricos y de docking mostraron por una parte que el sustrato dador adquiere una conformación distorsionada en la enzima con respecto a la que presenta en solución acuosa; por otra parte, los complejos enzima-sustrato se forman de una manera ordenada, asociándose primero el sustrato dador y luego el aceptor ya que el estado abierto de la enzima o conformación inactiva, resulta tener muy poca afinidad por el sustrato aceptor. Se sugiere que esta unión ordenada de los sustratos se relaciona con un mecanismo de prevención de entrada de agua para evitar la hidrólisis del UDP-azúcar. Así, la conformación distorsionada del nucleótido azúcar al colocarse en el sitio activo podría ser crucial para conseguir la alta eficiencia de la reacción catalizada.<sup>38</sup>

Cabe decir que no existe correlación entre el plegamiento adoptado y la estereoquímica de la reacción catalizada, de modo que existen GTs-inv y GTs-ret en los diferentes plegamientos. Es interesante destacar también, que incluso presentando poca diversidad en el tipo de plegamiento, las GTs muestran una gran variabilidad estructural en el dominio al que se une el sustrato aceptor, lo que da lugar al amplio abanico de sustratos aceptores utilizados por estas enzimas, manteniendo cada GT una alta especificidad para dicho sustrato.

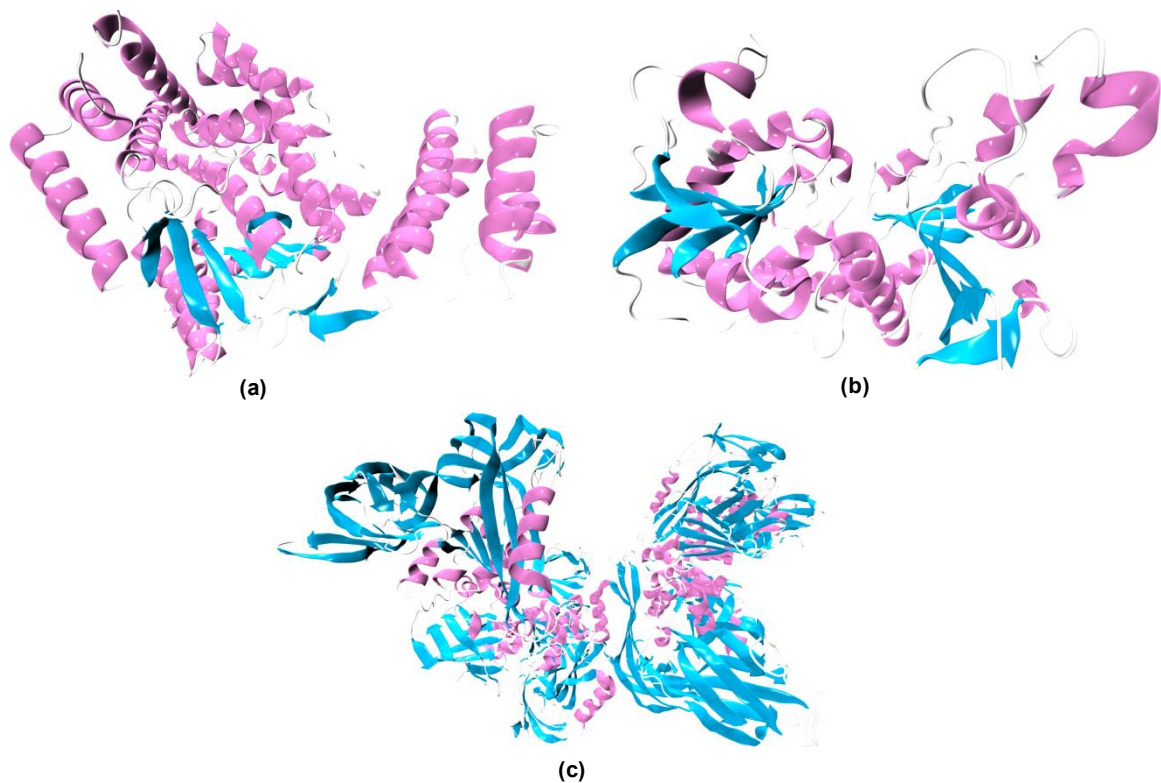
### 1.3.3 Mecanismos Catalíticos de las GTs

Para empezar, es importante comentar que la determinación de estructuras tridimensionales de las GTs ha sido un campo marcadamente obstaculizado por las dificultades para expresarlas en cantidades grandes, purificarlas y cristalizarlas. La primera estructura cristalográfica de una GT se publicó en 1994, pero en los últimos años ha habido un progreso considerable en el área. Así, por ejemplo, en 2006 ya se disponía de más de 100 estructuras correspondientes a 23 GTs distintas y cubriendo 17 familias; en 2012, el número de familias de GTs para las que se había resuelto la estructura tridimensional de al menos un representante ya se elevaba a 38.<sup>53</sup>

El número de plegamientos exhibidos por las GTs es bastante reducido habiéndose caracterizado inicialmente sólo dos tipos, el GT-A y el GT-B (Figura 7), a los que se cree pertenecen todas las GTs que utilizan nucleótidos monofosfatos dadores de azúcar. Además de esto, nuevos análisis de *threading* han predicho que muchas de las familias no caracterizadas de GTs deberían adoptar uno de estos dos plegamientos. Por un lado, esto podría reflejar la restricción que supone poseer un motivo de interacción/asociación para el nucleótido y por otra parte, apunta a la posible evolución desde un grupo reducido de secuencias progenitoras.<sup>45</sup> El plegamiento GT-A fue primeramente observado en la enzima *Bacillus subtilis* SpsA de la familia GT2 y consiste en un “sandwich”  $\alpha/\beta/\alpha$  cuya lámina  $\beta$  está formada por 7 u 8 hebras rodeadas por hélices  $\alpha$ , (Figura 7a) lo cual es reminiscente de plegamientos tipo Rossmann típicamente exhibidos por complejos proteína-nucleótido.<sup>36</sup> El plegamiento GT-A suele ser considerado como un único dominio plegado pero aun así se observa la presencia de un dominio N-terminal al cual se asocia el nucleótido y otro dominio C-terminal para la asociación del sustrato aceptor. El dominio del aceptor se aloja en una fisura rodeado por un lazo desordenado, mientras que el sustrato dador se aloja dentro de una fisura más profunda interaccionando directamente con varios residuos.<sup>54</sup> Frecuentemente, el motivo DXD es considerado como una característica distintiva de las GTs que muestran el plegamiento GT-A<sup>45</sup> y de hecho, en la mayoría de las GT-A está presente y participa en la unión del cofactor que comúnmente es un  $Mn^{2+}$ . Uno o dos de los aspartatos del motivo coordinan al catión divalente mientras que el resto del motivo interacciona con el sustrato dador. Sin embargo, estos no son motivos estrictamente conservados a lo largo de las enzimas que poseen el plegamiento GT-A, ya que existen enzimas con este mismo tipo de plegamiento que no tienen el motivo como por ejemplo, la GT-inv  $\beta$ 1,6-N-acetilglucosaminiltransferasa núcleo 2 de tipo Leucocito (C2GnT-L), una GT-A de la familia GT14;<sup>44</sup> o la GT-inv bifuncional sialiltransferasa de *Campylobacter jejuni* (CstII) de la familia GT42, la primera enzima independiente de metal para la que se reportó un plegamiento GT-A.<sup>43</sup> En un principio, diferentes análisis que involucraron elementos de la estructura secundaria y la topología de las láminas  $\beta$  de CstII,<sup>55</sup> condujeron a considerar que esta enzima podía representar un nuevo tipo de plegamiento el cual denominaron  $\alpha$ -2,3/8-sialiltransferasa CstII (SCOP fold); posteriormente tanto CstII como Cst-I, una GT-inv procedente de la misma bacteria y que pertenece también a la familia GT42,<sup>56</sup> se acabaron considerando simplemente variantes del plegamiento GT-A ya que sus diferencias se limitan a diferentes ordenamientos dentro de la lámina  $\beta$  central característica de este plegamiento.<sup>45,53</sup> De esta manera, aunque la gran mayoría de las GT-A poseen el motivo DXD, existen ejemplos de GT-A que carecen de la firma DXD.

El primer plegamiento GT-B conocido fue también la primera estructura cristalográfica conseguida para una GT dependiente de nucleósidos-difosfato y corresponde a la bacteriófaga T4  $\beta$ -glucosiltransferasa (BGT).<sup>57</sup> Este plegamiento resultó ser análogo al mostrado por la glucógeno fosforilasa y ha sido observado en varias otras GTs también. El plegamiento GT-B, está formado por dos dominios tipo Rossmann separados por una fisura profunda en la cual se localiza el sitio activo. (Figura 7b) El sustrato dador se aloja en la cima de la ranura interaccionando únicamente con residuos del dominio C-terminal y como la ranura está alineada con varios residuos cargados positivamente, el sustrato aceptor (una doble hebra de ADN) se puede alojar en esta misma cavidad.<sup>54</sup> Es importante añadir que también existen variantes de este plegamiento como por ejemplo, la enzima  $\alpha$ 1,6-fucosiltransferasa de mamífero, cuyo dominio catalítico es muy parecido al plegamiento GT-B.<sup>58</sup>

Además de los plegamientos ya mencionados, en el año 2003 fue sugerido un tercer tipo de plegamiento para las GTs, el denominado GT-C (Figura 7c). Mediante análisis iterativos de secuencias usando programas como BLAST,<sup>41</sup> se predijo que 8 familias de GTs (GT22, GT39, GT48, GT50, GT53, GT57, GT58 y GT59) podrían adoptar el plegamiento GT-C. Posteriormente, las familias GT66, GT83, GT85 y GT86 fueron también predichas para adoptar este plegamiento por la base de datos CAZy.<sup>45</sup> El plegamiento GT-C se relaciona con grandes proteínas integrales de membrana ubicadas en el retículo endoplasmático, exhibiendo de 8 a 13 hélices transmembrana y un sitio activo localizado en la región de un lazo de gran longitud. Las familias para las que se ha predicho adoptar este último tipo de plegamiento usan un lípido-fosfato como complejo activado dador de azúcar, con la excepción de los miembros de las familias GT48 y GT53 que usan UDP-glucosa y UDP-L-arabinosa, respectivamente. Como se mencionó anteriormente, en el año 2007 se determinó la estructura del dominio C-terminal soluble de *Pyrococcus furiosus oligosacaryltransferasa* STT3 asignada a la familia GT66.<sup>40</sup> Esta estructura reveló que la secuencia de aminoácidos en la cual se basó el alineamiento no incluye al lazo en el que se predijo que estaría el sitio activo, por consiguiente no ha podido ser demostrada ninguna actividad catalítica. De esto se deduce que la mayor similitud de estas enzimas podría no estar en el dominio catalítico sino en la región transmembrana; en otras palabras, no queda claro que la predicción de la topología GT-C pueda ser asignada con éxito a nuevas secuencias de GTs y lo único que indica de manera inequívoca, es la presencia de un gran componente trans-membrana.<sup>45</sup>



**Figura 8.** Representación de la estructura tridimensional de diferentes GTs. Cada plegamiento es representativo de una superfamilia y es conocido como: (a) GT-A código PDB 4dmw GT44, (b) GT-B código PDB 1jg7 GT63 y (c) GT-C código PDB 2zag.

En resumen, el conjunto de información estructural y bioinformática acumulada dan información básicamente de tres cosas:

- i. Superfamilia GT-A: es ampliada debido a que se encontró que las GTs dependientes de un azúcar nucleósido-difosfato y las nucleótido-transferasas comparten rasgos de su estructura tridimensional y forman una familia monofilética (descienden de una única secuencia o grupo de secuencias progenitoras que no comparten con ningún otro grupo).
- ii. Superfamilia GT-B: a pesar de que muestran dos dominios tipo Rossmann y plegamientos tridimensionales parecidos, no se encontraron semejanzas de secuencia con la superfamilia GT-A ni viceversa. Esto apunta a que estas enzimas han evolucionado independientemente (convergencia) o que su último ancestro común, si es que existió, es extremadamente antiguo.
- iii. Superfamilia GT-C: ésta relaciona 8 familias CAZy de GTs. La arquitectura predicha para el plegamiento GT-C es observada en grandes proteínas hidrofóbicas integrales de membrana, localizadas en el retículo endoplasmático sin embargo, la topología GT-C podría no ser asignada con éxito a nuevas secuencias de GTs.<sup>41,45,59-61</sup>

Futuros análisis de las familias clasificadas como GTs a las que aún no se les ha podido asignar ningún tipo de plegamiento por análisis de secuencia, deberían poder aportar nueva información del origen evolutivo de las GTs.



Varias GTs dependientes de nucleótidos-azúcar, que son las enzimas que se estudian en esta tesis, y que pueden mostrar un plegamiento GT-A o GT-B poseen al menos un lazo flexible que tiene un importante papel en la catálisis. Debido al alto grado de flexibilidad de este lazo, en la mayoría de los casos no es posible trazarlo ni en la apo ni en la holo enzima pero se encuentra siempre en la vecindad del sitio de unión del nucleótido-azúcar.<sup>38</sup> Un caso especial es la enzima  $\beta$ 1,4-galactosiltransferasa T1 ( $\beta$ 4Gal-T1, GT-A inversora),<sup>62-64</sup> en la que casi todos los residuos que forman los dos lazos flexibles han podido ser trazados en ambas formas, la apo y el complejo enzima-nucleótido-azúcar más metal. En la forma apo los residuos del sitio activo de ambos lazos están lejos de la cavidad activa mientras que en la holo los residuos se acercan al nucleótido formando interacciones polares, electrostáticas o hidrofóbicas (CH- $\pi$  stacking); esto da origen a dos tipos de conformaciones, la abierta y la cerrada, respectivamente. Este cambio conformacional inducido por la unión del sustrato ha sido observado para otras enzimas como por ejemplo: la  $\alpha$ 1,3-galactosiltransferasa ( $\alpha$ 3GT, GT-A retenedora),<sup>65,66</sup> la  $\alpha$ 1,4-galactosiltransferasa (LgtC, GT-A retenedora),<sup>67</sup> la enzima N-acetilglucosaminiltransferasa I (GnT I, GT-A inversora),<sup>68</sup> la enzima polipéptido N-Acetilgalactosaminiltransferasa-2 (ppGalNAcT-2, GT-A retenedora) trabajo en el cual se acuñó el término "conformación activa" para la conformación cerrada e "inactiva" para la abierta,<sup>69</sup> y también dos de las enzimas que se estudian en esta tesis,  $\alpha$ 1,4-N-Acetilhexosaminiltransferase que puede transferir GalNAc y GlcNAc (EXTL2, GT-A retenedora)<sup>70</sup> y la Glucosil-3-fosfoglicerato sintasa (GpgS, GT-A retenedora).<sup>71</sup> El reordenamiento de este lazo inducido por la unión del sustrato dador y ubicado en el extremo C-terminal de la secuencia, crea una especie de tapa sobre el sustrato dador facilitando la formación de interacciones entre el UDP y diferentes residuos; esta nueva conformación cerrada crea una cavidad que servirá para unir el sustrato aceptor.<sup>36,42,65</sup> Adicionalmente, posteriores estudios estructurales, calorimétricos y de docking mostraron por una parte que el sustrato dador adquiere una conformación distorsionada en la enzima con respecto a la que presenta en solución acuosa; por otra parte, los complejos enzima-sustrato se forman de una manera ordenada, asociándose primero el sustrato dador y luego el aceptor ya que el estado abierto de la enzima o conformación inactiva, resulta tener muy poca afinidad por el sustrato aceptor. Se sugiere que esta unión ordenada de los sustratos se relaciona con un mecanismo de prevención de entrada de agua para evitar la hidrólisis del UDP-azúcar. Así, la conformación distorsionada del nucleótido azúcar al colocarse en el sitio activo podría ser crucial para conseguir la alta eficiencia de la reacción catalizada.<sup>38</sup>

Cabe decir que no existe correlación entre el plegamiento adoptado y la estereoquímica de la reacción catalizada, de modo que existen GTs-inv y GTs-ret en los diferentes plegamientos. Es interesante destacar también, que incluso presentando poca diversidad en el tipo de plegamiento, las GTs muestran una gran variabilidad estructural en el dominio al que se une el sustrato aceptor, lo que da lugar al amplio abanico de sustratos aceptores utilizados por estas enzimas, manteniendo cada GT una alta especificidad para dicho sustrato.



antibióticos de tipo vacomicina.<sup>97</sup> Adicionalmente, en las GT-B también aparecen varias enzimas que usan como base catalítica un Asp o Glu, por ejemplo: la enzima BGT, la primera GT para la que fue resuelta una estructura por difracción de rayos X y el único miembro de la familia GT63,<sup>57,98,99</sup> u otras enzimas como WaaC de la familia GT9,<sup>100</sup> la  $\alpha$ 1,3-fucosiltransferasa de *Helicobacter pylori* (FucT) de la familia GT10<sup>101</sup> o la  $\alpha$ 1,3-fucosiltransferasa V (FucT V).<sup>102</sup> Curiosamente, en el caso de la  $\alpha$ 1,6-fucosiltransferasa de mamífero (FUT8) de la familia GT23, para la cual no se había podido identificar un residuo que pudiera actuar como base catalítica,<sup>58,103</sup> se propuso que podría usar como base el mismo nucleótido que es parte del sustrato dador (grupo saliente). Además, hay otra O-fucosiltransferasa que utiliza como aceptor Ser/Thr, POFUT1 de la familia GT65, para la cual también se ha propuesto eso.<sup>104</sup> Este rol catalítico para el nucleótido fue inicialmente sugerido para explicar la estereoquímica del producto sintetizado por las GT-ret y surgió como resultado de la examinación de las primeras estructuras cristalográficas: LgtC<sup>67</sup> y OtsA.<sup>105,106</sup> Interesantemente, Vetting *et. al.* basándose en la estructura y modelización del sustrato de MshA, una GT-ret de plegamiento GT-B implicada en la síntesis del micotiol, utilizó el término “catálisis asistida por sustratos” para referirse a la abstracción del protón del sustrato aceptor por el nucleótido que forma parte del dador lo cual por cierto, también es una forma de estabilización del grupo saliente que es el tema que se comenta a continuación.<sup>42</sup>

Respecto a la estrategia mecánica para estabilizar el nucleótido liberado luego de la disociación nucleótido-azúcar, casi todas las GT-A se ayudan del catión metálico divalente con algunas excepciones ya mencionadas anteriormente: la  $\beta$ -1,6-GlcNAc transferasa C2GnT-L que posee una Arg378 y una Lys401 para estabiliza al grupo saliente,<sup>44</sup> y la sialiltransferasa CstII que posee dos Tyr interaccionando con el grupo saliente a través de sus grupos hidroxilos.<sup>43</sup>

### 1.3.3.2 Mecanismo de Retención

Con la resolución de las dos primeras estructuras cristalográficas de GTs-ret, LgtC<sup>67</sup> (GT-A) y OtsA<sup>105,106</sup> (GT-B), se pudo verificar que no había un Nuc bien posicionado en la cara  $\beta$  del azúcar que indicará que el mecanismo de doble desplazamiento fuera posible. LgtC es una  $\alpha$ 4-galactosiltransferasa de *Neisseria Meningitidis* que cataliza la transferencia de  $\alpha$ -Galactosa ( $\alpha$ Gal) desde UDP-Gal hasta la lactosa terminal de un Lipooligosacarido (LOS). La estructura obtenida por difracción de rayos X contiene un análogo del sustrato dador, UDP-2FGal, un análogo del sustrato aceptor, 4'-deoxilactosa, y el cofactor metálico Mn<sup>2+</sup>. Al examinar la cara  $\beta$  del azúcar, sólo se encuentra el oxígeno de la cadena lateral de Gln189 posicionada a 3.5Å del carbono anomérico. Medidas cinéticas sobre el mutante Q189A de LgtC indicaron que el valor del  $k_{cat}$  corresponde a un 3% del  $k_{cat}$  de la enzima nativa, lo cual es considerado una actividad residual relativamente alta teniendo en cuenta que el residuo es esencial para la catálisis. Las mutaciones del residuo nucleofílico en las GHs-ret redujeron la actividad residual de la mutante a casi cero, indicando que el residuo es esencial para la catálisis.<sup>28</sup> En vista de este resultado, se comenzó a considerar la posibilidad de un mecanismo alternativo tipo-S<sub>N</sub>i. En el caso de OtsA de *Escherichia coli*, una de las estructuras cristalográficas contiene UDP y el sustrato aceptor glucosa-6-fosfato (Glc-6-P)<sup>105</sup>, mientras que las otras dos contienen el sustrato dador UDP-Glc o UDP-

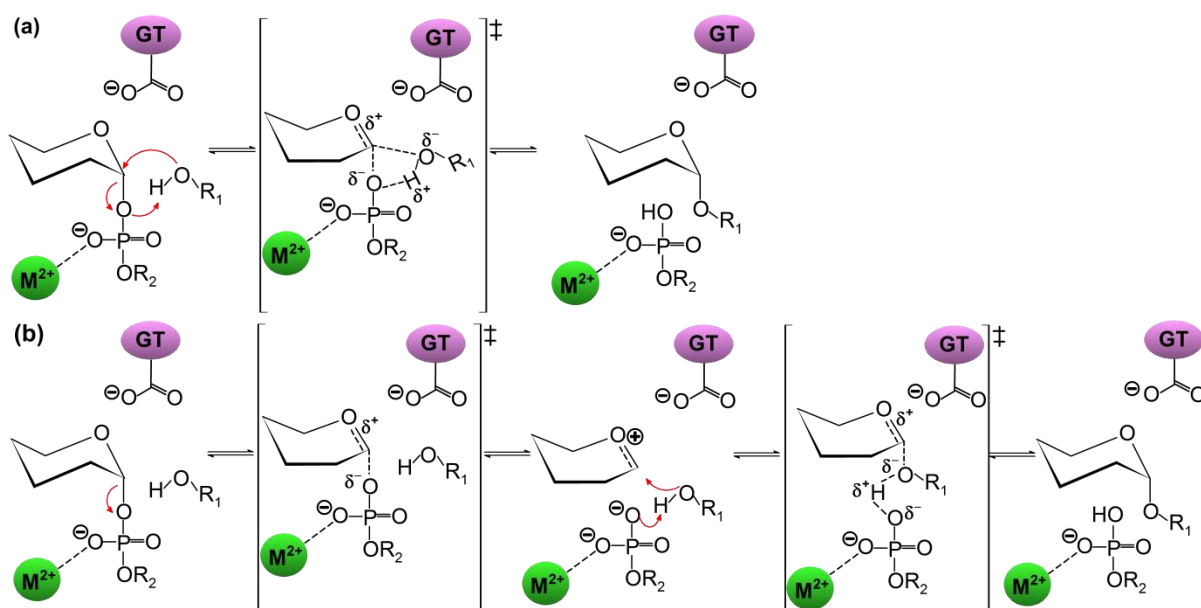
2FGlc.<sup>106</sup> Nuevamente, no se observó un potencial nucleófilo bien posicionado para la formación del CGE por lo que el mecanismo alternativo tipo-S<sub>N</sub>i volvió a surgir como una posibilidad. Cabe destacar que más tarde fue sintetizada una segunda mutante para LgtC, Q189E. Al igual que la mutante Q189A, la  $k_{cat}$  se redujo al 3% del valor correspondiente a la enzima nativa y la  $K_m$  de UDP-Gal no cambió significativamente.<sup>107</sup> Hasta aquí, el experimento resulta inconcluyente como los anteriores pero el curioso hecho de que se identificara un CGE formado con Asp190, que en la enzima nativa está a 8.9 Å del carbono anomérico, resultó bastante inquietante pues además de mostrar la gran plasticidad de estas enzimas, encendió la duda acerca de si es posible o no un mecanismo de doble desplazamiento.

Los siguientes complejos binarios de GTs-ret que se fueron cristalizando carecen de un residuo nucleofílico bien posicionado para el ataque, o bien poseen una cadena lateral a distancia apropiada pero el residuo no es buen nucleófilo. Entre algunos ejemplos que vale la pena mencionar está la enzima glucogenina<sup>108</sup> (GYG) de la familia GT8 (como LgtC) y de plegamiento GT-A que participa en la formación de nuestro polímero almacenador de energía por excelencia, el glucógeno; esta enzima tiene un Asp163 localizado a 6.1 Å del carbono anomérico; las enzimas UDP GalNAc:polipéptido  $\alpha$ -N-acetilgalactosaminiltransferas (ppGalNAcTs) de la familia GT27 y plegamiento GT-A, muestran una Gln346 a 6.1 Å del carbono anomérico en ppGalNAcT-10<sup>109</sup> y una Asn335 en ppGalNAcT-2,<sup>110</sup> aunque no es posible indicar una distancia respecto al carbono anomérico pues el ligando no contiene el azúcar. Cabe mencionar que posteriores simulaciones de MD y QM/MM sobre ppGalNAcT-2 han señalado que no hay un residuo nucleofílico bien posicionado para la formación de un CGE, pero que el sustrato aceptor Thr7 está bien orientado para el ataque nucleofílico por lo que se trataría de un mecanismo de ataque frontal.<sup>69,111,112</sup> (Las últimas 2 refs se discutirán en la siguiente sección) En el dominio catalítico de la toxina B de *Clostridium difficile* (ToxB) de la familia GT44 y plegamiento GT-A, hay una Asn384 a 3.5 Å del carbono anomérico;<sup>113</sup> en la manosilglicerato sintasa (MGS) de *Rhodothermus marinus* de la familia GT78 y plegamiento GT-A, hay una Leu163 a 4.3 Å del carbono anomérico;<sup>114</sup> la Kre2p/Mnt1p de la familia GT15 y plegamiento GT-A no muestra un residuo nucleofílico cerca del carbono anomérico<sup>115</sup> y se ha reportado que podría seguir un mecanismo tipo-S<sub>N</sub>i.<sup>116</sup> (La última ref se discutirá en la siguiente sección) La enzima EXTL2 de la familia GT64 y de plegamiento GT-A, muestra un Asp246 y una Arg293 a 5.3 Å y a sólo 4.0 Å del carbono anomérico, respectivamente. Aunque la Arg293 está mejor posicionada para el ataque que el Asp246, la probabilidad de que una Arg actúe como nucleófilo es bajísima considerando su naturaleza electrónica.<sup>70</sup> EXTL2 es una de las enzimas que se investigan en esta tesis y la primera en la que se encuentra una pareja Asp/Arg en la cara  $\beta$  del azúcar.

Una comparación de las estructuras representativas de GTs-ret de plegamiento GT-A ha mostrado características conservadas en la región que rodea al nucleótido y al sustrato aceptor, cara  $\alpha$  del azúcar. La situación de la cara  $\beta$  del azúcar es diferente, cuya única característica estructural conservada es una cadena lateral carboxilato de un Asp o un Glu (Asp/Glu) a distancia de interacción de los hidroxilos 4 y 6 del azúcar, y de un residuo con carga positiva Arg/Lys.<sup>45</sup> Además, Arg/Lys interacciona con el O3 del azúcar y con la cadena lateral del primer Asp del motivo DXD, mientras que el último coordina al metal.<sup>45</sup> Es interesante mencionar algunos ejemplos correspondientes al Asp/Glu conservado en la cara  $\beta$  del azúcar de algunas GTs-ret: en GYG (GT8) el Asp conservado corresponde a Asp163, y se encuentra a 6.1 Å del carbono anomérico; en  $\alpha$ 3GalT (GT6) se trata del Glu316, que



Dicho esto, el mecanismo tipo  $S_Ni$ , actualmente también conocido como mecanismo de ataque frontal, requiere tener al nucleófilo y al grupo saliente en la misma región o cara del azúcar y que el grupo fosfato que se desprende del sustrato dador actúe como base catalítica estableciendo esta interacción previamente a la ruptura del enlace nucleótido-azúcar (Figura 11). Existen antecedentes del grupo fosfato actuando como base y tal rol le fue asignado en el mecanismo de la farnesil difosfato sintasa (FPP por su sigla en inglés) proveniente de la *Escherichia Coli*<sup>119</sup> y en el de la *Trypanosoma cruzi*.<sup>120</sup>



**Figura 11.** Mecanismo de ataque nucleofílico frontal o tipo- $S_Ni$  en el que puede existir o no un residuo nucleofílico en la cara  $\beta$  del azúcar; el metal puede ser reemplazado por aminoácidos como Lys, Arg o hélices dipolares dependiendo de si trata de una GT-A o una GT-B, respectivamente. El mecanismo resulta en la retención de la configuración del carbono anomérico y podría ocurrir (a) en una única etapa en la que la ruptura nucleótido-azúcar y la deprotonación-ataque del aceptor ocurran de manera asincrónica, involucrando un TS tipo-oxocarbenio (b) vía un intermedio IP de corta vida y alta energía al que seguiría el ataque del nucleófilo ( $D_N^i A_{Niss}$ ). Aunque la abreviación -ss" significa -solvent separate" IP, aquí se usa para indicar que el nucleótido y el azúcar están lo suficientemente distanciados el uno del otro para permitir la entrada del nucleófilo; si se tratara de un IP íntimo no habría espacio para el nucleófilo que se acerca. Abreviaciones:  $R_1$  representa al sustrato aceptor y puede ser un azúcar, una proteína, antibiótico o cualquier otra aglicona;  $R_2$  un nucleósido o un nucleósido-fosfato.

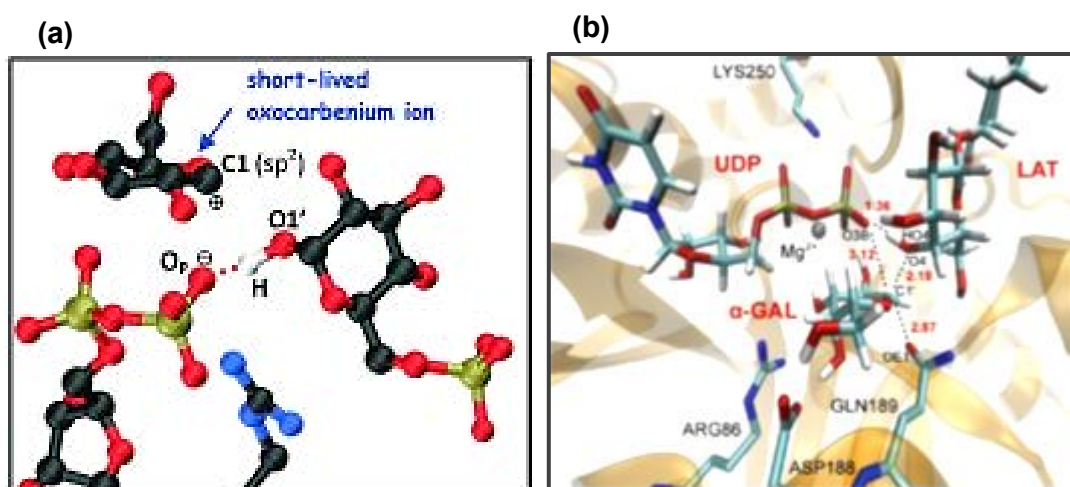
En 2004, Tvaroška *et. al.*<sup>121</sup> presentaron un estudio computacional del mecanismo de LgtC en él utilizaban un modelo reducido. En este trabajo se evaluó el mecanismo tipo- $S_Ni$  con UDP actuando como base catalítica y por lo tanto activando al aceptor. Aunque la primera estimación de la barrera de energía resultó ser de aproximadamente 31 kcal/mol, la inclusión de un medio dieléctrico continuo y de correcciones energéticas condujeron a una estimación final de entre 10-24 kcal/mol. Este estudio pone en evidencia las limitaciones de simular el sitio activo de estas enzimas sin considerar el ambiente proteico aun describiendo la región a nivel cuántico.

## 1.4 Estado del Debate en el Momento de Iniciar la Tesis

A continuación se discutirán en detalle los estudios computacionales del mecanismo de retención que existían al iniciar y desarrollar esta investigación doctoral.

Cuando esta investigación doctoral se inició a finales del año 2013, a nivel computacional se habían realizado estudios antes mencionados que consideraban pequeños modelos del sitio activo que no tenían en cuenta el ambiente enzimático<sup>121,122</sup> y también comenzaban a aparecer los primeros trabajos basados en modelos completos enzima-sustratos que describen explícitamente el ambiente enzimático y el solvente a nivel QMMM.<sup>123-126</sup> Estos trabajos, junto con los nuevos resultados experimentales más relevantes que se reportaron, se resumen a continuación.

En el año 2011, Ardèvol *et. al.* presentaron el primer trabajo en el que fue estudiado un modelo completo de una GT usando simulaciones de metadinámica QM/MM.<sup>123</sup> Se trataba de OtsA, una GT-B que cataliza la formación de uniones glicosídicas  $\alpha$ 1-1 entre UDP-Glc y glucosa-6-fosfato (Glc-6P por su sigla en inglés) para formar trehalosa-6-fosfato. El proceso encontrado en este trabajo fue descrito como un mecanismo de ataque frontal que podría ser interpretado como un mecanismo en dos etapas vía un intermedio iónico de muy corta vida y alta energía: UDP—oxocarbenio (Figura 12a).



**Figura 12.** Representación de la especie oxocarbenio identificada en diferentes simulaciones del mecanismo de ataque frontal para (a) OtsA: par iónico (IP) UDP—oxocarbenio<sup>123</sup> y (b) LgtC: TS oxocarbenio.<sup>125</sup>

Sin embargo, la estabilidad del IP con respecto a los máximos de energía contiguos es de  $\pm 2$  kcal/mol, indicativo de una superficie de energía libre bastante plana. También se concluyó que la interacción UDP(O<sub>P</sub>)-Glc-6F(HO1') es muy importante ya que además de actuar como base catalítica activando la nucleofilicidad del aceptor, estabiliza la salida de UDP. Ésta interacción tiene un valor de 1.87 Å en el reactivo y de 1.38 Å en el intermedio (Figura 12a). Las distancias UDP(O<sub>P</sub>)-C1 y Glc-6F(O1')-C1 en el IP son de 2.84 Å y 2.79 Å, respectivamente. El producto o ataque del aceptor, se produce de manera simultánea a la transferencia protónica de Glc-6P(HO1') hasta UDP(O<sub>P</sub>), y la barrera de energía libre calculada es de 23.0 kcal/mol.

Entre 2010-2011 se publicaron trabajos experimentales muy importantes, dos de ellos están relacionados con OtsA mientras que el otro utilizó también una enzima GT-ret de plegamiento GT-B, MshA. Notablemente, los trabajos de OtsA tienen éxito en comprobar que la interacción entre el nucleósido difosfato -NDP por su sigla en inglés- con el hidroxilo atacante del aceptor es necesaria y está presente en el TS-mimic de OtsA (un TS-mimic es una aproximación a las interacciones y estructura del estado de transición usando una imitación del producto natural que en este caso es, validoxilamina).<sup>127</sup> El siguiente estudio confirmó lo recién descrito a través de medidas de efecto cinético isotópico (KIE) sobre el O del grupo saliente, cuyos valores indican claramente la existencia del enlace de hidrógeno entre el hidroxilo nucleofílico (Nuc) y el grupo saliente.<sup>128</sup> En el caso de MshA, los experimentos cinéticos de estado estacionario señalan que un análogo del sustrato dador, UDP-5FGlcNAc, inhibe eficazmente a la enzima con una constante de Michaelis 150 veces más pequeña que la calculada para el sustrato natural,  $1.4 \pm 0.2 \mu\text{M}$  vs.  $210 \mu\text{M}$ , respectivamente. Estos resultados confirman que hay un desarrollo de carga positiva en el azúcar durante la reacción de transferencia con retención de la configuración.<sup>129</sup>

Muy poco tiempo después se publicó el trabajo de nuestro grupo (tesis de H. Gómez) sobre LgtC, que fue el primer estudio que presentó un modelo completo de una GT-ret con plegamiento GT-A. Se utilizaron métodos QMMM para calcular perfiles de energía potencial asociados a la reacción que cataliza esta enzima,<sup>125</sup> que es la transferencia de  $\alpha$ -galactosa desde UDP (UDP-Gal) hasta su aceptor lactosa (LAT) formando una unión glicosídica  $\alpha 1-4$ . Esta GT-ret difiere de OtsA en su tipo de plegamiento y en que muestra un motivo DXD junto a un metal divalente necesario para la catálisis. En este estudio fue modelado el mecanismo de ataque frontal y también el mecanismo de doble desplazamiento. Para el primero se encontró que la sustitución nucleofílica ocurre en una única etapa con una barrera de energía de 14.6 kcal/mol (a nivel QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM)) sin pasar por un intermedio UDP—oxocarbenio (Figura 12b). En esta especie, las distancias UDP(O3B)-Gal(C1'), LAT(O4')-Gal(C1') y UDP(O3B)-LAT(HO4) son de 3.12 Å, 2.18 Å y 1.36 Å, respectivamente. Otras importantes conclusiones de este trabajo indicaron que el mecanismo de ataque frontal es altamente disociativo, el enlace UDP-Gal se rompe temprano en la reacción mientras el aceptor se aproxima lentamente al carbono anomérico de la galactosa (C1'). Desde el inicio de la reacción está presente la interacción entre UDP(O3B)-LAT(HO4)= 1.67Å por lo que el aceptor ya está bien posicionado para que UDP lo active y para facilitar la ruptura del enlace UDP-Gal. Por último, la transferencia de LAT(HO4) hasta el grupo UDP ocurre simultáneamente con la formación del enlace glicosídico pero asincrónicamente a la ruptura UDP(O3B)-Gal(C1'). Además de la importante interacción UDP(O3B)-LAT(HO4) que es necesaria para la activación de nucleófilo, existen otras interacciones entre sustratos que se señalaron como claves en este mecanismo: UDP(O3B)-LAT(HO3) y UDP(O1B)-Gal(HO2'). Ambos enlaces hidrógenos facilitan la salida de UDP y son parte de la estrategia de catálisis asistida por sustratos.

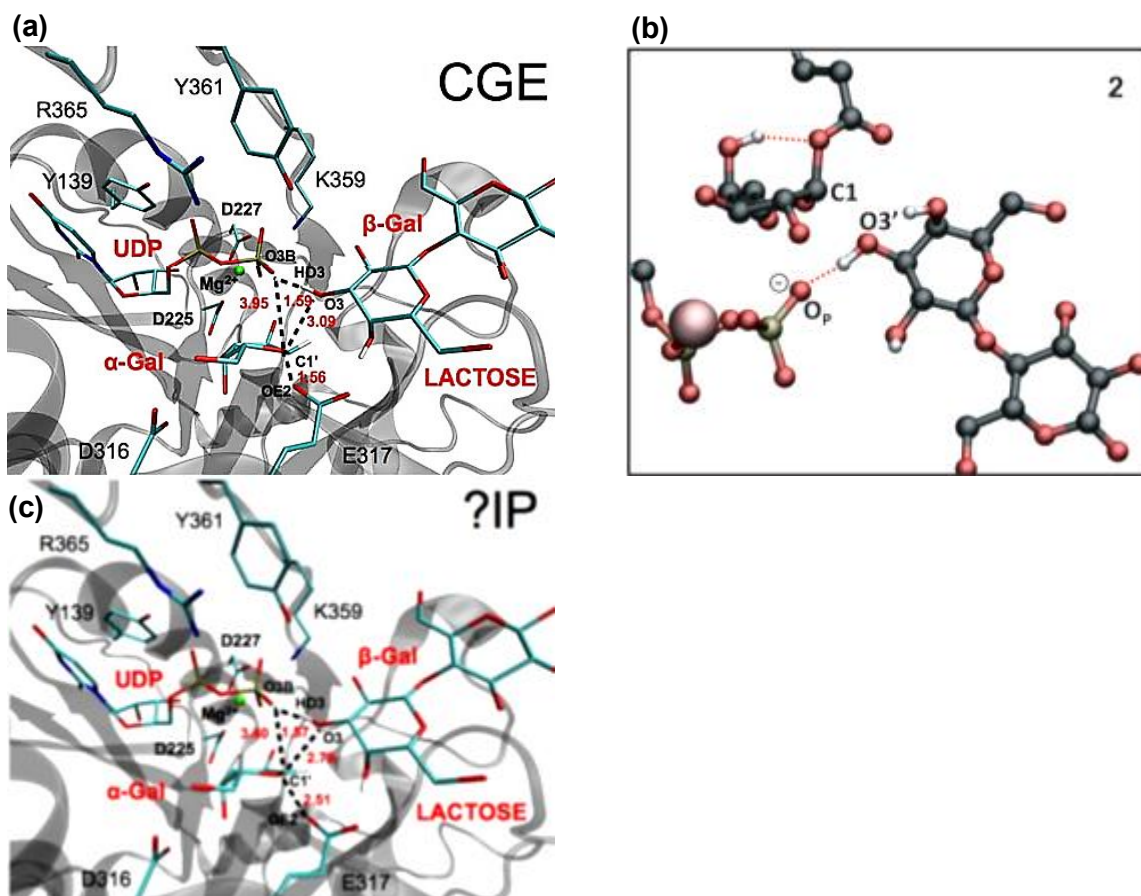
Por otra parte, todos los intentos por encontrar un CGE fallaron. Aunque la coordenada de reacción que se usó para simular la reacción consiguió aproximar al posible nucleófilo (Gln189) hasta quedar a 1.58Å del carbono anomérico, la optimización de esta estructura colapsó al reactante.

En resumen, aunque OtsA y LgtC poseen diferentes plegamientos, diferentes sustratos y distintas arquitecturas estructurales en el sitio activo, se encontró que ambas siguen un mecanismo de ataque frontal que involucra la formación de especies oxocarbenio. En el



primer caso, la especie oxocarbenio fue caracterizada como un intermedio de alta energía y de muy corta vida en el que el UDP está a 2.85 Å del carbono anomérico mientras que en LgtC, la especie oxocarbenio es caracterizada como un TS exhibiendo por consiguiente, el valor más alto de energía a lo largo del perfil con el UDP ubicado a 3.12 Å del carbono anomérico. En este mecanismo, la salida de UDP ocurre de manera asincrónica al ataque del aceptor, algo que Lairson *et. al* ya habían anticipado en su trabajo de analogías mecanísticas entre enzimas modificadoras de carbohidratos.<sup>130</sup> Un mecanismo que pase a través de un intermedio CGE no fue abordado en el caso de OtsA debido a que esta enzima carece de un residuo en el sitio activo que pueda llevar a cabo esta función. Aunque en LgtC sí existe un residuo nucleofílico que podría formar tal intermedio, la simulación de esta reacción no condujo a tal especie; es posible que la interacción estable que forma Gln189 con LAT (Gln189(NH1)-LAT(O6)), reduzca la habilidad nucleofílica de este residuo. Cabe mencionar que experimentos adicionales realizados por Gómez *et. al.*, les permitieron concluir que Gln189 estabiliza a la especie oxocarbenio a través de la interacción Gln189(OE1)-Gal(C1').

Más tarde, la enzima  $\alpha$ 3-GalT perteneciente a la familia GT6 y de plegamiento GT-A fue estudiada también en nuestro grupo dentro de la tesis de H. Gómez.<sup>124,131</sup> Esta enzima cataliza la transferencia de galactosa desde UDP-Gal hasta LAT formando el enlace glicosídico  $\alpha$ 1-3 y además, posee un residuo nucleofílico bien posicionado, el Glu317, que podría llevar a cabo la formación del CGE.



**Figura 13.** Representación del sitio activo de  $\alpha 3\text{GalT}$  obtenido desde distintas simulaciones del mecanismo de ataque frontal: (a) el intermedio CGE,<sup>124</sup> (b) el intermedio CGE<sup>126</sup> y (c) el intermedio oxocarbenio identificado como un IP UDP—oxocarbenio.<sup>124</sup>

La formación del CGE ocurre en la primera etapa del mecanismo a través de un TS1 que representa una barrera energética de 15.1 kcal/mol (nótese que esta barrera en realidad corresponde al promedio de varias barreras que se obtuvieron a partir de varias conformaciones del complejo de Michaelis provenientes de una dinámica molecular).<sup>124</sup> El TS1 muestra las siguientes distancias:  $\text{UDP}(\text{O3B})\text{-Gal}(\text{C1}') = 3.18 \text{ \AA}$ ,  $\text{LAT}(\text{O3})\text{-Gal}(\text{C1}') = 2.85 \text{ \AA}$  y  $\text{Glu317}(\text{OE2})\text{-Gal}(\text{C1}') = 2.60 \text{ \AA}$ ; la distancia entre  $\text{LAT}(\text{HO3})\text{-UDP}(\text{O3B})$  es de 1.58  $\text{ \AA}$ , 2.48  $\text{ \AA}$  más cercana que en el reactante lo cual nuevamente muestra el importante papel de esta interacción durante la ruptura del enlace UDP-Gal. Resulta evidente que en el TS1,  $\text{LAT}(\text{O3})$  aún no está lo suficientemente cerca de  $\text{C1}'$  para atacar a pesar de que  $\text{LAT}(\text{HO3})$  ya se ha orientado hacia UDP. En el CGE (Figura 13a), la distancia  $\text{UDP}(\text{O3B})\text{-Gal}(\text{C1}') = 3.95 \text{ \AA}$ ,  $\text{LAT}(\text{O3})\text{-Gal}(\text{C1}') = 3.09 \text{ \AA}$  y  $\text{Glu317}(\text{OE2})\text{-Gal}(\text{C1}') = 1.56 \text{ \AA}$ . Mientras tanto, la distancia  $\text{LAT}(\text{HO3})\text{-UDP}(\text{O3B})$  permanece casi invariante en 1.59  $\text{ \AA}$ . La subsecuente ruptura del CGE para formar el producto ocurrió con una barrera de energía de 15.9 kcal/mol.

En este trabajo también se estudió el mecanismo de ataque frontal, definiendo una coordenada de reacción diferente que lleva a un TS1 con una barrera energética de 16.2 kcal/mol. En el TS1, las distancias que separan a  $\text{UDP}(\text{O3B})$ ,  $\text{Glu317}(\text{OE2})$  y  $\text{LAT}(\text{O3})$  del carbono anomérico son de 2.48  $\text{ \AA}$ , 3.16  $\text{ \AA}$  y 2.87  $\text{ \AA}$ , respectivamente; mientras la distancia  $\text{LAT}(\text{HO3})\text{-UDP}(\text{HO3})$  es de 1.63  $\text{ \AA}$ . UDP está aún cerca del  $\text{C1}'$  por lo que serían

necesarios algunas reorganizaciones para conseguir que el aceptor se posicione correctamente. En orden a conseguir estas reorganizaciones, parece ser que la estrategia de la enzima es formar una especie intermedia que en este mecanismo sería un IP muy poco profundo o de alta energía y probablemente de muy corta vida. La distancia que separa a UDP(O3B), a Glu317(OE2) y a LAT(O3) del carbono anomérico en esta especie es de 3.40 Å, 2.51 Å y 2.79 Å, respectivamente; mientras la distancia LAT(HO3)-UDP(O3B) ha decrecido ligeramente hasta 1.57 Å. (Figura 13c) Finalmente, el TS2 aparece con una barrera energética de 16.3 kcal/mol. La superficie de energía potencial es bastante plana lo cual dificulta la caracterización de las especies oxocarbenio que surgen luego de que el enlace UDP-Gal se rompe. Sin embargo, los resultados apuntan a que un intermedio de reacción es necesario en orden a lograr una adecuada orientación de LAT. Cabe destacar que aparte de la interacción UDP(O3B)-LAT(HO3), sólo hay una interacción más que podría contribuir a la catálisis asistida por sustratos; la interacción intra-sustrato dador UDP(O1B)-Gal(HO2') que por cierto, también fue observada en LgtC. En este trabajo también se estimaron las contribuciones electrostáticas que el sustrato aceptor LAT podría brindar en  $\alpha$ 3Gal-T y en LgtC, encontrándose que en  $\alpha$ 3Gal-T la estabilización es muy pobre y en LgtC es notablemente buena. Adicionalmente, se calculó esta misma estimación para los residuos nucleofílicos de estas enzimas, Glu317 y Gln189, encontrándose ahora que Glu317 estabiliza notablemente la salida de UDP mientras Gln189 contribuye discretamente. Basándose en estos resultados, se concluye que debido a la limitada estabilización proporcionada por LAT en la enzima  $\alpha$ 3GalT, es necesaria una especie adicional que facilite la ruptura nucleótido-azúcar. De esta manera, el rol de Glu317 en la cara  $\beta$  de Gal es estabilizar la especie oxocarbenio, estrategia que bautizaron como catálisis asistida por nucleófilo.<sup>124</sup> Es importante añadir que al igual que Gln189 interacciona con LAT en LgtC, en  $\alpha$ 3GalT existe la interacción Glu317(OE1)-LAT(HO4). Una de las preguntas que se intenta responder en esta tesis es: ¿la posición estructuralmente análoga que ocupa Glu317 y Gln189 está exclusivamente relacionada con la estabilización de la especie oxocarbenio o podría relacionarse también con el sustrato aceptor?, si es así ¿qué implicancias mecanísticas tendría esto?. Esta discusión se retomará más adelante.

En el mismo año, Rovira *et. al.* estudiaron también la enzima  $\alpha$ 3GalT simulando en un caso la ruptura del enlace UDP(O<sub>P</sub>)-C1 y la formación del enlace glicosídico LAT(O3)-C1, y en el segundo caso las dos distancias anteriores más la deprotonación UDP(O<sub>P</sub>)-LAT(O3'H). En ambas simulaciones la galactosa colapsó con el Glu317 formando el CGE. La formación del CGE ocurre con una barrera de 23.0 kcal/mol y la estructura asociada a esta barrera muestra distancias UDP(O<sub>P</sub>)-Gal(C1)= 2.22 Å, LAT(O3')-Gal(C1)= 2.50 Å y Glu317(O)-Gal(C1)=3.06 Å. Lamentablemente, la distancia LAT(O3'H)-UDP(O<sub>P</sub>) no pudo ser encontrada pero se verifica que la interacción sólo se formó en el intermedio CGE (Figura 13 b). De hecho, parece ser que la formación de este intermedio permite una reorganización que logra posicionar al LAT(O3'H) en una adecuada orientación con respecto a UDP para que éste pueda actuar como base catalítica. Después de que el enlace Glu317(O)-C1 se rompe, la galactosa se mueve hasta el aceptor formándose el nuevo enlace glicosídico con una barrera de energía libre de 13.0 kcal/mol. Resulta curioso que la simulación que incluyó entre sus variables a la distancia LAT(O3'H)-UDP(O<sub>P</sub>) haya también colapsado al CGE, aunque también es lógico que bajo una configuración en la que UDP no interacciona con LAT(O3'H) desde el principio, la formación del CGE sea la única manera de facilitar la salida de UDP.

Cabe mencionar que cuando esta tesis ya se encontraba en desarrollo, Bobovská *et. al.*<sup>132</sup> abordaron el estudio de otro miembro de la familia GT6, la enzima GTB, encontrando que ambos mecanismo son posibles, el de doble desplazamiento y el mecanismo de ataque frontal. También sugieren que el principal rol de Glu303 (análogo a Glu317) podría ser la estabilización de la especie oxocarbenio evitando así la hidrólisis, además de posicionar adecuadamente el aceptor para el ataque nucleofílico. Esto último ya había sido sugerido en algunos trabajos experimentales<sup>66,133</sup> y también en dos de los trabajos computacionales de nuestro grupo sobre  $\alpha$ 3GalT incluidos en la tesis de H. Gómez.<sup>124,131</sup>

Posteriores estudios computacionales sobre Kre2p/Mnt1p<sup>116</sup> y ppGalNAcT2,<sup>111,112,134</sup> han mostrado que estas enzimas siguen un mecanismo de ataque frontal o tipo-S<sub>N</sub>i que implica la formación de una especie intermedia oxocarbenio de alta energía.

En resumen, durante los últimos años el mecanismo de ataque frontal ha ido ganando aceptación gracias a los estudios experimentales y computacionales que lo apoyan. La posibilidad de un mecanismo de doble desplazamiento dentro de la familia GT6 no se puede descartar completamente pues por una parte, la evidencia experimental no resulta concluyente y por otro lado, los estudios computacionales han indicado que este camino es energéticamente accesible para miembros de la familia GT6 en las que la interacción nucleótido-NUC no se ha establecido en el complejo ternario a partir del cual se inicia el estudio de reactividad.

## 2 Objetivos

Los carbohidratos son conocidos por llevar a cabo tareas esenciales en importantes y diversos procesos biológicos. En los organismos vivos, la modificación organizada de los carbohidratos la llevan a cabo enzimas como las GHs, las fosforilasas y las GTs, lo cual da origen a la enorme diversidad de oligo y polisacáridos encontrados en seres vivos. Los mecanismos catalíticos que usan las GHs o las GTs-inv han sido estudiados y establecidos, sin embargo el mecanismo catalizado por las GTs-ret aún es un área en la que existen algunos interrogantes respecto a los detalles mecanísticos.

El mecanismo de reacción de las GTs que catalizan la transferencia del azúcar manteniendo la configuración del carbono anomérico es un área difícil de abordar desde un punto de vista experimental y computacional. En primer lugar, los estudios estructurales y mecanísticos requieren de cantidades suficientes de formas activas de estas enzimas, sin embargo la obtención y purificación de formas activas es bastante complicada ya que se trata de enzimas asociadas a membranas. Las técnicas experimentales que fueron sistemáticamente aplicadas con éxito sobre las GHs, no han conducido a resultados concluyentes en las GTs, por lo que no existe evidencia sistemática para un mecanismo general en las GTs-ret. Sin embargo, esto no significa que algunos experimentos cinéticos puedan funcionar en algunas enzimas, y en este contexto la propuesta de ataque frontal ha ido ganando credibilidad durante los últimos años. Adicionalmente, los análisis bioinformáticos muestran que no hay una arquitectura conservada en la cara  $\beta$  del azúcar en la que se esperaría encontrar un residuo nucleofílico para la formación del CGE, por lo que resulta desafiante asignar roles a los residuos de esta región. Creemos que un entendimiento detallado de esta región de la enzima es fundamental para completar la visión general del mecanismo catalítico GTs-ret. Debido a lo anterior, los interrogantes que se busca responder son: **1.** ¿hay un mecanismo catalítico general para la síntesis del enlace glicosídico con retención de la configuración catalizado por las GTs-ret?, **2.** ¿Por qué los residuos de la cara  $\beta$  del azúcar que se transfiere no son conservados en las GTs-ret? y **3.** ¿Cuál es el rol de estos residuos en el caso de estar presentes?

Para responder estas preguntas se plantean los siguientes objetivos generales y específicos.

## 2.1 Objetivos Generales

Dar nueva información acerca de las estrategias mecanísticas usadas por las GTs-ret para sintetizar el enlace glicosídico con retención de la configuración del carbono anomérico.

## 2.2 Objetivos Específicos

1. Modelizar el complejo de Michaelis de la enzima en presencia del sustrato aceptor y dador.
2. Construir mutantes *in silico* a partir de los complejos de Michaelis modelizados.
3. Simular el mecanismo de ataque frontal y el mecanismo de formación del CGE usando los diversos modelos obtenidos (enzimas nativas y mutantes), y caracterizar los puntos estacionarios del mecanismo.
4. Identificar las interacciones y residuos que modulan las barreras energéticas en los diferentes modelos estudiados.
5. Analizar los factores que contribuyen a la catálisis, haciendo un enfoque especial en los residuos que ocupan la cara  $\beta$  del azúcar.

# 3 Fundamentos y Métodos Teóricos

## 3.1 Mecánica Molecular

Cuando se necesita obtener la superficie de energía potencial a partir de la energía electrónica para una determinada configuración nuclear, uno de los problemas con los que se encuentran los químicos computacionales es seleccionar el nivel de teoría adecuado para el problema que se desea estudiar.<sup>135</sup> Además, si se trata de sistemas muy grandes como enzimas o proteínas (> 30000 átomos), los métodos cuánticos son prohibitivos en tiempo y recursos computacionales, incluso cuando algunos electrones son omitidos explícitamente como en las aproximaciones semiempíricas.<sup>136</sup>

Los métodos de Campo de Fuerzas (FF derivado del inglés Force Field), abordan el cálculo de la energía electrónica escribiéndola como una función paramétrica de las coordenadas nucleares y ajustando estos parámetros a datos experimentales o computacionales *ab initio*.<sup>137</sup> [Jensen, pp22-24] Dado que los electrones no son considerados como partículas individuales (tratamiento no explícito), toda la información del sistema debe ser conocida (longitud de enlaces, ángulos de enlace, etc) pues no se resolverá una ecuación electrónica. En el caso de que se estudie un fenómeno independiente del tiempo, el problema se reduce a calcular la energía de una determinada geometría (minimización). En un campo de fuerzas, las moléculas se conciben como esferas unidas por resortes, considerando distintos tamaños y longitudes de enlace, respectivamente.

El fundamento de los métodos FF, también conocidos como Mecánica Molecular (MM), se basa en la observación de que las moléculas tienden a estar compuestas de unidades que son estructuralmente similares en diferentes moléculas. Por ejemplo, la longitud del enlace C-H es aproximadamente constante en todas las moléculas, tomando valores entre 1.06 y 1.10 Å. Las vibraciones de estiramiento del enlace C-H también son similares, entre 2900 y 3300  $\text{cm}^{-1}$ , lo cual implica que las constantes de fuerza C-H también son comparables. Adicionalmente, si los enlaces C-H son subdivididos en distintos grupos, esto es: un carbono involucrado en un enlace simple, doble y triple, la variación dentro de cada uno se hace aún más pequeña. Este concepto de transferibilidad también se aplica a características energéticas. Al graficar los calores de formación de alcanos lineales, por ejemplo:  $\text{CH}_3(\text{CH}_2)_n\text{CH}_3$ , en función de la longitud de la cadena ( $n$ ), se produce una línea recta que indica que cada grupo  $\text{CH}_2$  contribuye esencialmente con la misma cantidad de energía.<sup>137</sup> Así, en los FF la idea de transferibilidad es implementada a través de los tipos de átomos, los cuales dependen del número atómico y el tipo de enlace químico en el que está involucrado. Entre los FF más utilizados en la modelización de proteínas, ácidos nucleicos, lípidos y carbohidratos se encuentra: AMBER,<sup>138,139</sup> CHARMM,<sup>140,141</sup> OPLS<sup>142,143</sup> y GROMOS.<sup>144,145</sup>

La energía del campo de fuerza (ecuación 1) se escribe como una suma de términos, cada uno de ellos describiendo la energía necesaria para distorsionar a la molécula en una manera específica. Esto es; energía de estiramiento, primer término de la ecuación 1; energía de deformación angular, segundo término de la ecuación 1; energía torsional, tercer término de la ecuación 1; energía de términos impropios, cuarto término de la ecuación 1; energía de interacciones no enlazantes, quinto término de la ecuación 1; y un segundo término para interacciones no enlazantes, el potencial de Coulomb. La colección de funciones que conforman la función de energía potencial y el conjunto de parámetros que ajustan cada una de estas funciones, constituyen un método FF. A continuación se dan algunas ideas básicas de las expresiones que se utilizan para calcular el valor de la función de energía potencial.

### 3.1.1 Función de Energía Potencial

La mecánica molecular (MM) trata al sistema como un conjunto de cargas puntuales con radio que interactúan entre ellas a través de potenciales que dependen de varios términos. Una expresión típica de la energía potencial se da a continuación:

$$U(\vec{R}) = \sum_{\text{enlaces}} \frac{k_b}{2} (b - b_o)^2 + \sum_{\text{ángulos}} \frac{k_\theta}{2} (\theta - \theta_o)^2 + \sum_{\text{torsión}} V_n [1 + \cos(n\chi - \delta)] + \sum_{\text{impropios}} \frac{k_{imp}}{2} (\varphi - \varphi_o)^2 + \sum_{\text{noenl}} \left( 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right) + \frac{q_i q_j}{4\pi\epsilon_o r_{ij}} \quad [3.1]$$

Como se observa en la expresión precedente, la función de energía potencial incluye una colección de funciones simples que dan cuenta de un conjunto mínimo de interacciones que pueden describir la estructura molecular. En esta expresión,  $b$  representa la distancia de enlace,  $\theta$  es el ángulo de enlace,  $\chi$  es el ángulo diedro,  $\varphi$  es el ángulo impropio;  $r_{ij}$  es la distancia entre los átomos  $i$  y  $j$ ;  $k_b$ ,  $b_o$ ,  $k_\theta$ ,  $\theta_o$ ,  $k_{imp}$ ,  $\varphi_o$ ,  $V_n$ ,  $n$  y  $\delta$  son los parámetros de un campo de fuerzas y dan cuenta de la constante de fuerza de enlace, distancia de enlace en equilibrio, constante de fuerza de ángulo de enlace, valor de equilibrio del ángulo de enlace, constante de fuerza del ángulo impropio, valor de equilibrio del ángulo impropio, constante de fuerza del ángulo de torsión, multiplicidad y fase del ángulo, respectivamente. Los parámetros no enlazantes incluyen las cargas atómicas  $q_i$ , y  $\zeta_{ij}$  que es la distancia finita a la que la interacción no enlazante entre las partículas neutras es nula. Por último,  $\epsilon_o$  y  $\epsilon_{ij}$  son la constante dieléctrica del vacío y la profundidad del pozo, respectivamente.<sup>146</sup>

Enlaces, ángulos y deformaciones fuera del plano (ángulos diedros e impropios), son descritos por funciones armónicas con la excepción de los ángulos diedros que son descritos por un término sinusoidal. Las interacciones no enlazantes se tratan con el potencial de Lennard-Jones (LJ) para describir la repulsión entre los átomos y las interacciones de dispersión, y con el potencial coulombico para tratar las interacciones electrostáticas.



### 3.1.2 Análisis de los Términos de la Función de Energía Potencial

- i. Estiramiento y contracción de la energía de enlace:  $\sum_{\text{enlaces}} \frac{k_b}{2} (b - b_o)^2$

Esta expresión analiza la energía asociada a la desviación de una distancia de referencia sobre todos los pares de átomos directamente enlazados. La elección más simple para modelar este término de la función potencial es el oscilador armónico para cada enlace.

$$U_{i,j}^{alarg} = \frac{1}{2} k_{I,J} (b_{i,j} - b_{I,J}^o)^2 \quad [3.2]$$

Donde  $i,j$  representan a los átomos  $i$  y  $j$  y las letras mayúsculas  $I$  y  $J$  representan parámetros de los átomos  $i$  y  $j$ .  $k_{I,J}$  es la constante de fuerza asociada al enlace y da cuenta del costo energético que implica alejarlo de la posición de referencia  $b_{I,J}^o$ .

- i. Deformación del ángulo de enlace:  $\sum_{\text{ángulos}} \frac{k_\theta}{2} (\theta - \theta_o)^2$

Esta expresión tiene relación con la deformación del ángulo de enlace formado entre tres átomos y se calcula utilizando la siguiente función para cada ángulo:

$$U_{i,j,k}^{def} = \frac{1}{2} k_{I,J,K} (\theta_{i,j,k} - \theta_{I,J,K}^o)^2 \quad [3.3]$$

en donde  $k_{I,J,K}$  es la constante de fuerza asociada a la deformación del ángulo de enlace y por ende, indica el costo energético de deformar tal ángulo ( $\theta_{I,J,K}^o$ ).

- iii. Deformación torsional:  $\sum_{\text{torsión}} \frac{V_n}{2} [1 + \cos(n\chi - \delta)]$

Esta expresión se relaciona con la rotación alrededor del enlace 2-3, a los cuales se unen los átomos 1 y 4, respectivamente (1-2-3-4). Para determinar la contribución de este término a la energía potencial para cada diedro, se utiliza una serie de Fourier truncada:

$$U^{tor} = \frac{1}{2} \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\chi - \delta)] \quad [3.4]$$

Donde  $V_n$  está relacionado con la barrera rotacional alrededor del enlace, el parámetro  $n$  determina la multiplicidad (números de mínimos de la función a lo largo de una rotación de

360°), y  $\chi$  es el ángulo diedro formado entre la secuencia de átomos  $i, j, k, l$ .  $\delta$  es el parámetro que indica la fase del ángulo diedro.

#### iv. Términos no enlazantes

Las interacciones no enlazantes son las que consumen mayor tiempo en un cálculo y es por esto que para sistemas grandes éstas sumas son truncadas dejando fuera los pares de átomos que se encuentren a una distancia mayor a un valor arbitrario denominado radio de corte o cut-off. Existen varios métodos que se usan para abordar el cálculo de estas interacciones. En el caso de las interacciones de van der Waals (vdW) se utiliza el potencial de Lennard-Jones 12-6, el cual cae rápidamente según  $r^{-6}$  y  $r^{-12}$  en el caso de las fuerzas atractivas y repulsivas, respectivamente (ecuación 3.5). Por otra parte, las interacciones electrostáticas no disminuyen rápidamente con la distancia (ecuación 3.6). Existen distintos métodos que permiten llevar a cabo este cálculo más eficientemente, siendo uno de los más simples el uso de un cut-off de forma similar a como se hace con las interacciones de vdW. Sin embargo, debido a que estas interacciones son de alto rango se suelen aplicar otros métodos más sofisticados que incluyen el método PME<sup>147</sup> (del inglés Particle Mesh Ewald) o el algoritmo multipolo.<sup>148</sup>

#### a. Interacciones de van der Waals

Cada término de esta expresión es la suma de las interacciones atractivas débiles entre átomos distantes (fuerzas de dispersión de London) y las interacciones repulsivas entre átomos que se han acercado demasiado (Fuerzas de Pauli). Los campos de fuerza calculan este término con el potencial de Lennard-Jones 12-6.

$$U^{vdW} = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad [3.5]$$

En esta expresión no están incluidos los pares de átomos i-j enlazados. Las distancias entre

los pares de átomos i-j se calculan como  $r_{ij} = \left\| \vec{r}_j - \vec{r}_i \right\|$ . Los parámetros de las interacciones no enlazantes incluyen la profundidad del pozo o mínimo en la curva de energía de interacción,  $\varepsilon_{ij}$  y la distancia a la que la energía es cero,  $\zeta_{ij}$ , usado para el tratamiento las interacciones de van der Waals (vdW). Usualmente,  $\varepsilon_i$  y  $\zeta_i$  se obtienen para los tipos de átomos individuales y luego se combinan a través de reglas para obtener los parámetros  $\varepsilon_{ij}$  y  $\zeta_{ij}$ . Por último, suponiendo que se trata de un cálculo en presencia de solvente explícito,  $\varepsilon_0$  representa la permitividad.<sup>146</sup> El potencial de LJ se aproxima rápidamente a 0 a medida que  $r_{ij}$  aumenta, y usualmente es trucado a 0 luego de pasar el radio de corte.

#### b. Interacciones electrostáticas

Cada término de esta expresión es la suma de las interacciones electrostáticas entre pares de átomos, y su contribución a la energía total se calcula a través de la ley de Coulomb.

$$U^{elec} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad [3.6]$$

En donde  $\epsilon_0$  es la constante dieléctrica en el vacío.

En resumen, la ventaja principal de los métodos de FF es la rapidez con la que una determinada simulación puede ser realizada, lo cual permite aplicar estos métodos al estudio y modelización de macromoléculas como proteínas y ADN. Para sistemas con buenos parámetros disponibles, es posible hacer buenas predicciones de geometrías y energías en un tiempo corto. Sin embargo, si no se dispone de parámetros adecuados o éstos son de poca calidad, la aplicación de estos métodos FF se verá limitada.<sup>137</sup>

### 3.2 Métodos de la Química Cuántica

Si se requiere describir la distribución electrónica explícitamente por ejemplo, una reacción química, procesos de transferencia electrónica, etc, no hay otra opción que la mecánica cuántica (QM del inglés Quantum Mechanics)

Los postulados de la mecánica cuántica establecen que la denominada función de onda,  $\Psi$ , define el estado de cualquier sistema y que operadores apropiados actuando sobre  $\Psi$  darán el valor de la propiedad observable del sistema (valor propio) asociado a ese operador. La notación matemática para esto, también conocida como ecuación de valores propios, se da a continuación utilizando el operador Hamiltoniano, por lo que corresponde a la ecuación de Schrödinger.

$$\hat{H}\Psi_i = E_i\Psi_i \quad [3.7]$$

$E_i$  es la energía total del estado descrito por la función de onda  $\Psi_i$  y  $\hat{H}$  es el Hamiltoniano.

La forma típica del operador Hamiltoniano independiente del tiempo y sin considerar efectos relativistas, incluye 5 términos que contribuyen a la energía total del sistema: la energía cinética de electrones y núcleos, la atracción entre electrones y núcleos y las repulsiones interelectrónicas e internucleares:

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}} \quad [3.8]$$

en donde  $i$  y  $j$  abarcan a todos los electrones,  $k$  y  $l$  abarcan a todos los núcleos,  $\hbar$  es la constante de Planck dividida por  $2\pi$ ,  $m_e$  es la masa del electrón,  $m_k$  es la masa del núcleo  $k$ ,  $\nabla^2$  es el operador Laplaciana,  $e$  es la carga del electrón,  $Z$  es el número atómico y  $r_{ik}$  es la distancia entre las partículas  $i$  y  $k$ . Dado que la ecuación 3.7 es matemáticamente inviable para sistemas polielectrónicos, comúnmente se asumen algunas aproximaciones que

permiten abordar su resolución. La primera y más importante de estas se conoce como aproximación de Born-Oppenheimer, en la que se asume que es posible separar el movimiento electrónico del movimiento nuclear debido a que los protones y los neutrones son aproximadamente 1800 veces más pesados que los electrones, por lo que el movimiento de éstos es bastante más rápido que el movimiento nuclear en sí mismo.<sup>149</sup> Bajo la aproximación de Born-Oppenheimer, la parte electrónica de la función de onda puede desacoplarse de la parte nuclear y dado que las partículas nucleares se consideran fijas, la energía cinética nuclear es nula y sólo existe un término para su energía potencial.

$$\hat{H}_{elec} = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} = \hat{T}_e + \hat{V}_{N-e} + \hat{V}_{ee} \quad [3.9]$$

La ecuación 3.9 sólo contiene 3 términos (ver ecuación 3.8) que describen la energía cinética de los electrones ( $\hat{T}_e$ ), la interacción atractiva electrón-protón ( $\hat{V}_{N-e}$ ) y la repulsión electrón-electrón ( $\hat{V}_{ee}$ ). Cabe destacar que sin la aproximación Born-Oppenheimer no existiría el concepto de superficie de energía potencial o PES (del inglés Potential Energy Surface). La PES es la superficie que define la energía electrónica sobre todas las posibles configuraciones nucleares y se obtiene de la siguiente manera:

$$U(\vec{R}) = E_{elec}(\vec{R}) + V_{N-N}(\vec{R}) \quad [3.10]$$

donde  $U(\vec{R})$  es la energía total del sistema asumiendo que los núcleos tienen posiciones fijas en una configuración nuclear  $\vec{R}$ . A pesar de las ventajas que ofrece esta aproximación, aún se ha de resolver el término de correlación electrón-electrón, el cual es bastante complicado. A continuación, se resumirán los métodos que más se utilizan para tratar con sistemas polielectrónicos. Los métodos en los que la estructura electrónica es calculada a partir de la ecuación de Schrödinger y las constantes físicas fundamentales se conocen como métodos *ab initio* (del latín, desde el principio). En el caso de utilizar datos experimentales, se habla de métodos semiempíricos

### 3.2.1 El Método Variacional

Uno de los métodos más utilizados para obtener una aproximación a la energía del estado fundamental de un sistema con un Hamiltoniano independiente del tiempo es el método variacional. Este método se basa en el teorema variacional o teorema de Eckart que establece que si  $\Phi$  es cualquier función normalizada de las coordenadas de las partículas del sistema que se comporta bien y que satisface las condiciones de contorno del problema:

$$W = \langle \Phi | \hat{H} | \Phi \rangle \geq E_0 \quad [3.11]$$

en donde  $E_0$  es la energía exacta del estado fundamental del sistema, o en otras palabras, el valor propio más bajo de  $\hat{H}$ ,  $W$  se conoce como integral variacional. Este teorema establece que para cualquier sea  $\Phi$  (o función de prueba), el valor medio que se obtiene cuando  $\hat{H}$  opera sobre  $\Phi$  siempre será mayor o igual que la energía del estado fundamental,  $E_0$ . Si estos valores llegaran a coincidir, entonces  $\Phi$  también coincidiría con  $\Psi_0$  y el valor medio calculado corresponde efectivamente a  $E_0$ .

### 3.2.2 El Método Hartree-Fock

El método Hartree-Fock (HF) es un método variacional. Como ya se ha mencionado antes, la ecuación de Schrödinger no puede ser resuelta exactamente para un sistema con  $N$ -electrones y se deben utilizar aproximaciones. La aproximación más simple es considerar que los  $N$ -electrones son independientes y generar una función de onda  $N$ -electrónica representada por el producto antisimétrico de  $N$  funciones mono-electrónicas u spin-orbitales,  $\chi_i$ . El producto antisimétrico de los spin-orbitales se conoce también como determinante de Slater. La función mono-electrónica u orbital es el producto de una función espacial  $\varphi_i$  dependiente de coordenadas espaciales y una función de spin,  $\alpha(s)$  o  $\beta(s)$ . Para construir el determinante de Slater debe escogerse un conjunto de spin-orbitales, y es muy conveniente que este conjunto sea ortonormal. La aproximación fundamental que subyace este método es que cada electrón se mueve en un campo eléctrico promedio creado por el resto de los electrones. Luego, el procedimiento de optimización de spin-orbitales para todos los electrones se realiza sujeto a la restricción de ortonormalidad, lo que quiere decir que las expresiones serán derivadas haciendo que los spin-orbitales cumplan las siguientes condiciones:

$$\int \varphi_i^*(r) \varphi_j(r) dr = \delta_{ij} \quad [3.12a]$$

$$\begin{aligned} \langle \alpha | \alpha \rangle &= \langle \beta | \beta \rangle = 1 \\ \langle \alpha | \beta \rangle &= \langle \beta | \alpha \rangle = 0 \end{aligned} \quad [3.12b]$$

Esta aproximación es la más simple y común, aunque debe ser dicho que hay otras aproximaciones que no asumen esta restricción pero sustancialmente más complejas.

Aplicando el teorema variacional, el valor de expectación del Hamiltoniano resultante será el más bajo que se puede obtener para una función de onda mono-determinantal. La ecuación de HF tiene la siguiente forma:

$$\hat{f}\chi_i = \varepsilon_i \chi_i \quad [3.13]$$

en donde  $\hat{f}$  es el operador de Fock, un operador mono-electrónico, que incluye el potencial medio que los electrones y el núcleo crean sobre un electrón;  $\chi_i$  son los spin-orbitales canónicos y  $\varepsilon_i$  es la energía del  $i^{\text{ésimo}}$  spin-orbital. Dado que en la ecuación 3.13 el operador de Fock depende de los spin-orbitales que se busca determinar, la ecuación se debe resolver iterativamente, lo que se conoce como técnica de campo auto-consistente o SCF (del inglés Self-Consistent Field).

### 3.2.2.1 Aproximación LCAO y las Ecuaciones Roothaan-Hall

Hasta ahora, las aproximaciones mencionadas son puramente formales y las ecuaciones de HF son ecuaciones integro-diferenciales. En 1951 Roothaan y Hall propusieron usar un conjunto base de orbitales a través de la expansión de los orbitales usando la Combinación Lineal de Orbitales Atómicos o LCAO (del inglés Linear Combination of Atomic Orbitals), lo cual conduce a un conjunto de ecuaciones algebraicas que se puede resolver con métodos matriciales del algebra lineal. Los orbitales moleculares se expanden como:

$$\varphi_i = \sum_{\mu} c_{\mu i} \varphi_{\mu} \quad [3.14]$$

donde las  $\varphi_{\mu}$  son las funciones de base. La implementación de la aproximación LCAO en la ecuación 3.13 da lugar a las ecuaciones de Roothaan-Hall. Esto transforma las ecuaciones HF 3.13 derivadas del método variacional en ecuaciones lineales en las que los parámetros variacionales son los coeficientes  $c_{\mu i}$  de la expansión 3.14. Antes de aplicar esta aproximación se debe decidir si adoptar o no una restricción de spin. Para sistemas con número par de electrones en su estado fundamental, capa cerrada, se adopta la restricción de spin o RHF (del inglés Restricted Hartree-Fock); para sistemas de capa abierta se puede escoger mantener la restricción de spin, método conocido como restricción de spin de capa abierta o OSRHF (del inglés Open-Shell Restricted Hartree-Fock), o no adoptar una restricción de spin, en cuyo caso se habla de método HF no restringido o UHF (del inglés Unrestricted Hartree-Fock). Teniendo esto en cuenta y asumiendo una restricción de spin para sistemas de capa cerrada, a continuación se da la expresión matricial de la ecuación 3.13, lo cual conduce a las ecuaciones de Roothaan-Hall:

$$FC = SCE \quad [3.15]$$

En donde  $F$  es la representación matricial del operador de Fock en el conjunto base de orbitales,  $C$  es la matriz de los coeficientes de la expansión para la correspondiente  $\varphi$ ,  $S$  es la matriz de solapamiento y  $E$  es la matriz diagonal de las energías  $\varepsilon$  de los orbitales. Los elementos de la matriz de Fock son:

$$F_{\nu\mu} = \left\langle \phi_\nu \left| \hat{f} \right| \phi_\mu \right\rangle = H_{\nu\mu}^{core} + G_{\nu\mu} = H_{\nu\mu}^{core} + \sum_{\lambda\sigma} P_{\lambda\sigma} G_{\nu\mu\lambda\sigma} \quad [3.16]$$

en donde  $H_{\nu\mu}^{core}$  y  $G_{\nu\mu}$  son las representaciones matriciales de las partes mono- y bi-electrónicas, respectivamente, del operador de Fock. Los elementos de  $H_{\nu\mu}^{core}$  son las integrales mono-electrónicas  $H_{\nu\mu}$ , y  $G_{\nu\mu}$  depende de las integrales bi-electrónicas y también de los coeficientes;  $G_{\nu\mu\lambda\sigma}$  contiene a las integrales bi-electrónicas y  $P_{\nu\zeta}$  es un elemento de la matriz densidad  $P$ , los cuales dependen de los coeficientes  $c_{b\nu}$  y  $c_{b\zeta}$  según la siguiente ecuación:

$$P_{\lambda\sigma} = 2 \sum_b^{N/2} c_{\lambda b} c_{\sigma b}^* \quad [3.17]$$

La expresión de energía es la siguiente:

$$E_o = \frac{1}{2} \sum_\nu \sum_\mu P_{\nu\mu} (H_{\nu\mu}^{core} + F_{\nu\mu}) \quad [3.18]$$

Luego de que se define la matriz densidad de prueba,  $F$  se calcula y la ecuación matricial puede ser resuelta iterativamente hasta que la energía converge dentro de la precisión deseada.<sup>150</sup>

El límite HF se define como aquellas soluciones de las ecuaciones HF que se obtendrían al utilizar un conjunto de bases infinito. Dado que esto no es una posibilidad real, algunas veces el límite HF puede ser extrapolado con validez aceptable.<sup>135</sup> Utilizando un conjunto de base lo suficientemente grande, el método HF podría proporcionar hasta el 99 % del total de la energía electrónica.<sup>137</sup> Dado que el límite HF es imposible de llevar a la práctica, el conjunto de funciones de base es truncado hasta un número finito de funciones de base que se escogen en dependencia de la naturaleza química del proceso que se desea estudiar. Un conjunto mínimo de funciones de base contiene sólo una función por orbital atómico. Sin embargo, existen conjuntos de base más grandes en los que se pueden utilizar dos o tres funciones para describir un orbital atómico, conjuntos doble- $\zeta$  y triple- $\zeta$ , etc, respectivamente. Adicionalmente, también se puede considerar la distorsión de orbitales atómicos añadiendo funciones de polarización. Por último, otras características químicas asociadas a especies aniónicas o radicales pueden ser consideradas al añadir funciones difusas que describirán de una mejor manera el comportamiento de aquellos electrones más distantes del núcleo.

Como se mencionó al principio del párrafo anterior, la consideración de un conjunto base lo suficientemente grande podría proveer hasta el 99 % de la energía electrónica total. El 1 % restante representa la energía de correlación electrónica, la cual corresponde a la diferencia entre la energía exacta no relativista del sistema y la energía correspondiente al límite HF. La mayor limitación de la teoría HF que se construye en base a las ecuaciones de Roothaan-Hall, es la naturaleza mono-electrónica del operador de Fock que no permite considerar la correlación electrónica debido a que cada electrón ve a todos los demás como

un campo eléctrico promedio. Muchos esfuerzos han sido realizados para incluir la  $E_{corr}$  en los llamados métodos post-HF (por ejemplo: los métodos Moller-Plesset, Coupled Cluster). Además, los métodos basados en la Teoría de los Funcionales de la Densidad o DFT (del inglés Density Functional Theory) (sección 3.2.3) también abordan el cálculo de la energía de correlación.

### 3.3 La Teoría de los Funcionales de la Densidad

Los cálculos convencionales de las propiedades de moléculas basados en la resolución de la ecuación de Schrödinger tienen una alta complejidad matemática debido a que involucran la descripción del movimiento individual de los electrones. La función de onda depende de una coordenada de spin y tres coordenadas espaciales para cada electrón (asumiendo que los núcleos están en posiciones fijas), por lo que a mayor número de electrones, más complicado es el problema. Además, la función de onda en sí misma no tiene un significado físico por lo que la pregunta que surge en este punto es: ¿Sería posible determinar la energía electrónica a partir de algún observable físico de la molécula? Hasta ahora el Hamiltoniano depende de las posiciones y números atómicos de los núcleos y el número total de electrones. Al depender del número total de electrones, es intuitivo suponer que un observable útil sería la densidad electrónica, dado que la integración de la densidad electrónica sobre todo el espacio, da el número total de electrones, N:

$$N = \int \rho(r) dr \quad [3.19]$$

El establecimiento riguroso de la teoría DFT como una metodología QM llegó en 1964, cuando Hohenberg y Kohn probaron dos teoremas críticos. El primer teorema establece que:

*Cualquier observable de un estado estacionario fundamental no degenerado puede ser, en teoría, exactamente calculado a partir de la densidad electrónica del estado fundamental. En otras palabras, cualquier observable puede ser escrito como un funcional de la densidad electrónica del estado fundamental.*

En el lenguaje DFT, los electrones interactúan entre sí y con un potencial externo. En las primeras aproximaciones, este potencial externo era una carga positiva uniformemente distribuida (el gas electrónico uniforme). Por otra parte, el potencial externo en una molécula es la atracción nuclear que viene dada por la expresión habitual. De esta manera, el primer teorema de Hohenberg-Kohn establece que para un potencial externo específico ( $V_{Ne}(\vec{r})$ ), las propiedades del estado fundamental de un sistema polielectrónico son exclusivamente determinadas por la  $\rho(\vec{r})$ . En este contexto, la energía electrónica  $E_o$  puede ser expresada como:

$$E_o[\rho_o] = \int \rho_o(\vec{r}) V_{Ne}(\vec{r}) d\vec{r} + F[\rho_o(\vec{r})] \quad [3.20]$$



en donde  $V_{Ne}(\vec{r})$  corresponde al potencial de energía debida a la atracción electrón-núcleo y  $F[\rho_o(\vec{r})]$  es el funcional universal dado que no dependen del potencial externo:

$$F[\rho_o(\vec{r})] = T[\rho_o(\vec{r})] + J[\rho_o(\vec{r})] + E_{ncl}[\rho_o(\vec{r})] \quad [3.21]$$

en donde  $T[\rho_o(\vec{r})]$  da cuenta de las contribuciones individuales a la energía cinética,  $J[\rho_o(\vec{r})]$  se refiere a la interacción de Coulomb clásica y  $E_{ncl}[\rho_o(\vec{r})]$  incluye interacciones electrón-electrón no clásicas que dan cuenta de la energía de correlación de intercambio y coulombica, así como las correcciones de auto-interacción.

$J[\rho_o(\vec{r})]$  tiene una expresión clásica, sin embargo  $T[\rho_o(\vec{r})]$  y  $E_{ncl}[\rho_o(\vec{r})]$  son expresiones desconocidas. Los primeros intentos de deducción de funcionales para la energía cinética y de intercambio consideraron un gas electrónico uniforme, una suposición que podría ser válida para los electrones de valencia en ciertos sistemas metálicos, pero que resulta bastante ineficiente para átomos y moléculas.<sup>137</sup>

El segundo teorema de Hohenberg-Kohn usa el principio variacional para obtener  $E[\rho]$  y puede ser expresado como:

*La densidad electrónica de un estado fundamental no degenerado puede ser calculada exactamente, en teoría, determinando la densidad que minimiza la energía del estado basal.*

$$E_o[\rho_o(\vec{r})] \leq \int \rho_{pr} V_{Ne}(\vec{r}) dr + T[\rho_{pr}(\vec{r})] + J[\rho_{pr}(\vec{r})] + E_{ncl}[\rho_{pr}(\vec{r})] \quad [3.22]$$

Este principio prueba que cualquiera sea la densidad de prueba, ésta resultará en una energía mayor o igual que la energía exacta del estado fundamental. De este modo, para obtener la densidad exacta del estado basal se tendrá que encontrar aquella densidad que minimice la energía:

$$\left[ \frac{\delta E[\rho]}{\delta \rho} \right] = 0 \quad [3.23]$$

Aunque esta ecuación da una receta para minimizar la energía, sigue existiendo el problema que genera no tener a priori una expresión exacta para los términos no clásicos dentro del funcional  $F[\rho_o(\vec{r})]$ . Además, no existe ninguna receta para escoger de una manera racional una densidad de prueba y por otro lado, no existe una aproximación que permita calcular la energía directamente a partir de la densidad sin recursos asociados a la función de onda. Tal aproximación apareció por primera vez en 1965.<sup>135</sup>

### 3.3.1 El Método Kohn-Sham

El fundamento del uso de los métodos DFT en la química computacional fue la introducción de orbitales sugerida por Kohn y Sham (KS).<sup>151</sup> La idea crucial del método Kohn-Sham es que toma como punto de partida un sistema ficticio de electrones no interaccionantes que se mueven bajo un potencial externo ( $v_s(r)$ ), cuya densidad electrónica en el estado fundamental es la misma que la de un sistema real en el que los electrones sí interaccionan. Cuando este potencial externo es aplicado sobre un sistema da una función de onda ( $\psi_s$ ) que tiene la misma densidad del sistema real. Dado que los electrones no interaccionan, la aproximación de orbitales que conduce a las ecuaciones HF, daría un resultado exacto. El Hamiltoniano de este sistema sólo contiene términos mono-electrónicos y tiene la siguiente forma:

$$\hat{H}_s = \sum_{i=1}^N -\frac{1}{2} \nabla_i^2 + \sum_{i=1}^N v_s(\vec{r}_i) = \sum_{i=1}^N \hat{h}^{KS}(i) \quad [3.24]$$

en donde  $\hat{h}^{KS}$  es el Hamiltoniano mono-electrónico Kohn-Sham. La función de onda exacta para este sistema es el determinante de Slater (ecuación 3.25) y los orbitales moleculares de este determinante se obtienen resolviendo la ecuación de Schrödinger (ecuación 3.26):

$$\psi_s = \frac{1}{\sqrt{N!}} |\chi_1(1)\chi_2(2)\cdots\chi_N(N)| \quad [3.25]$$

$$\left[ -\frac{1}{2} \nabla^2 + v_s(\vec{r}_i) \right] \chi_i = \varepsilon_i \chi_i \quad [3.26]$$

En donde  $\chi_i$  y  $\varepsilon_i$  son los orbitales Kohn-Sham y las energías mono-electrónicas, respectivamente. Estas ecuaciones mono-electrónicas son resueltas iterativamente y dado que el Hamiltoniano depende de la densidad electrónica, es necesario tener una densidad de prueba. Como en el método HF, el procedimiento para resolver estas ecuaciones se inicia a partir de un conjunto de orbitales moleculares de prueba,  $\{\chi_i(\vec{r})\}$ , con los cuales es posible determinar la densidad electrónica de prueba usando la ecuación 3.27. Esta densidad sirve para obtener el potencial externo y para resolver subsecuentemente la ecuación 3.26, un procedimiento que se repite hasta obtener la convergencia deseada.

Por otra parte, continuando bajo la suposición de este sistema ideal en el que los electrones no interaccionan, la densidad y energía cinética exacta toman la forma siguiente:

$$\rho(\vec{r}) = \sum_{i=1}^N |\chi_i(\vec{r})|^2 \quad [3.27]$$

$$T_s[\rho] = -\frac{1}{2} \sum_{i=1}^N \langle \chi_i | \nabla_i^2 | \chi_i \rangle \quad [3.28]$$

Sin embargo, dado que se trata de un sistema modelo, la energía cinética del sistema real,  $T[\rho]$ , es diferente a lo que se muestra en la ecuación 3.28. Para tener en cuenta estas diferencias, el funcional  $F[\rho(\vec{r})]$  (ecuación 3.21, se quita el subíndice —basumiendo que seguimos en el estado fundamental) debe ser reformulado definiendo los siguientes términos:

$$F[\rho(\vec{r})] = T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + (T[\rho(\vec{r})] - T_s[\rho(\vec{r})]) + E_{ncl}[\rho(\vec{r})] \quad [3.29]$$

en esta ecuación;  $T_s[\rho(\vec{r})]$  es la energía cinética electrónica de un sistema que tiene la misma densidad del sistema real, pero en el que las interacciones interelectrónicas son ignoradas;  $J[\rho(\vec{r})]$  es la energía de repulsión Coulombica;  $T_c[\rho(\vec{r})] = (T[\rho(\vec{r})] - T_s[\rho(\vec{r})])$ , esta expresión es la diferencia entre la energía cinética del sistema real y la del sistema de referencia, también conocida como energía cinética de correlación. El último término de esta ecuación corresponde a la energía de correlación e intercambio que combinada con  $T_c[\rho(\vec{r})]$ , conforman la energía de correlación e intercambio total.

$$E_{xc}[\rho] = (T[\rho(\vec{r})] - T_s[\rho(\vec{r})]) + E_{ncl}[\rho(\vec{r})] \quad [3.30]$$

$E_{xc}$  incluye todas las contribuciones a la energía para las cuales no se tiene una expresión simple en función de la densidad electrónica: la energía de intercambio, la energía de correlación, la porción de energía necesaria para corregir  $T_s[\rho(\vec{r})]$  y la corrección asociada al error de auto-interacción o SIE (del inglés Self-Interaction Error).

Un aspecto básico de la teoría DFT, y que marca la diferencia entre el número de funcionales existentes, es encontrar una buena expresión para el funcional de correlación e intercambio cuya forma exacta no se conoce.<sup>150</sup> Cuando se discute la naturaleza de diferentes funcionales, es útil escoger alguna de las notaciones más comúnmente usadas en el campo. Por ejemplo, la dependencia funcional de  $E_{xc}$  con respecto a la densidad electrónica es expresada como una interacción entre la densidad electrónica y una "densidad de energía"  $\varepsilon_{xc}$  que depende de la densidad electrónica, concretamente.

$$E_{xc}[\rho(\vec{r})] = \int \rho(\vec{r}) \varepsilon_{xc}[\rho(\vec{r})] d\vec{r} \quad [3.31]$$

La densidad de energía  $\varepsilon_{XC}$  siempre es tratada como la suma individual de contribuciones de correlación e intercambio. Notar que dos diferentes densidades están involucradas en este término: la densidad electrónica por unidad de volumen y la densidad de energía por partícula.<sup>135</sup> Es costumbre separar  $E_{XC}$  en dos partes, la parte de intercambio  $E_X$  y la parte de correlación

$$E_{XC} = E_X[\rho] + E_C[\rho] = \int \rho(\vec{r}) \varepsilon_X[\rho(\vec{r})] d\vec{r} + \int \rho(\vec{r}) \varepsilon_C d\vec{r} \quad [3.32]$$

Como se mencionó antes, un aspecto fundamental de la teoría DFT es encontrar buenas aproximaciones para la energía de correlación e intercambio. La más simple y al mismo tiempo tremendamente útil es la aproximación LDA (del inglés Local Density Approximation). El siguiente nivel lo conforman las aproximaciones que se conocen como no-locales o aproximaciones de gradiente generalizado GGA (del inglés Generalized Gradient Approximation). A continuación se revisan brevemente estas aproximaciones junto a los funcionales meta-GGA y los denominados funcionales híbridos

### 3.3.2 Aproximación de la Densidad Local o LDA

En la aproximación LDA, el sistema modelo lo representa un gas de electrones uniforme en el que los electrones se mueven en un campo de carga positiva de modo que el sistema total es neutro. En la práctica, se asume que  $\rho(\vec{r})$  varía tan lentamente con la posición que se puede considerar constante. Aunque no se conoce una expresión explícita para la parte de correlación,  $\varepsilon_C^{LDA}[\rho]^*$ , esta aproximación en la que la  $\rho(r)$  es constante permite obtener una expresión para la parte de intercambio.

$$\varepsilon_X^{LDA}[\rho] = -\frac{9}{4} \alpha \left( \frac{3}{8\pi} \right)^{\frac{1}{3}} \rho^{\frac{1}{3}} \quad [3.33]$$

Esta es la energía de correlación por partícula en la que  $\alpha=2/3$ , lo que finalmente nos permite escribir:

$$E_X^{LDA}[\rho] = -\frac{9}{4} \alpha \left( \frac{3}{8\pi} \right)^{\frac{1}{3}} \int \rho^{\frac{4}{3}} d\vec{r} \quad [3.34]$$

Cuando las densidades de spin  $\alpha$  y  $\beta$  son diferentes, se utiliza la aproximación LSDA (del inglés Local Spin Density Approximation),<sup>152</sup> en la que  $\rho(\vec{r})$  es dividida en una parte  $\rho^\alpha(\vec{r})$  y otra  $\rho^\beta(\vec{r})$ .

A pesar de la simplicidad de la aproximación LDA, su aplicación da lugar a resultados razonables para sistemas en los que la densidad electrónica se mantiene aproximadamente constante. Sin embargo, para átomos o moléculas en los que hay variaciones de densidad importantes, también trabaja razonablemente bien. La principal razón de lo anterior es la cancelación sistemática de errores, lo cual se asocia al hecho de que la aproximación subestima  $E_x(\rho)$  y sobreestima  $E_c(\rho)$ , de manera que comúnmente genera un valor correcto de  $E_{xc}(\rho)$ .<sup>153</sup>

### 3.3.3 Aproximación GGA y meta GGA

En la aproximación LDA se asume que los efectos de la energía de correlación e intercambio son locales y dependen únicamente del valor de la densidad electrónica en cada punto. La introducción de gradientes de la densidad en la descripción de los efectos de correlación e intercambio da origen a la aproximación GGA. En esta aproximación, no sólo se tiene en cuenta el valor de la densidad en cada punto sino también, cómo varía esta densidad alrededor de cada punto,  $\vec{\nabla}\rho(\vec{r})$ . Esto permite describir en alguna medida, la heterogeneidad de la densidad electrónica real. De este modo, la expresión general para la energía de correlación e intercambio tiene la siguiente forma:

$$E_{xc}^{GGA}[\rho] = \int f(\rho(\vec{r}), \vec{\nabla}\rho(\vec{r})) d\vec{r} \quad [3.35]$$

Dos tendencias filosóficas imperaron en el desarrollo de los funcionales basados en la aproximación GGA (o funcionales GGA). La primera tiene una naturaleza empírica dado que se basa en un procedimiento de ajuste numérico considerando un conjunto grande de moléculas prueba, idea que fue propuesta por Becke. [150-155H] Algunos ejemplos los proporcionan los funcionales de intercambio de Becke88 (B88),<sup>154</sup> el funcional modificado de Perdew-Wang (mPW),<sup>155</sup> [J.Chem.Phys.1998,108,664-675] el funcional OptX (O),<sup>156</sup> el funcional X;<sup>157</sup> y los funcionales de intercambio y correlación Perdew-Wang (PW).<sup>155</sup> La segunda tendencia dentro de los métodos GGA considera que los funcionales de correlación e intercambio deben ser desarrollados estrictamente desde la mecánica cuántica y fue propuesta por Perdew.<sup>158,159</sup> Algunos de los funcionales que forman parte de este segundo grupo son: el funcional de intercambio Becke86 (B86),<sup>160</sup> el funcional de correlación Perdew86 (P86),<sup>161</sup> el funcional de intercambio y correlación Perdew-Burke-Ernzerhof (PBE)<sup>162</sup> y el funcional PBE modificado (mPBE).<sup>162,163</sup>

Los funcionales meta-GGA representan el siguiente paso en el desarrollo de funcionales DFT. En esta aproximación, los funcionales de energía de correlación e intercambio contienen, además de la densidad y su gradiente, la densidad de energía cinética (ecuación 3.36) y/o la Laplaciana de la densidad electrónica (ecuación 3.37), que siguen a continuación.

$$\tau(\vec{r}) = \frac{1}{2} \sum_{i=1}^N |\vec{\nabla}\chi_i(\vec{r})|^2 \quad [3.36]$$

$$E_{XC}^{meta-GGA}[\rho] = \int f(\rho, \vec{\nabla}\rho, \tau, \nabla^2\rho) d\vec{r} \quad [3.37]$$

En la ecuación 3.36 se ve que en los funcionales meta-GGA hay una dependencia específica con respecto a los orbitales de KS. Por otra parte, la presencia de  $\nabla^2\rho$  en la ecuación 3.37 indica que se deberán calcular las segundas derivadas de las funciones de base. Algunos ejemplos de estos funcionales los proveen el funcional BB95, mPWB95, PBEK CIS, TPSS<sup>164</sup> y el VSXC.<sup>165</sup>

### 3.3.4 Hiper GGA o Métodos Híbridos

Los métodos hiper-GGA incorporan los efectos de correlación e intercambio de los métodos GGA y algún porcentaje del intercambio exacto derivado de un cálculo HF. Una primera aproximación para obtener la energía de correlación e intercambio fue propuesta por Becke en 1993, en la cual se considera la misma proporción de la energía intercambio exacto HF,  $E_X^{HF}$ , y la energía intercambio y correlación derivada de métodos DFT,  $E_{XC}^{DFT}$ .

$$E_{XC} = \frac{1}{2} E_{XC}^{DFT} + \frac{1}{2} E_X^{HF} \quad [3.38]$$

El uso de un coeficiente igual a 0.5 en la ecuación 3.38 define el denominado método H&H (derivado del inglés —half-and-half<sup>o</sup>), pero hay otros métodos basados en la conexión adiabática. Quizá uno de los métodos más comúnmente usados es el de Becke, que contiene tres parámetros:

$$E_{XC} = E_{XC}^{LSDA} + a_o(E_X^{exacto} - E_X^{LSDA}) + a_x\Delta E_X^{B88} + a_c\Delta E_C^{PW91} \quad [3.39]$$

En donde  $a_o=0.20$ ,  $a_x=0.72$  y  $a_c=0.81$ . Estos valores son los que mejor ajustan el funcional B3PW91 a un conjunto de datos termodinámicos. Otros funcionales híbridos relevantes son por ejemplo: B3LYP, MPW1K, PBE0, sólo por mencionar algunos.<sup>150</sup> Finalmente, el más reciente desarrollo de funcionales hasta ahora, son los denominados funcionales híbridos meta-GGA. En estos se combinan los funcionales meta-GGA con la energía de intercambio HF. Algunos de los funcionales que pertenecen a esta categoría son: el MPWB1K, TPSSh o el M05-2X.<sup>166</sup> El último de estos ha sido desarrollado por el grupo del Profesor Truhlar en la Universidad de Minnesota. Este funcional incorpora la densidad electrónica de spin ( $\rho$ ), el gradiente de la densidad ( $\vec{\nabla}\rho$ ), la densidad de energía cinética ( $\tau$ , ecuación 3.36) y por supuesto, el intercambio HF:

$$E_{XC}^{M05-2X}[\rho^\alpha, \rho^\beta] = \int \rho \varepsilon_X^{UEG}(\rho) F_{XC}(\rho^\alpha, \rho^\beta, \vec{\nabla}\rho^\alpha, \vec{\nabla}\rho^\beta, \tau^\alpha, \tau^\beta) d\vec{r} \quad [3.40]$$

En donde UEG quiere decir gas de electrones uniforme (del inglés Uniform Electron Gas). La densidad electrónica total viene dada por la suma de  $\rho^\alpha$  y  $\rho^\beta$ , y  $\epsilon_x^{UEG}(\rho) = -\left(\frac{3}{4}\pi\right)(3\pi^2\rho)$  corresponde a la energía de intercambio de una gas de electrones uniforme por electrón con spin no polarizado ( $\rho^\alpha = \rho^\beta$ ). Por último,  $F_{xc}$  da cuenta de los efectos de correlación y heterogeneidad.<sup>167</sup>

El funcional M05-2X describe apropiadamente cinética termoquímica e interacciones no covalentes tales como interacciones débiles, enlaces de hidrógeno, interacciones  $\pi$ - $\pi$  e interacciones de energía de nucleótidos. También describe de buena manera a los carbohidratos, incluyendo energías de disociación de especies fosfato-azúcar. Esta es la razón por la que se utiliza como método de referencia en cuanto a los valores energéticos.

El desarrollo de mejores potenciales para la correlación y el intercambio es actualmente un campo de investigación muy activo. Hay muchos funcionales disponibles que se mantienen en constante optimización. Tanto los funcionales híbridos GGA como los híbridos meta-GGA incorporan ajustes a datos experimentales y han sido muy exitosos en el análisis de diferentes problemas químicos. Sin embargo, esta misma característica les quita universalidad y por consiguiente, la elección del funcional es un paso importante al utilizar estos métodos.

## 3.4 Métodos Semiempíricos

Aunque las teorías HF y DFT están basadas en formalismos rigurosos, en la práctica aparecen limitaciones en la implementación de estos métodos. En décadas pasadas, cuando la capacidad computacional era mínima, llevar a cabo un cálculo HF sin aproximaciones era una tarea ingente y complicada incluso para pequeñas moléculas. Hoy en día, cuando se trata de sistemas muy grandes como biomoléculas el costo computacional HF sigue siendo elevado ya que el cálculo escala con la cuarta potencia del número de funciones de base usadas. Los métodos semiempíricos aumentan considerablemente la eficiencia computacional a costa de resolver menos integrales y parametrizar parte de las integrales que sí se consideran. Debido a esto, suelen funcionar mejor con sistemas para los que hay información experimental disponible.

### 3.4.1 Métodos Semiempíricos Hartree-Fock

En un cálculo HF, el paso computacional más demandante es la evaluación de un conjunto de integrales bi-electrónicas (también conocidas como integrales de 4-índices) que dan cuenta de la energía de interacción electrónica. El objetivo de los métodos semiempíricos es poder aproximar el valor de algunas integrales en lugar de calcularlas. Por ejemplo, las

integrales Coulombicas que miden la repulsión entre dos electrones que ocupan una región del espacio definida por las funciones de base que los describen, podrían ser aproximadas a cero si se asume que la función base asociada a un electrón en la integral mono-electrónica está demasiado distante de la función base del otro electrón. Lo mismo puede aplicarse para las integrales mono-electrónicas que dan cuenta de la atracción nuclear.<sup>135</sup>

En este contexto, la primera aproximación fue considerar explícitamente sólo los electrones de valencia, mientras que los efectos de los electrones más internos se tienen en cuenta reduciendo la carga nuclear. Además, los electrones de valencia son descritos por un conjunto mínimo de funciones base tipo Slater. De esta manera, una de las ideas centrales en los métodos semiempíricos HF la estableció la aproximación ZDO (del inglés Zero Differential Overlap), en la que se ignoran todos los productos de dos funciones de base diferentes que dependen de las mismas coordenadas electrónicas. Así todas las integrales mono-electrónicas de dos centros y todas las bi-electrónicas de 3 ó 4 centros son nulas. Para compensar estas aproximaciones, el resto de las integrales son convertidas en parámetros que pueden ser asignados a través de cálculos o datos experimentales.

En el método CNDO (Complete Neglect of Differential Overlap) se aplica la aproximación ZDO por completo. Es decir, el solapamiento diferencial  $\langle \phi_\mu(1) | \phi_\nu(1) \rangle$  se hace cero si  $\mu$  es distinto de  $\nu$ , incluso si ambas funciones de base están centradas en el mismo átomo. En el método NDDO (Neglect of Diatomic Differential Overlap) sólo se desprecia el solapamiento diferencial diatómico, reteniendo pues todas las integrales con solapamiento monoatómico, sean monocéntricas o dicéntricas. La mayoría de los métodos semiempíricos actuales están basados en el método original NDDO. Entre los más populares destacan por ejemplo: el método MNDO,<sup>168</sup> AM1<sup>169</sup> o PM3;<sup>170</sup> los tres se basan en la aproximación NDDO y se diferencian en la forma en que se trata la repulsión entre los átomos internos.

### 3.4.2 Métodos Semiempíricos DFT

Dentro de la teoría DFT, Adamson, Gill y Pople propusieron una parametrización de funcionales GGA puros diseñada específicamente para obtener buenos resultados utilizando conjuntos de base pequeños. Se trata de los funcionales de densidad empírica o EDF1 (del inglés Empirical Density Functional 1). La idea que motivó el desarrollo de estos modelos fue mejorar la eficiencia de los cálculos, lo cual podría ser aplicable a problemas que aún eran intratables para funcionales más rigurosos. Uno de estos modelos es usualmente conocido con la abreviación DFTB (del inglés Density Functional Tight-Binding).

Inicialmente se adoptó una aproximación no-SCF, esto es: una vez escogida una densidad electrónica, se optimizan los orbitales KS minimizando variacionalmente los coeficientes del conjunto de base sin calcular una nueva densidad a partir de estos. La aproximación DFTB supone que la densidad de carga de la molécula puede ser adecuadamente representada por la suma de las densidades de carga de sus átomos individuales y por consiguiente, no considera características como la polarización de enlace que puede ocurrir entre átomos con diferentes electronegatividades. Para tener en consideración tal polarización se desarrolló un nuevo método que se basa en una aproximación de segundo orden a la

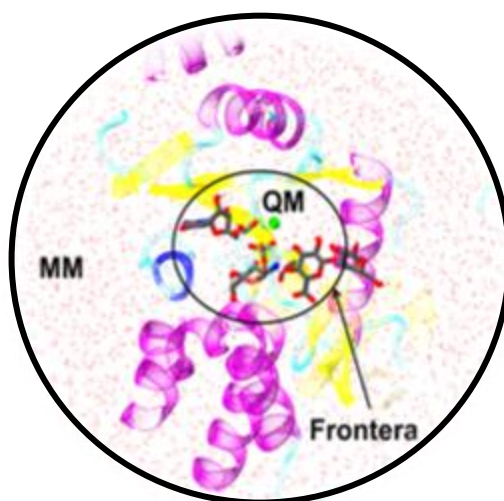


fluctuación de la densidad electrónica alrededor de la densidad fija  $\rho_o(r)$ ; además, la auto-consistencia (SCF) es reintroducida en esta aproximación mejorada, por lo que este método es comúnmente conocido como SCC-DFTB (del inglés Self-Consistente Charge Density Functional Tight-Binding).<sup>171-173</sup> Estrictamente, el método SCC-DFTB es una extensión de la teoría estándar TB (del inglés tight-binding)<sup>174</sup> aplicada a un modelo de carga consistente dentro de una expansión de segundo orden de la energía.

La aproximación SCC-DFTB es casi tan rápida como el método semiempírico NDDO<sup>135</sup> y a pesar de que las fuerzas de dispersión no son tratadas en una manera óptima, está siendo muy utilizada en el estudio de biomoléculas.

### 3.5 Aproximación QM/MM

Cuando se desea abordar el estudio de macromoléculas, en el caso de esta tesis enzimas, y el proceso químico de interés ocurre en una región específica mientras el resto de la enzima permanece sin sufrir cambios durante el proceso químico, resulta muy conveniente modelizar el sistema a través de la aproximación QM/MM. El formalismo QMMM permite combinar los mejor de los métodos QM y MM previamente revisados pues la región de interés, donde ocurre la formación o rotura de enlaces, puede tratarse a nivel QM mientras simultáneamente se trata el resto del sistema a nivel MM. (Figura 14)



**Figura 14.** En un cálculo QMMM sobre por ejemplo, una enzima, el sistema es dividido en un subsistema QM y MM. En el primero (barras y esfera) se encuentran la o las moléculas que participan en la reacción química y que serán descritas con el método QM que se escoja. En el subsistema MM se encuentra el resto del sistema, en este caso aguas y enzima, las cuales serán descritas por un método MM.

La metodología QMMM permite utilizar prácticamente cualquier combinación de métodos QM y MM. En la práctica, muchas de las aplicaciones disponibles para realizar este tipo de

cálculos utilizan la teoría DFT como método QM, dado que se obtienen buenos resultados dentro de un tiempo de cálculo menos demandante que métodos post-HF. Por otra parte, los métodos semiempíricos han sido frecuentemente utilizados y siguen siendo considerados importantes en los cálculos QMMM.

El poder de cálculo actual permite además implementar métodos de alto nivel *ab initio* en sistemas biomoleculares, aunque normalmente se utilizan para hacer estimaciones puntuales de energía sobre geometría previamente optimizadas. La teoría detrás de la aproximación QMMM ha sido descrito por varios autores,<sup>175-177</sup> pero aquí se dará sólo una breve revisión de los puntos más importantes que de esta metodología.

### 3.5.1 Expresión de Energía QM/MM

En la aproximación QMMM el sistema debe dividirse en dos subsistemas (Figura 14), cada uno de los cuales conforma la región o el subsistema QM y MM. Estas dos regiones interactúan muy fuertemente por lo que la energía del sistema completo no es simplemente la suma de las energías de los subsistemas individuales. Hay dos grandes esquemas para llevar a cabo el cálculo QMMM: el esquema sustractivo y el esquema aditivo. En la primera, por ejemplo el método ONIOM en Gaussian,<sup>178</sup> la energía total,  $E_{QM/MM}^{sustractivo}(C)$ , tiene la siguiente forma:

$$E_{QM/MM}^{sustractivo}(C) = E_{MM}(C) + E_{QM}(I) - E_{MM}(I) \quad [3.41]$$

en donde C indica el sistema completo, e I señala a la parte interna o subsistema QM. En el esquema sustractivo son necesarios tres cálculos: un cálculo MM sobre el sistema completo, un cálculo QM sobre la parte interna (subsistema QM) y un cálculo MM sobre el subsistema QM.<sup>176</sup> La principal ventaja de este esquema es que el acoplamiento entre los subsistemas no es tratado explícitamente y por consiguiente, se pueden aplicar los métodos QM y MM estándar sin modificaciones adicionales. La otra aproximación a la energía QMMM usada dentro de los métodos QMMM es el esquema aditivo (utilizado en esta tesis). La expresión de energía general en este esquema, queda descrita de la siguiente manera:

$$E_{QM/MM}^{aditivo}(C) = E_{MM}(E) + E_{QM}(I+U) + E_{QM/MM}(I,E) \quad [3.42]$$

donde E se refiere a la parte externa del sistema (subsistema MM) e I señala a la parte interna o subsistema QM; U indica la frontera de acoplamiento entre los dos subsistemas QM y MM. El término  $E_{QM/MM}(I,E)$  incluye los términos de interacción de los dos subsistemas E e I, y el término  $E_{QM}(I+U)$  da cuenta del subsistema QM "dividido" y es calculado a nivel QM.<sup>177</sup> Lo que define un método QMMM en particular, es la forma exacta de la expresión  $E_{QM/MM}(I,E)$ , que de manera general se expresa como sigue:

$$E_{QM/MM}(I, E) = E_{QM/MM}^b + E_{QM/MM}^{vdW} + E_{QM/MM}^{el} \quad [3.43]$$

El primer término de la parte derecha de esta ecuación recoge los términos de interacción enlazantes de los subsistemas. Las interacciones no enlazantes quedan descritas por el potencial de van der Waals y las interacciones electrostáticas por el potencial de Coulomb (Ecuación 3.5 y 3.6, respectivamente).

### 3.5.2 Esquemas de Interacción Electrostática

La interacción electrostática entre la densidad de carga QM y el modelo de cargas usados en la región MM puede ser abordada con distintos niveles de sofisticación que se caracterizan por el grado de mutua polarización entre ambas partes.<sup>177</sup> En el caso del esquema de acoplamiento mecánico (del inglés mechanical embedding), las interacciones entre ambos subsistemas son tratadas como se haría en un modelo de cargas MM (cargas puntuales atómicas rígidas). Aunque esto resulta computacionalmente eficiente, existen algunas desventajas como por ejemplo; las cargas del subsistema MM no interactúan con las densidades de carga en el subsistema QM, por lo que esta última no es afectada por su ambiente electrostático (no hay polarización); dependiendo del proceso que se desea simular, la distribución de carga en el subsistema QM podría llegar a modificarse y sería razonable actualizar las cargas, sin embargo esto podría generar discontinuidades en la PES. Por último, la derivación de cargas puntuales para los átomos del subsistema QM no es un proceso trivial y además de ser una tarea pesada, no necesariamente reproduce la distribución de carga real del sistema.

En el esquema de acoplamiento electrostático (del inglés electrostatic embedding), las deficiencias del esquema anterior se soslayan ya que el cálculo QM se lleva a cabo en presencia del modelo de cargas MM, incorporando éstas últimas como un término mono-electrónico en el Hamiltoniano QM. Esto se expresa a continuación utilizando unidades atómicas (u.a.)

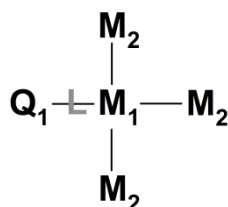
$$\hat{H}_{QM/MM}^{el} = - \sum_i^N \sum_{J \in E}^L \frac{q_J}{\left| \vec{r}_i - \vec{R}_J \right|} + \sum_{\alpha \in I+U}^M \sum_{J \in E}^L \frac{q_J Q_\alpha}{\left| \vec{R}_\alpha - \vec{R}_J \right|} \quad [3.44]$$

En donde  $q_J$  se refiere a las cargas puntuales del modelo de cargas MM ubicadas en la posición  $\vec{R}_J$ ;  $Q_\alpha$  indica las cargas nucleares de los átomos QM en  $\vec{R}_\alpha$ ; y  $\vec{r}_i$  denota la posición de los electrones. Los índices  $i$ ,  $J$  y  $\alpha$  suman sobre los  $N$  electrones, las cargas puntuales MM y los  $M$  núcleos QM, respectivamente. De esta manera, la región QM es automáticamente polarizada por los cambios en la distribución de cargas MM y las interacciones electrostáticas son tratadas a nivel QM consiguiéndose una mayor precisión. Sin embargo, este esquema no está exento de problemas pues en la región de unión o

frontera de ambos subsistemas, las cargas MM de la parte externa están en la vecindad inmediata de la densidad de carga QM, lo cual puede generar sobrepolarización sobre todo cuando la frontera se localiza directamente sobre un enlace covalente, que es lo que ocurre la mayoría de las veces en reacciones enzimáticas.<sup>177</sup> A continuación, se revisan brevemente las distintas aproximaciones que permiten abordar el problema de los átomos frontera.

### 3.5.3 Tratamiento de Átomos Frontera

Si tratamos con un modelo en el que alguna de las particiones entre el subsistema QM y MM queda a lo largo de un enlace covalente, el tratamiento de los átomos frontera de los subsistemas QM (átomo  $Q_1$ ) y MM (átomo  $M_1$ ) puede ser dividido en dos grandes grupos.



**Figura 15.** Representación de los átomos frontera de un sistema en el que la partición QM-MM se realiza a lo largo del enlace covalente  $Q_1-M_1$ . En el subsistema QM queda el átomo  $Q_1$  y en el subsistema MM quedan los átomos  $M_1$  (directamente conectado al átomo  $Q_1$ ), y los átomos  $M_2$  (separados del átomo  $Q_1$ ) por un enlace.

El primero utiliza el esquema del *link-atom*, en el que se introduce un centro atómico adicional (L) que no es parte del sistema real y que usualmente es un átomo de hidrógeno. El átomo L (Figura 15) se enlaza covalentemente al átomo  $Q_1$  completando su valencia y así el cálculo QM se realizará sobre un sistema electrónicamente saturado. El enlace  $Q_1-M_1$  se tratará a nivel MM. Por definición, el *link-atom* no es ni un átomo QM ni MM pues no está presente en ninguno de los subsistemas originales y la forma de definir su posición es hacer que sus coordenadas dependan de los átomos entre los que se encuentra ( $Q_1-L-M_1$ ), evitando de esta forma, añadir un nuevo grado de libertad al sistema. Usualmente, el *link-atom* es puesto a lo largo de la línea imaginaria que conecta el átomo  $Q_1$  con el átomo  $M_1$ .

Como se mencionó anteriormente, la densidad de carga QM tenderá a ser sobrepolarizada por el modelo de cargas puntuales rígidas MM, lo cual es aún más marcado cuando la carga puntual representa un átomo frontera. Existen diferentes aproximaciones que permiten minimizar la sobrepolarización que puede generar el átomo  $M_1$  dentro del esquema del *link-atom*; en una de ellas se borran del Hamiltoniano las cargas MM en la región del átomo L;<sup>177</sup> también se pueden ignorar las integrales mono-electrónicas asociadas con el átomo L.<sup>177</sup> Por otra parte, existen esquemas que permiten mantener la carga del átomo  $M_1$  (Figura 16) y a menudo el dipolo en la región frontera, e incluso así eliminar la sobrepolarización que podría generar el átomo  $M_1$ . Dentro de este grupo de esquemas, uno de los más usados es el que conoce como esquema de desplazamiento de carga o CS (del inglés Charge-

Shift),<sup>179-181</sup> y que de hecho, es el esquema que se ha utilizado en esta tesis. Básicamente, la carga del átomo  $M_1$  en el esquema CS es uniformemente desplazada hacia los átomos  $M_2$  con el objetivo de compensar la modificación del dipolo inicial  $M_1-M_2$ .

Adicionalmente, existen otras metodologías que evitan añadir un nuevo átomo al sistema y en lugar de esto, intentan imitar las propiedades electrónicas del fragmento MM utilizando pseudo-potenciales monovalentes (o potenciales efectivos) localizados en el átomo  $M_1$ . De esta manera, el potencial efectivo participa en el cálculo MM y a la vez satura la valencia incompleta del átomo  $Q_1$ .<sup>182-184</sup> Por último, también está la aproximación FLO (del inglés Frozen Localized Orbitals), cuya idea básica es localizar en cualquiera de los átomos frontera un conjunto de orbitales adecuadamente orientados y fijos, es decir, los orbitales no participan de los ciclos SCF. Algunos de los esquemas que forman parte de esta aproximación son; el campo de auto-consistencia local o LSCF (del inglés Local Self-Consistent Field);<sup>185</sup> el esquema de orbitales fijos, un variante del procedimiento LSCF;<sup>186</sup> y los orbitales híbridos generalizados o GHO (del inglés Generalized Hybrid Orbitals).<sup>187</sup>

La elección de los átomos frontera es muy importante y de manera general, se pueden mencionar algunos consejos para definir su ubicación. La más importante es localizar la partición QM-MM tan lejos como sea posible de los átomos químicamente activos, lo cual viene define el tamaño del subsistema QM. Finalmente, el enlace covalente en el que se hace la partición debería ser apolar y simple, por ejemplo, un enlace covalente entre carbonos alifáticos.

### 3.5.4 Otras Interacciones QM/MM no Enlazantes

Además de las interacciones electrostáticas (sección 3.5.2), las interacciones vdW y otras interacciones de enlace también contribuyen al término de acoplamiento QM-MM (ecuación 3.43). Las interacciones de vdW son típicamente tratadas con el potencial de Lennard-Jones (ecuación 3.5), por lo que se necesitan parámetros para los átomos del subsistema QM. La forma más común de obtener los parámetros de dispersión para estos átomos es usar los que ya se han derivado para el campo de fuerza que se esté empleando. Sin embargo, también existe la opción de derivar algunos parámetros Lennard-Jones específicos para complementar y mejorar los resultados.<sup>186,188,189</sup> De la misma manera, en el caso de las interacciones enlazantes (deformación de ángulos de enlace, estiramiento de enlaces, etc) también se suelen usar los parámetros estándar del campo de fuerza. En caso de que alguno de los parámetros que se requieren no estén derivados por el modelo MM, estos pueden ser complementados con términos enlazantes adicionales.

En cualquier caso, siempre es conveniente definir la partición QM-MM tan lejos como sea posible del centro químico reactivo del sistema.

## 3.6 Dinámica Molecular

La Dinámica Molecular o MD (del inglés Molecular Dynamics) es importante en el estudio de enzimas pues permite investigar diferentes fenómenos como por ejemplo: cambios conformacionales, generación de estructuras para la modelización mecanística y también para cálculos QMMM de perfiles de energía libre de reacción, el efecto de la temperatura, y en general la naturaleza dinámica de las enzimas.<sup>176</sup> La MD se basa en las ecuaciones de movimiento de Newton para predecir la evolución temporal de la estructura de sistemas moleculares sobre una superficie de energía potencial MM o QMMM. Una dinámica molecular permite no sólo seguir los cambios conformacionales a escala atómica sino que también estudiar la estabilidad de las interacciones entre biopolímeros como las enzimas y sus sustratos.

### 3.6.1 Algoritmo de la MD

Para resolver una MD es fundamental evaluar la evolución temporal del sistema (trayectorias) mediante la resolución de la segunda ley de Newton:

$$\vec{F}_i(r_1, \dots, r_N) = m_i \frac{d^2 \vec{r}_i}{dt^2} \quad [3.45]$$

en donde  $\vec{r}_i$  y  $m_i$  indican el vector de posición y la masa de la partícula  $i$ , respectivamente.  $\vec{F}_i$  denota la fuerza que actúa sobre el átomo  $i$  y se calcula a partir de la siguiente ecuación:

$$-\frac{\partial V(\vec{r}^N)}{\partial \vec{r}_i} = \vec{F}_i \quad [3.46]$$

en donde  $V(\vec{r}^N)$  es la energía potencial de un sistema de  $N$  partículas (MM o QMMM) y  $\vec{F}_i$  es la fuerza que actúa sobre el átomo  $i$ -ésimo. La simulación de MD consiste en resolver las  $N$  ecuaciones diferenciales (ecuación 3.45) teniendo en cuenta que la fuerza viene dada por la ecuación 3.46. Para llevar a cabo esto se necesita emplear un procedimiento numérico de integración de las ecuaciones diferenciales para lo cual se suele emplear el método integrador de Verlet,<sup>190</sup> el cual propaga la posición de las partículas sin tener en cuenta las velocidades.

$$\vec{r}^{(N)}(t + \Delta t) = 2\vec{r}^{(N)}(t) - \vec{r}^{(N)}(t - \Delta t) + \vec{a}(t)(\Delta t)^2 \quad [3.47]$$

en donde  $\vec{r}^{(N)}$  es el vector posición de las  $N$  partículas,  $t$  es el tiempo,  $\Delta t$  es el paso de integración (usualmente 2 fs) y  $a$  es la aceleración o segunda derivada de la posición del

vector  $\vec{r}^{(N)}$ . De este modo, la posición de una partícula en un tiempo subsecuente ( $t + \Delta t$ ) es calculada a partir de su posición actual, su posición previa y la aceleración de la partícula; siendo esta última calculada a partir de la ecuación 3.46 y la segunda Ley de Newton. Por otra parte, si se necesita controlar la temperatura, se utiliza un algoritmo que tenga en cuenta las velocidades. En este caso, los vectores de posición y velocidad se propagan a través de otros métodos como el algoritmo de leap-frog.<sup>191</sup>

### 3.6.2 Restricciones y Paso de Integración

Los tiempos a los cuales ocurren cambios en una enzima o proteína están dentro de un rango que va desde los  $10^{-15}$  s (fluctuaciones atómicas rápidas), hasta  $10^{-7}$ - $10^4$  s en el caso de plegamiento de proteínas. Un paso de integración apropiado debería ser menor que el tiempo que tarda el movimiento más rápido del sistema que se simula. En el caso de la simulación de biomoléculas, generalmente no se suele estar interesado en los movimientos más rápidos de los átomos como las vibraciones de enlaces y esta es la razón por la cual es una práctica habitual no tener en cuenta estas vibraciones en la función de energía potencial. Esto permite aumentar la eficiencia del cálculo por medio de la restricción o fijación de las vibraciones de enlaces más rápidas, las que involucran átomos de hidrógeno, a través de algoritmos como el de SHAKE<sup>192,193</sup> o el de RATTLE.<sup>194</sup>

Las restricciones también pueden ser aplicadas para otros propósitos que incluyen los denominados eventos raros, ya que son procesos en los que se desea explorar un espacio de las fases menos estable,<sup>195</sup> o simplemente por conveniencia, por ejemplo cuando se desea fijar la posición de ciertos átomos con algún fin investigativo. Adicionalmente, de especial interés en esta tesis es la aplicación de restricciones para calcular el Potencial de Fuerza Media o PMF (del inglés Potential of Mean Force). Esta técnica se utiliza para estimar las barreras de energía libre en reacciones enzimáticas y se describirá brevemente en la sección 3.6.5.

### 3.6.3 Condiciones Periódicas de Contorno y Radios de Corte

El uso de condiciones periódicas de contorno o PBCs (del inglés Periodic Boundary Conditions) es una técnica habitual dentro de las simulaciones MD, en la que se asume que el sistema es una celda unitaria dentro de algún cristal de tipo cúbico u ortorrómbico.<sup>135</sup> Si suponemos un sistema de 1000 átomos organizados en una caja cúbica de  $10 \cdot 10 \cdot 10$ , cerca de la mitad de los átomos se ubicarán en las caras exteriores del cubo, lo cual tendrá efectos en las propiedades que se desean medir. Para evitar estos artefactos, la caja cúbica se rodea de réplicas de sí misma de modo que cada átomo interacciona con el átomo más cercano o imagen en el arreglo periódico. Así, esta técnica asegura que el sistema es periódico en todas las direcciones. La recomendación habitual respecto al tamaño de la celda de simulación, es que las dimensiones deberían ser al menos tan largas como el radio de corte o cut-off empleado en la simulación.

Por otra parte, cuando se desea investigar fenómenos confinados a una región pequeña del sistema, por ejemplo el sitio activo en una enzima, no es necesario aplicar las condiciones periódicas de contorno. En estos casos, la parte reactiva del sistema es posicionada en el centro de una esfera y los átomos que quedan dentro de esta esfera se pueden mover mientras las capas externas forman una especie de barrera que mantiene la estructura del sistema.

### 3.6.4 Condiciones Experimentales y Colectivos Termodinámicos

Los experimentos suelen llevarse a cabo a temperatura (T), volumen (V) y número de partículas constante (N), colectivo canónico NVT; o a temperatura, presión (P) y N constante, colectivo canónico NPT. Si al iniciar una simulación MD se desea mantener constante cualquiera de estas variables, es necesario aplicar técnicas que se conocen comúnmente como el termostato de Nosé-Hoover.<sup>196,197</sup> o el termostato/baróstato de Berendsen.<sup>198</sup> Las simulaciones realizadas en esta tesis aplicaron dinámica de Langevin para controlar la temperatura y el método que combina el pistón de Langevin<sup>199</sup> con el método de Nosé-Hoover para mantener la presión constante.<sup>200</sup>

### 3.6.5 Potencial de Fuerza Media

Muchas de las simulaciones MD en sistemas biológicos se hacen para calcular los cambios de energía libre a lo largo de un conjunto seleccionado de grados de libertad. De interés especial en esta tesis es el cálculo de perfiles de energía libre para calcular la diferencia de energía libre entre dos puntos de interés, lo cual permite obtener una estimación de las estabildades relativas de los correspondientes estados del sistema. En catálisis enzimática, una vez que se ha definido una coordenada de reacción apropiada, es posible determinar la barrera de energía libre de activación a través del cálculo del PMF [J.Chem.Phys.1977,23,187-199] a lo largo de esa coordenada de reacción,  $\xi(\vec{r}^{(N)})$ . Si  $P(z)$  es la densidad de probabilidad clásica evaluada en  $\xi = z$  y usando un colectivo NPT, se tiene

$$P(z) \frac{1}{A} \frac{1}{h^{3N}} \int e^{-[H(\vec{r}^{(N)}, \vec{p}^{(N)}) + PV]/k_B T} \delta[\xi(\vec{r}^{(N)}) - z] d\vec{r}^{(N)} d\vec{p}^{(N)} \quad [3.48]$$

en donde  $A$  es una constante de normalización,  $h$  es la constante de Planck,  $\vec{r}^{(N)}$  y  $\vec{p}^{(N)}$  denota el conjunto de  $3N$  coordenadas cartesianas atómicas y el momento conjugado, respectivamente;  $H$  es el Hamiltoniano nuclear clásico (cuya parte potencial es obtenido a través de cálculos QMMM),  $k_B$  es la constante de Boltzmann,  $\delta$  es función del delta de Dirac y el término PV se añade por el colectivo NPT que se usa. Notar que a la integral 3.48



sólo contribuyen las configuraciones nucleares que corresponden a  $\xi = z$ . Así, la expresión para el potencial de fuerza media en  $\xi = z$  es:

$$W(z) = -k_B T \ln P(z) \quad [3.49]$$

El potencial de fuerza media,  $W(z)$ , fue introducido por primera vez por Kirkwood<sup>201</sup> en 1935 y es un concepto clave en las teorías de mecánica estadística moderna de líquidos y macromoléculas. La barrera de energía libre de activación puede

A menudo no es práctico calcular las cantidades expresadas a través de las ecuaciones 3.48 y 3.49 directamente de la MD y para soslayar esta tarea, existen técnicas de muestreo especialmente diseñadas para calcular el PMF desde las trayectorias de MD. Una de estas técnicas es la denominada US (del inglés Umbrella Sampling),<sup>202,203</sup> la cual ha sido utilizada en esta tesis y se revisa brevemente a continuación.

### 3.6.6 Método Umbrella Sampling (US)

En este método, el sistema microscópico de interés se simula en presencia de unas ventanas artificiales sometidas a un potencial,  $w(\xi)$ , que se introducen para mejorar el muestreo configuracional en la vecindad de  $\xi$ . La idea de esta técnica es poder muestrear zonas del espacio de las fases que en una simulación de MD normal (a temperatura ambiente) no son visitadas por ser de alta energía, por ejemplo, la zona en la que aparece un estado de transición. Las simulaciones sometidas a este potencial (biased simulations) se generan usando la energía potencial  $[U(R) + w(\xi)]$ . Típicamente, los potenciales sirven para confinar las variaciones de la coordenada de reacción  $\xi$  dentro de un pequeño intervalo alrededor de un valor de referencia de  $\xi$ , lo cual ayuda a conseguir un muestreo configuracional eficiente en esta región. Una de las opciones más utilizadas para generar los colectivos sometidos a la restricción es el uso de potenciales armónicos:

$$w_i(\xi) = \frac{1}{2} K (\xi - \xi_i)^2 \quad [3.50]$$

Estos potenciales se centran en diferentes valores sucesivos de  $\xi_i$  y  $K$  determina la extensión del muestreo configuracional accesible alrededor de  $\xi_i$ .<sup>204</sup> Dado que el muestreo configuracional está confinado a pequeñas regiones durante una simulación sometida a un potencial, para obtener el PMF a lo largo de todo el rango de interés definido por  $\xi$  (ecuación 3.50), es necesario llevar a cabo simulaciones de muchas ventanas centradas en valores de  $\xi_i$  a lo largo de todo el rango abarcado por  $\xi$ . Finalmente, a los resultados de

las ventanas se les quita la contribución debida a  $w(\xi)$  y se recombinan en orden a obtener la estimación final del PMF.

Algunos de los métodos más usados para analizar los resultados del método US son por ejemplo; el método WHAM (del inglés Weighted Histogram Analysis Method),<sup>205</sup> y el método UI (del inglés Umbrella Integration).<sup>206,207</sup>

### **3.7 Metodología de Docking**

El Docking es una técnica computacional dirigida a la predicción de los modos de unión más favorables entre un ligando y un determinado receptor, así como a la estimación de la correspondiente energía libre del complejo ligando-receptor. El proceso comienza con la aplicación de algoritmos de Docking que posicionan el ligando en el sitio activo. Estos algoritmos son complementados por funciones de puntuación o SF (del inglés Scoring Functions) diseñadas para predecir la actividad biológica de los ligandos evaluados (modos de unión), a través de la evaluación de las interacciones que se llegan a establecer entre estos.

#### **3.7.1 Métodos de Búsqueda y Flexibilidad Molecular**

El tratamiento de la flexibilidad del ligando puede ser dividida en tres categorías básicas: métodos sistemáticos (búsqueda conformacional, bases de datos); métodos aleatorios o estocásticos (Monte Carlo, algoritmos genéticos); y métodos de simulación (MD, minimización de energía). Los métodos estocásticos realizan cambios aleatorios de un ligando o un conjunto de ligandos y las nuevas posiciones de los ligandos se evalúan en base a una función de probabilidad.<sup>208</sup> Algunas de las aproximaciones más populares se basan en implementaciones alternativas del método Monte Carlo,<sup>209,210</sup> por ejemplo AutoDock,<sup>209</sup> mientras que otras aproximaciones son implementadas en base a algoritmos genéticos<sup>211,212</sup> como por ejemplo, DOCK y GOLD.<sup>213</sup>

La flexibilidad de la proteína es un tópico menos avanzado que la flexibilidad del ligando pero existen varias aproximaciones que han aplicado un modelo flexible al menos en una región del receptor. Un ejemplo de esto lo proveen las librerías de rotámeros (aproximación usada en algunos experimentos de esta tesis).<sup>214,215</sup> La idea de usar librerías de rotámeros de cadenas laterales de aminoácidos, es motivada con la intención de modelizar el espacio conformacional de la proteína a partir de un número limitado de conformaciones observadas experimentalmente.<sup>215</sup> Luego, el algoritmo remueve de manera recursiva las conformaciones que no contribuyen a conseguir una estructura de mínima energía.

### 3.7.2 Funciones de Puntuación

El diseño de SF confiables es de importancia fundamental en esta metodología debido a que permiten discriminar entre posiciones correctas e incorrectas de los ligandos que se evalúan. Las SF que se implementan en los algoritmos de Docking realizan varias aproximaciones y simplificaciones para evaluar los complejos obtenidos y por consiguiente, no tienen en cuenta todas las contribuciones que determinan el reconocimiento molecular. Esencialmente, hay tres clases de SF: las basadas en campo de fuerza, las empíricas y las que se basan en conocimientos estructurales a priori, esto es, están diseñadas para reproducir estructuras experimentales más que energías de asociación.

Dentro de las SF basadas en métodos MM, los parámetros del potencial de Lennard-Jones suelen variar dependiendo de la "dureza" que se desee para el potencial. Potenciales Lennard-Jones 12-6 pueden resultar demasiado repulsivos impidiendo contactos cercanos entre receptor y ligando (D-Score).<sup>216</sup> También se pueden encontrar potenciales Lennard-Jones 8-4 que suavizan los contactos repulsivos (G-Score).<sup>216</sup> Los términos electrostáticos son tratados dentro de una formulación Coulombica, utilizando una función dieléctrica que depende de la distancia y que disminuye las contribuciones de las interacciones carga-carga. Adicionalmente, se pueden encontrar SF que incluyen términos de entropía torsional para los ligandos (G-Score), u otras que incorporan términos de enlace de hidrógeno explícitos proteína-ligando. Este es el caso GOLD<sup>217</sup> y AutoDock.<sup>211</sup>

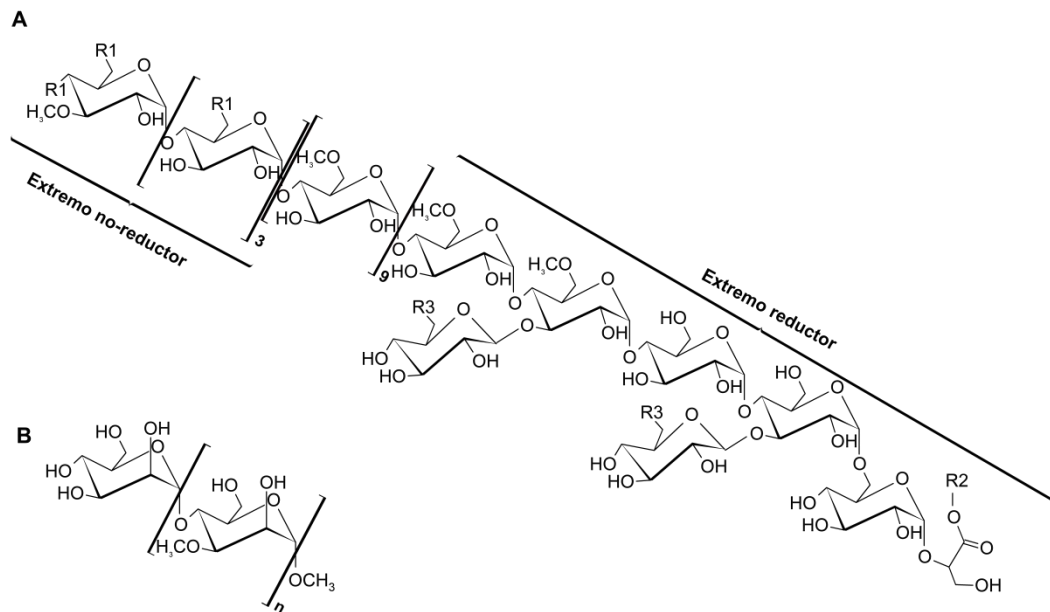
Finalmente, aunque las metodologías de Docking y las funciones de puntuación (SF) se basan en muchas simplificaciones, los cálculos de Docking han sido aplicados durante cerca de dos décadas en la investigación farmacéutica.

# 4 Glucosil-3-fosfoglicerato Sintasa: GpgS

## 4.1 Aspectos Generales

La enzima sintasa de glucosil-3-fosfoglicerato (GpgS), es una GT que participa en la biosíntesis del lipopolisacárido (LPS) de metilglucosa en micobacteria, género que incluye al patógeno causante de la tuberculosis y al de la lepra, catalizando la primera etapa de glicosilación. En esta tesis se ha estudiado dicha enzima en colaboración con el grupo de Biología Estructural del Prof. Ikerbaske Marcelo Guerin (CICbioGune, Bilbao) y con el grupo de Enzimología del Prof. Antoni Planas (IQS, Universidad Ramón Llull, Barcelona), además de la participación del grupo de Síntesis Orgánica y Catálisis Homogénea (ISQCH) del Prof. Pedro Merino en la Universidad de Zaragoza, los cuales se han encargado principalmente de la parte experimental.

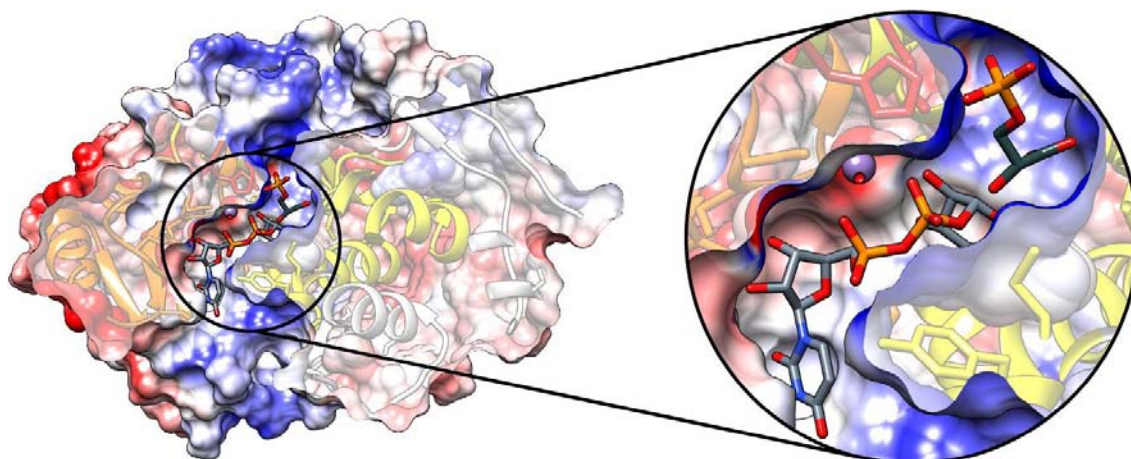
GpgS ha sido obtenida desde *Mycobacterium tuberculosis*, precisamente el microorganismo asociado a la tuberculosis.<sup>218</sup> La micobacteria sintetiza dos inusuales moléculas pertenecientes al grupo de los polisacáridos polimetilados (PSPM). *In vitro*, se ha detectado que estas moléculas tienen un rol regulador en la biosíntesis de ácidos grasos. Uno de los PSPM producidos por la micobacteria es el lipopolisacárido de 6-O-metilglucosa (MGLP), una molécula que tiene 10 monómeros de 6-O-metilglucosa conectados a través de enlaces  $\alpha$ -(1 $\rightarrow$ 4) y unidos a un final no-reductor tetrasacárido, la 3-O-metil-D-glucosa-( $\alpha$ -(1 $\rightarrow$ 4)-D-glucosa)<sub>3</sub>- $\alpha$ -(1 $\rightarrow$ ) (Figura 16A). El final reductor también es un tetrasacárido constituido por  $\rightarrow$ 4)-( $\alpha$ -(1 $\rightarrow$ 4)-D-glucosa)<sub>3</sub>- $\alpha$ -(1 $\rightarrow$ 6)-D-glucosa- $\alpha$ -(1 $\rightarrow$ ) unido a la posición 2 del ácido D-glicérico) (Figura 16A). El otro PSPM generado es el polisacárido de 3-O-metilmanosa o MMP (Figura 16B), cuyo final reductor es una aglucona, el grupo  $\alpha$ -metil. Es importante mencionar que a los PSPM también se les han atribuido roles como el de servir de protectores de ácidos grasos contra la degradación y como transportadores de éstos, facilitando la síntesis de moléculas muy grandes e insolubles como el éster de ácido micólico.<sup>219</sup>



**Figura 16.** Estructura química de: A: lipopolisacárido 6-O-metilglucosa (MGLP) procedente de *Mycobacterium bovis* BCG. Los sustituyentes R1, R2 y R3 representan a grupos acilos que en el caso de R1 puede corresponder a un acetato, propionato o isobutirato; R2 es un octanoato; y R3 es succinato.<sup>71</sup> B: polisacárido de 3-O-metilmanosa o MMP, compuesto de 10-13 (n) unidades de 3-O-metilmanosa conectadas por enlaces α-(1→4).

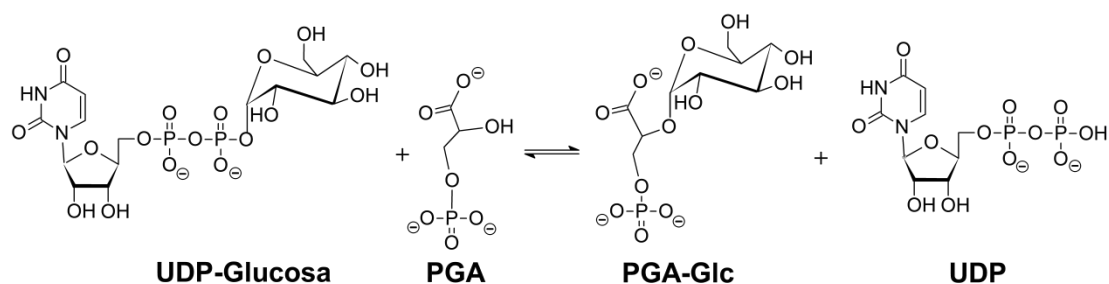
El laboratorio del Prof. Guerin ha resuelto varias estructuras cristalográficas para GpgS, publicadas en nuestro trabajo conjunto. A diferencia de las estructuras de GpgS que se habían publicado anteriormente, en este caso se ha conseguido cristalizar en una única estructura la enzima nativa junto a sus sustratos naturales, es decir, el complejo ternario nativo.<sup>220</sup> Se trata del primer complejo ternario nativo resuelto para una GT-ret.

GpgS exhibe un plegamiento GT-A, es miembro de la familia GT81 y al igual que la mayoría de los miembros de esta superfamilia, utiliza un catión divalente como cofactor esencial para llevar a cabo la catálisis. Aunque diversos experimentos han probado que GpgS es activa para un amplio rango de metales divalentes como el Mg<sup>2+</sup>, Mn<sup>2+</sup>, Ca<sup>2+</sup>, Co<sup>2+</sup> y Fe<sup>2+</sup>, la actividad transferasa *in vitro* es maximizada usando Mg<sup>2+</sup>.<sup>71</sup> En cada uno de los monómeros de GpgS es posible observar un túnel bien definido y cargado positivamente a través del cual se podrían unir los sustratos UDP-Glucosa (UDP-Glc) y 3-fosfoglicerato (PGA). De hecho, los mapas de densidad electrónica del catión divalente y ambos sustratos son nítidos y los muestran en el centro del túnel en el que la reacción química ocurre (Figura 17).



**Figura 17.** Localización del sitio activo de GpgS. Cada monómero muestra una cavidad o túnel cargado positivamente (azul), en el que se alojan los sustratos y el metal.

GpgS cataliza la transferencia de glucosa, desde UDP-Glc hasta la posición 2 del 3-fosfoglicerato (PGA), formando glucosil-3-fosfoglicerato (Glc-PGA) (Figura 18).



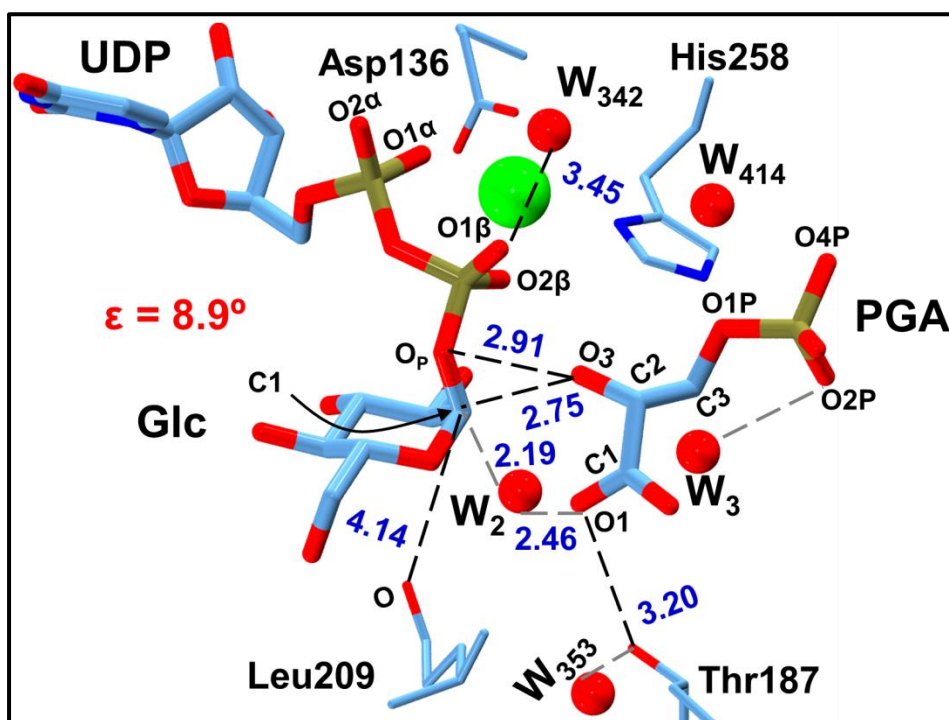
**Figura 18.** GpgS transfiere una glucosa desde UDP-Glc hasta la posición 2 del 3-fosfoglicerato (PGA), formando PGA-Glc.

La inspección de los residuos del sitio activo no permite identificar un posible residuo nucleofílico que pueda formar el intermedio CGE en lo que constituiría la primera etapa del mecanismo de doble desplazamiento. Debido a esto, es esperable que la reacción ocurra a través del mecanismo de sustitución nucleofílica tipo- $S_Ni$  o de ataque frontal. Curiosamente, en algunas de las estructuras de complejos ternarios obtenidas por el grupo del Prof. Guerin, aparecían unas densidades electrónicas cercanas al carbono anomérico de la Glucosa y del grupo carboxilato de PGA. Estas posibles densidades podrían relacionarse con la presencia de moléculas de agua en esta zona. De hecho, si estas densidades son asignadas, una de las posibles moléculas de agua queda bastante cerca del carbono anomérico (apartado 4.1.1), lo cual nos llevó a pensar en la posibilidad de que la molécula de agua actúe como nucleófilo en un mecanismo de doble desplazamiento, siendo activada por el propio carboxilato de PGA. En este capítulo presentamos el primer estudio QMMM del mecanismo de reacción de GpgS, enfocando la discusión en posibles estrategias mecanísticas, así como en el rol que diferentes residuos del sitio activo, incluidas las posibles moléculas de agua, podrían desempeñar en tales estrategias. La discusión se basa

en varios modelos que hemos construido con el objetivo de indagar tanto como sea posible en las características estructurales que modulan las barreras energéticas obtenidas

#### 4.1.1 Datos Cristalográficos de GpgS: Conformaciones de PGA y Posibles Aguas en el Sitio Activo

El comienzo de la colaboración con el Prof. Guerin para el estudio de GpgS se remonta a unos años atrás, durante la realización de una tesis previa en el grupo. Aquellos estudios computacionales se iniciaron con estructuras cristalográficas de complejos binarios pero ante la importante novedad de disponer de la estructura del complejo ternario nativo de GpgS (GpgS·Mn<sup>2+</sup>·UDP-Glc·PGA), no progresaron. En su lugar, se iniciaron nuevos estudios utilizando la nueva estructura. En la primera estructura que resolvieron del complejo ternario nativo (información cristalográfica no publicada, Figura 19), el O3<sub>PGA</sub> se encuentra a 2.75 Å del C1<sub>Glc</sub> y a 2.91 Å de O<sub>P</sub>. Se observa que la molécula de agua de coordinación W<sub>342</sub> se ubica a 3.45 Å del O1<sub>β</sub>, y además hay una molécula de agua cristalográfica (W<sub>414</sub>) que aparece a 2.78 Å de O4<sub>P</sub> y a 4.52 Å de O1<sub>β</sub>. El oxígeno nucleofílico de PGA (O3) se dispone de tal manera que prácticamente forma un plano con los átomos C2, C1 y O1 de PGA, un ángulo diedro que se puede identificar con la letra ε (Figura 19).



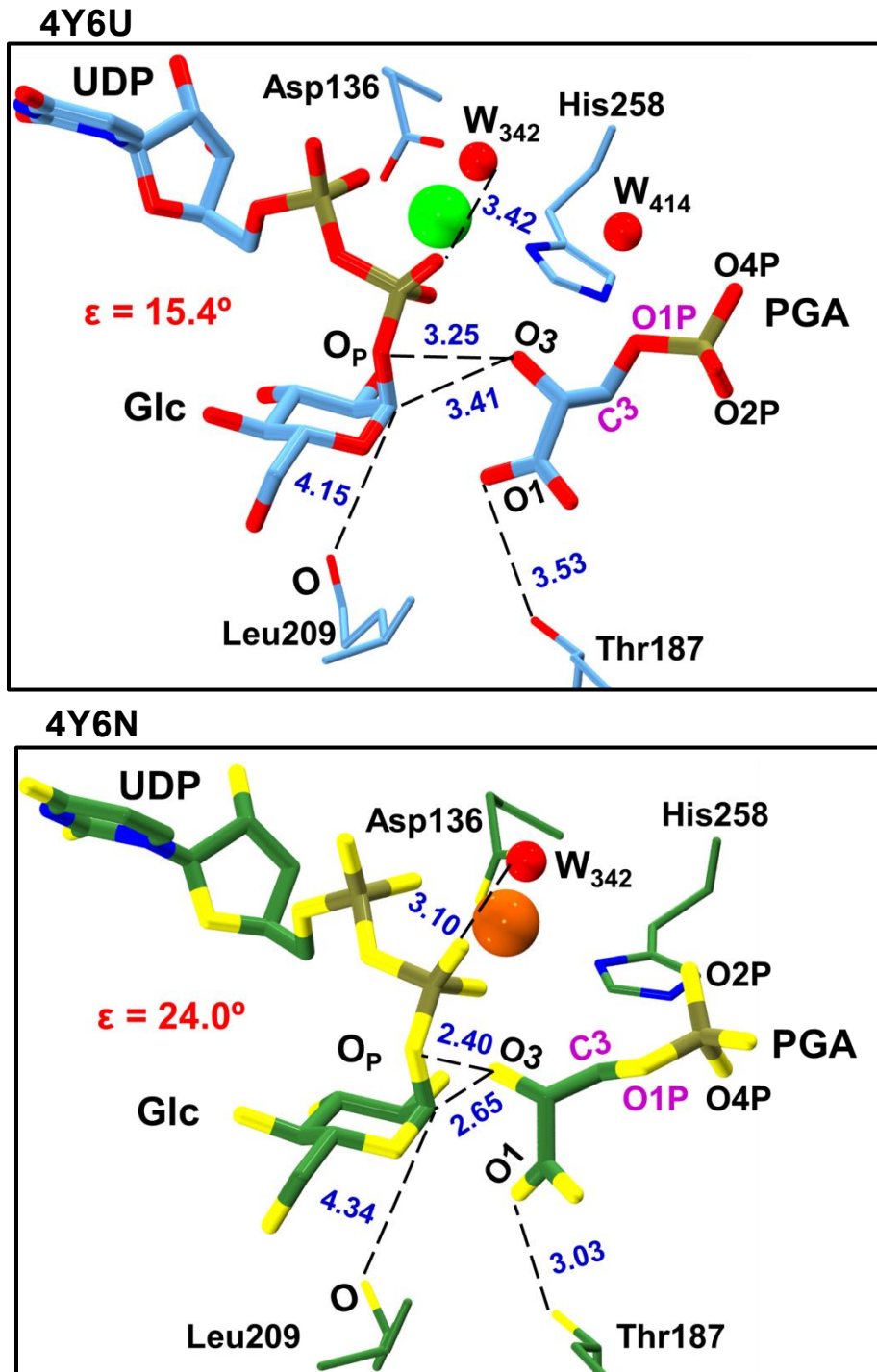
**Figura 19.** Representación de la **estructura 0** ó complejo ternario GpgS·Mn<sup>2+</sup>·UDP-Glc-PGA a través de esferas y barras (modelo cristalográfico generado a partir de datos no publicados). Se señalan los nombres de algunas de las moléculas que forman parte de sitio activo; el sustrato dador o UDP-Glc; el sustrato aceptor o PGA; Leu209, en la cara β de Glc; Thr187, en posición de interacción con O1<sub>PGA</sub>; y las 3 posibles moléculas de agua (W<sub>2</sub>, W<sub>3</sub> y W<sub>353</sub>). También se dibuja el catión metálico (esfera verde) y su primera esfera de coordinación que incluye a la molécula de agua W<sub>342</sub>, el Asp136, la His258 y dos oxígenos del grupo difosfato (O2<sub>β</sub> y O1<sub>α</sub>).

Adicionalmente, se dibuja la molécula de agua  $W_{414}$  que está en posición de interactuar con el grupo fosfato de PGA a través de  $O_{4P}$ . Por último, se indican también algunos identificadores de átomos, distancias relevantes (Å) y un ángulo de torsión que resulta útil para definir la conformación de PGA.

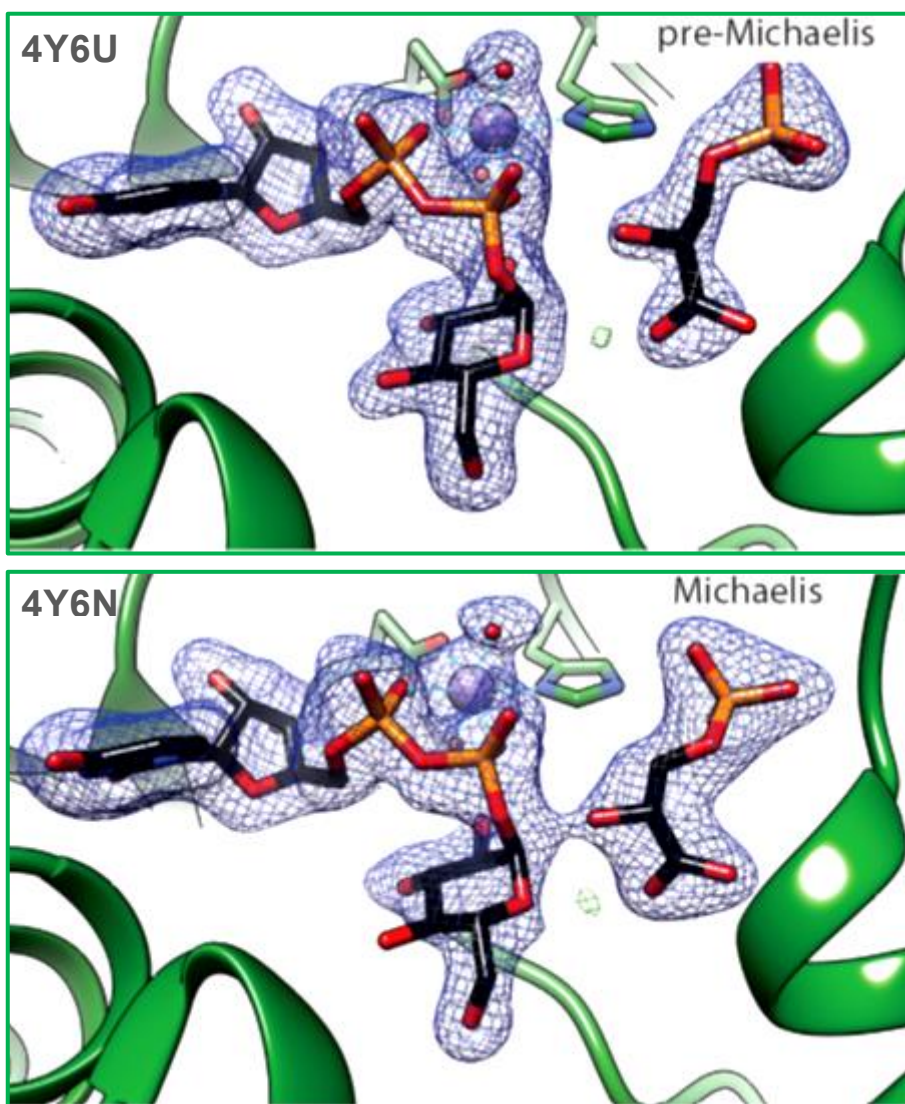
Al analizar la cara  $\beta$  de la Glucosa, se encuentra a la Leu209 con su oxígeno carbonílico apuntando hacia el centro del anillo, a 4.14 Å del carbono anomérico. También se incluye a la Thr187, que es el único residuo en posición de interactuar directamente con el grupo carboxilato de PGA ( $COO^-_{PGA}$ ). Resulta curioso que sólo la Thr187 interactuara con éste teniendo en cuenta su carga neta negativa. Los datos de difracción de rayos X indican unas densidades electrónicas bastante cercanas al carbono anomérico de Glc y al  $COO^-_{PGA}$ . Esto sugiere la presencia de moléculas de agua, a las cuales nos referiremos como  $W_2$ ,  $W_3$  y  $W_{353}$ ; la  $W_2$  se ubicaba a tan sólo 2.19 Å de  $C1_{Glc}$  (según modelo cristalográfico, Figura 19) y a 2.46 Å del  $O1_{PGA}$ . Las otras dos moléculas de agua quedaban en posiciones más cercanas al  $COO^-_{PGA}$  que al C1; la molécula de agua  $W_{353}$  se ubica a 2.54 Å del oxígeno del grupo hidroxilo de Thr187 mientras que  $W_3$ , está a 3.71 Å y 4.23 Å del  $O2P_{PGA}$  y  $O1_{PGA}$ , respectivamente (Figura 19). La posibilidad de una molécula de agua en la vecindad del carbono anomérico hizo resurgir la alternativa de un mecanismo de doble desplazamiento para la formación del enlace glicosídico. En este caso, a falta de un residuo nucleofílico enzimático, la molécula  $W_2$  podría ejercer este rol siendo activada por el mismo grupo carboxilato de PGA a través de  $O1$ . Así pues, se prosiguió con el estudio computacional que dio como resultado el Modelo 0 (apartado 4.2.2) y el Modelo 1 (apartado 4.2.5).

Paralelamente, el grupo del Prof. Guerin continuó con el refinamiento de estructuras consiguiendo nuevos complejos ternarios con variaciones en la conformación de PGA y en estas densidades que podrían indicar la presencia de aguas. Destacan dos estructuras, que denominaron complejo de pre-Michaelis (código PDB 4Y6U) y de Michaelis (código PDB 4Y6N), en referencia a la distancia  $O3_{PGA}-C1_{Glc}$  que presentan: 3.4 y 2.6 Å, respectivamente. La resolución a la que se obtuvieron dichas estructuras cristalográficas (> 2.0 Å) no permitía asignar con certeza esta densidad electrónica a las moléculas de agua, por lo que las estructuras finalmente publicadas y depositadas en la base de datos PDB (4Y6U y 4Y6N), no contemplan la presencia de estas moléculas de agua (Figura 20 y 21).





**Figura 20.** Representación del complejo de pre-Michaelis (4Y6U y carbonos celestes) y del complejo de Michaelis (4Y6N, carbonos verdes). Se indican las diferencias estructurales más importantes, por ejemplo: la orientación relativa de los átomos C3 y O1P de PGA,  $\epsilon$  y las distancias inter-sustratos (Å) en cada caso.



**Figura 21.** Mapas de densidad electrónica ( $2mF_o-DF_c$ ) para el complejo de pre-Michaelis GpgS·Mn<sup>2+</sup>·UDP-Glc·PGA-2 contorneado a  $1\zeta$  (código PDB 4Y6U), y el complejo de Michaelis GpgS·Mn<sup>2+</sup>·UDP-Glc·PGA-1 (PDB 4Y6N) contorneado a  $1\zeta$

Con el objetivo de estudiar el mecanismo de GpgS en ausencia de estas moléculas de agua y analizar las implicaciones mecanísticas que esto conllevaría, se construyó el Modelo 2, para el que también se realizó el estudio computacional completo y cuyos resultados se presentan más adelante en este capítulo.

Cabe destacar que la estructura cristalográfica utilizada para la construcción del modelo computacional inicial (Modelo 0 a partir de la estructura 0 (Figura 19), ver apartado 4.2.2), presenta los sustratos en conformaciones parecidas a las observadas en el complejo de pre-Michaelis 4Y6U (Figura 20). De acuerdo con este hecho, las barreras de energía obtenidas a partir del Modelo 0 indican que las interacciones del sitio activo aún no son óptimas para catalizar la reacción. La continuación del trabajo de modelización que realizamos, nos condujo a una estructura de complejo ternario catalíticamente competente

(Modelo 1, apartado 4.2.3), en el que UDP-Glc y PGA están dispuestos de forma muy semejante al complejo de Michaelis 4Y6N (Figura 20). En el caso de PGA, esto ocurre más marcadamente a partir de C3<sub>PGA</sub>, el átomo enlazado al grupo fosfato de PGA. Además, la distancia C1<sub>Glc</sub>-O<sub>P</sub> en el Modelo 1 y el complejo de Michaelis 4Y6N son 2.91 Å y 2.65 Å, respectivamente, distancias < 3.0 Å que no habían sido observadas en los modelos anteriores. Estas similitudes nos permitieron juzgar que este modelo computacional, y los resultados que habíamos obtenido con él, eran válidos para el estudio mecanístico y no se construyó un modelo adicional basado en 4Y6N. De todos modos, como se explicará más adelante, se generó un complejo ternario *in silico* (Híbrido 1.B) para comprobar la validez de dicha aproximación (apartado 4.2.7).

## 4.2 Modelos y Metodología

### 4.2.1 Construcción del Modelo Inicial

El modelo inicial (Modelo 0) fue obtenido en base a la estructura cristalográfica de un complejo ternario nativo de GpgS, GpgS·Mn<sup>2+</sup>·UDP-Glc·PGA (estructura 0, Figura 19). Como se acaba de comentar, por una cuestión cronológica no se corresponde exactamente a ninguna de las estructuras finalmente depositadas por el grupo del Prof. Guerin en la base de datos PDB, pero es muy parecida al complejo de pre-Michaelis GpgS·Mn<sup>2+</sup>·UDP-Glc·PGA-2 (código PDB 4Y6U, Figura 20),<sup>220</sup> aunque con una molécula de agua (W<sub>2</sub>) cerca del carbono anomérico (Figura 19). Específicamente, la densidad electrónica sugería la presencia de una molécula de agua a 2.19 Å de C1<sub>Glc</sub>; una posible segunda molécula de agua (W<sub>3</sub>) aparecía a 4.23 Å del carboxilato de PGA, pero a más de 5 Å del carbono anomérico; y una tercera molécula de agua (W<sub>353</sub>) a 4.63 Å del carboxilato de PGA e interaccionando con la Thr187 que está a 3.20 Å de PGA (Figura 19). En esta estructura, los lazos que contienen a los residuos 167-182 y 297-300 no son visibles y por este motivo, utilizamos las coordenadas generadas por el grupo del Prof. Antoni Planas (Universidad Ramon Llull), que ellos obtuvieron modelizando dichos lazos con el programa MODELLER<sup>221</sup> y llevando a cabo una simulación MD de varios centenares de ns. De esta manera, la construcción del modelo incluye diferentes contribuciones que dejan en evidencia la complejidad de esta enzima que deriva de la plasticidad asociada a las GTs-ret. Algunos de los procedimientos necesarios para la construcción del modelo, son la asignación de parámetros, topologías y estados de protonación. Así, con la excepción de la His258 a la cual le fue asignado el estado de protonación HSE, el estado de protonación del resto de los aminoácidos ionizables fue asignado a través del procedimiento PROPKA3.1<sup>222-224</sup> implementado en el servidor PDB2PQR.<sup>225,226</sup> El Mn<sup>2+</sup> fue reemplazado por Mg<sup>2+</sup> ya que su tratamiento computacional es más sencillo y tal reemplazo ha sido validado en trabajos anteriores de nuestro grupo<sup>125</sup> y de otros grupos también.<sup>116,122</sup> Además, en el caso de GpgS la actividad transferasa es maximizada usando este metal.<sup>71</sup> La topología, carga y parámetros de la proteína, Mg<sup>2+</sup>, moléculas de agua y uridina difosfato (UDP) corresponden a los del campo de fuerza CHARMM22,<sup>141,227,228</sup> y en caso de los carbohidratos fue usada la versión CHARMM36 del campo de fuerza que ya incluye

parámetros para las uniones glicosídicas.<sup>229-231</sup> La topología y parámetros del sustrato aceptor PGA, se consiguieron a través de la página web ParamChem<sup>232</sup> que usa el campo de fuerza Charmm Generalized Force Field (CGenFF) versión 2b8,<sup>233,234</sup> y el programa CGenFF versión 0.9.7.1.<sup>235,236</sup> Los parámetros utilizados para PGA se muestran en las siguientes tablas.

**Tabla 3.** Se indican los nombres de cada uno de los átomos que representan a la molécula de PGA (Columna: Nombre de átomo), junto a el tipo de átomo que le corresponde dentro del campo de fuerzas CHARMM. Además, se dan las cargas (u.a.).

| Nombre de átomo | Tipo de átomo CHARMM | Carga     |
|-----------------|----------------------|-----------|
| C1              | CG203                | 0.634000  |
| O1              | OG2D2                | -0.758000 |
| O2              | OG2D2                | -0.758000 |
| C2              | CG311                | 0.046000  |
| O3              | OG311                | -0.621000 |
| C3              | CG321                | -0.182000 |
| O1P             | OG303                | -0.403000 |
| P               | PG2                  | 1.099000  |
| O2P             | OG2P1                | -0.899000 |
| O3P             | OG2P1                | -0.899000 |
| O4P             | OG2P1                | -0.899000 |
| H2              | HGA1                 | 0.090000  |
| H32             | HGA2                 | 0.090000  |
| H33             | HGA2                 | 0.090000  |
| HO3             | HGP1                 | 0.370000  |

En las siguientes tablas se dan los parámetros del campo de fuerzas para los enlaces, ángulos de enlace y ángulos diedros.

**Tabla 4.** Parámetros del campo de fuerzas asociados a los enlaces a través de la siguiente ecuación:  $V_B = k_B (b - b_o)^2$ . En esta expresión,  $k_B$  es la constante de fuerza del enlace y  $b_o$  es la distancia de enlace de equilibrio.

| Tipo de átomo | $k_B$ [kcal/(mol·Å <sup>2</sup> )] | $b_o$ [Å] |
|---------------|------------------------------------|-----------|
| CG203 CG311   | 200.00                             | 1.520     |
| CG203 OG2D2   | 525.00                             | 1.260     |
| CG311 CG321   | 222.50                             | 1.538     |
| CG311 OG311   | 428.00                             | 1.420     |
| CG311 HGA1    | 309.00                             | 1.111     |
| CG321 OG303   | 320.00                             | 1.440     |
| CG321 HGA2    | 309.00                             | 1.111     |
| OG2P1 PG2     | 400.00                             | 1.520     |
| OG303 PG2     | 150.00                             | 1.655     |
| OG311 HGP1    | 545.00                             | 0.960     |

**Tabla 5.** Parámetros del campo de fuerzas asociados a los ángulos de enlace por medio de las dos ecuaciones siguientes:  $V_\theta = k_\theta (\theta - \theta_o)^2$  y  $V_{UB} = k_{UB} (S - S_o)^2$  en las que  $k_\theta$  es la constante de fuerza del ángulo de enlace;  $\theta_o$  es el valor de equilibrio del ángulo;  $k_{UB}$  es la constante de fuerza del potencial Urey-Bradley; y  $S_o$  es la distancia (en Å) entre los átomos 1-3.

| Tipo de átomo |       |       | $k_\theta$ [kcal/(mol·rad <sup>2</sup> )] | $\theta_o$ [°] | $k_{UB}$ [kcal/(mol·Å <sup>2</sup> )] | $S_o$ [Å] |
|---------------|-------|-------|---|----------------|---------------------------------------|-----------|
| CG311         | CG2O3 | OG2D2 | 40.00                                     | 116.00         | 50.00                                 | 2.35300   |
| OG2D2         | CG2O3 | OG2D2 | 100.00                                    | 128.00         | 70.00                                 | 2.25870   |
| CG2O3         | CG311 | CG321 | 52.00                                     | 108.00         |                                       |           |
| CG2O3         | CG311 | OG311 | 45.00                                     | 109.00         |                                       |           |
| CG2O3         | CG311 | HGA1  | 50.00                                     | 109.50         |                                       |           |
| CG321         | CG311 | OG311 | 75.70                                     | 110.00         |                                       |           |
| CG321         | CG311 | HGA1  | 34.50                                     | 110.10         | 22.53                                 | 2.17900   |
| OG311         | CG311 | HGA1  | 45.90                                     | 108.89         |                                       |           |
| CG311         | CG321 | OG303 | 75.70                                     | 110.10         |                                       |           |
| CG311         | CG321 | HGA2  | 33.43                                     | 110.10         | 22.53                                 | 2.17900   |
| OG303         | CG321 | HGA2  | 60.00                                     | 109.50         |                                       |           |
| HGA2          | CG321 | HGA2  | 35.50                                     | 109.00         | 5.40                                  | 1.80200   |
| CG321         | OG303 | PG2   | 20.00                                     | 120.00         | 35.00                                 | 2.33000   |
| CG311         | OG311 | HGP1  | 50.00                                     | 106.00         |                                       |           |
| OG2P1         | PG2   | OG2P1 | 104.00                                    | 121.00         |                                       |           |
| OG2P1         | PG2   | OG303 | 88.90                                     | 111.00         |                                       |           |

A continuación, se dan los parámetros para los ángulos de torsión o diedros descritos por la siguiente ecuación  $V_{Diedro}$ .

**Tabla 6.** Parámetros para los ángulos diedros descritos por la siguiente ecuación:  $V_{Diedro} = k_\xi (1 + \cos(n\xi - \delta))$ , en la que  $k_\xi$  es la constante de fuerza del ángulo diedro,  $n$  es la multiplicidad de la función,  $\xi$  es el ángulo diedro y  $\delta$  es el desplazamiento de la fase.

| Tipo de átomo |       |       |       | $k_\xi$ [kcal/mol] | $n$ | $\delta$ [°] |
|---------------|-------|-------|-------|--------------------|-----|--------------|
| OG2D2         | CG2O3 | CG311 | CG321 | 0.0500             | 6   | 180.00       |
| OG2D2         | CG2O3 | CG311 | OG311 | 0.5500             | 2   | 180.00       |
| OG2D2         | CG2O3 | CG311 | HGA1  | 0.0500             | 6   | 180.00       |
| CG2O3         | CG311 | CG321 | OG303 | 0.2000             | 3   | 0.00         |
| CG2O3         | CG311 | CG321 | HGA2  | 0.2000             | 3   | 0.00         |
| OG311         | CG311 | CG321 | OG303 | 0.2000             | 3   | 0.00         |
| OG311         | CG311 | CG321 | HGA2  | 0.1950             | 3   | 180.00       |
| HGA1          | CG311 | CG321 | OG303 | 0.1950             | 3   | 0.00         |
| HGA1          | CG311 | CG321 | HGA2  | 0.1950             | 3   | 0.00         |
| CG2O3         | CG311 | OG311 | HGP1  | 0.2200             | 1   | 0.00         |
| CG2O3         | CG311 | OG311 | HGP1  | 0.2300             | 2   | 180.00       |
| CG2O3         | CG311 | OG311 | HGP1  | 0.4200             | 3   | 0.00         |
| CG321         | CG311 | OG311 | HGP1  | 0.3000             | 1   | 0.00         |
| CG321         | CG311 | OG311 | HGP1  | 0.3000             | 3   | 0.00         |

|      |       |       |       |        |        |          |
|------|-------|-------|-------|--------|--------|----------|
| HGA1 | CG311 | OG311 | HGP1  | 0.0000 | 3      | 0.00     |
|      | CG311 | CG321 | OG303 | PG2    | 0.6000 | 1 180.00 |
|      | CG311 | CG321 | OG303 | PG2    | 0.6500 | 2 0.00   |
|      | CG311 | CG321 | OG303 | PG2    | 0.0500 | 3 0.00   |
|      | HGA2  | CG321 | OG303 | PG2    | 0.0000 | 3 0.00   |
|      | CG321 | OG303 | PG2   | OG2P1  | 0.1000 | 3 0.00   |

Después de asignar los parámetros y topología, el homodímero fue solvatado en una caja de moléculas de agua descritas por el modelo TIP3P;<sup>237</sup> las dimensiones de la caja de agua fueron de  $111 \cdot 82 \cdot 72 \text{ \AA}^3$ , y 26 cationes  $\text{Na}^+$  fueron añadidos para neutralizar la carga del sistema que quedó conformado por 62571 átomos. Finalmente, el sistema completo fue minimizado para liberar cualquier choque estérico. Cabe recordar que en este modelo inicial está incluida la molécula de agua  $W_2$  que se encuentra cerca de la cara  $\beta$  del azúcar, muy próxima al carbono anomérico (apenas  $2.19 \text{ \AA}$ ). Como se ha comentado, la posible presencia de una molécula de agua en esta posición inició la sospecha de que pudiera estar involucrada en la reacción. Debido a que la distancia cristalográfica es bastante pequeña, se tuvo especial cuidado en el proceso de minimización inicial, para lo cual decidimos combinar la aproximación QMMM y MM antes de proceder a la dinámica molecular. Los detalles se dan a continuación.

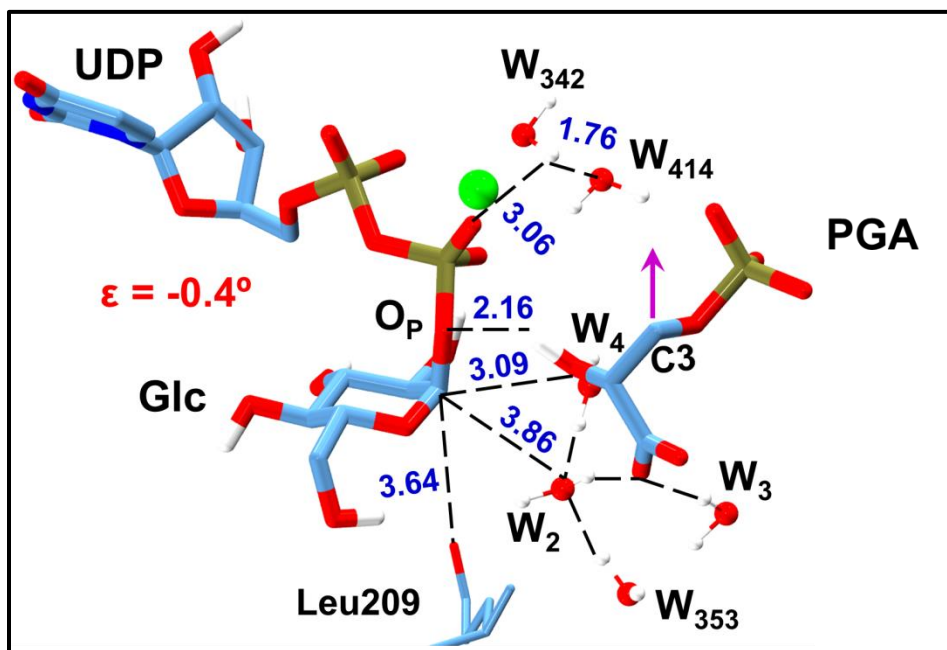
#### 4.2.2 Minimizaciones y Simulaciones de Dinámica Molecular: Obtención del Modelo 0

La minimización del sistema inicial contempló 2000 pasos de optimización (algoritmo de gradientes conjugados) únicamente para los átomos de hidrógeno y las aguas de solvatación, dejando el resto del sistema fijo. Después de la minimización, se continuó con una MD de 10 ps, utilizando para ambas el campo de fuerzas CHARMM. A continuación, se empleó la aproximación QM(BP86/SVP)/MM(CHARMM) - los detalles de la partición QM y MM se darán en el apartado 4.2.3 - para minimizar sólo los hidrógenos de PGA y UDP-Glc, y las posibles moléculas de agua cristalográficas que se encontraban en la vecindad del carbono anomérico de Glc ( $W_2$ ,  $W_3$  y  $W_{353}$ ); esto se hizo primero con el monómero A y luego con el monómero B. Lo que se explica a continuación, fue realizado manteniendo este orden. Se prosiguió con la minimización de todos los átomos dentro de un radio de  $10 \text{ \AA}$  con respecto al carbono anomérico ( $C1_{\text{Glc}}$ ), dejando fijos los átomos exteriores a la esfera de  $10 \text{ \AA}$ ; además, se aplicó una restricción armónica con una constante de fuerza de  $2 \text{ Hartree} \cdot \text{Bohr}^{-2}$  para la distancia  $C1_{\text{Glc}}-O_{W2}$ , evitando de esta manera un distanciamiento brusco entre estos dos átomos que se encuentran a tan sólo  $2.19 \text{ \AA}$ . Después se minimizaron los átomos exteriores a la esfera de  $10 \text{ \AA}$  y se finalizó con la relajación completa del sistema, primero a nivel QM(BP86/SVP)/MM(CHARMM) y después a nivel MM(CHARMM). La estructura obtenida se usó como punto de partida para la simulación de MD. Este complejo fue gradualmente calentada hasta  $300 \text{ K}$ , aplicando potenciales armónicos para restringir el movimiento de los átomos pesados de la proteína, sustratos y las aguas en la vecindad del carbono anomérico. Las restricciones posicionales sobre los

átomos de la proteína fueron paulatinamente disminuidas hasta cero, seguido de una MD de 500 ps. Las restricciones posicionales sobre el resto del sistema fueron también gradualmente reducidas, extendiendo la simulación hasta 8 ns. La molécula de agua  $W_2$  se mantuvo a una distancia promedio de  $3.99 \pm 0.50$  Å del carbono anomérico ( $C1_{Glc}$ ) y en los primeros 2 ns de la MD, otra molécula de agua ( $W_4$ ) se aproximó al sitio activo, permaneciendo ahí hasta el final de la simulación. En total tres moléculas de agua,  $W_2$ ,  $W_3$  y  $W_4$ , permanecen invariablemente cerca de la cara  $\beta$  del azúcar interactuando con  $COO^-_{PGA}$ . Además, la molécula agua  $W_{353}$  permanece formando un enlace de hidrógeno con  $W_2$  pero no con  $PGA$  (Figura 22).

Las minimizaciones y dinámicas realizadas bajo la aproximación MM se llevaron a cabo con el programa NAMD2.9,<sup>238,239</sup> para los análisis estructurales, visualización y generación de figuras se utilizó el programa VMD.<sup>240</sup> Se aplicaron condiciones Periódicas de Contorno, dinámica de Langevin para controlar la temperatura y el método que combina el pistón de Langevin<sup>199</sup> y el método de Nosé-Hoover<sup>200</sup> para mantener el sistema a 1 bar de presión. Se aplicó el algoritmo SHAKE<sup>193</sup> sobre todas las distancias O-H de moléculas de agua,<sup>192</sup> lo que permitió utilizar un paso de integración de 2 fs para la integración de las ecuaciones de movimiento.

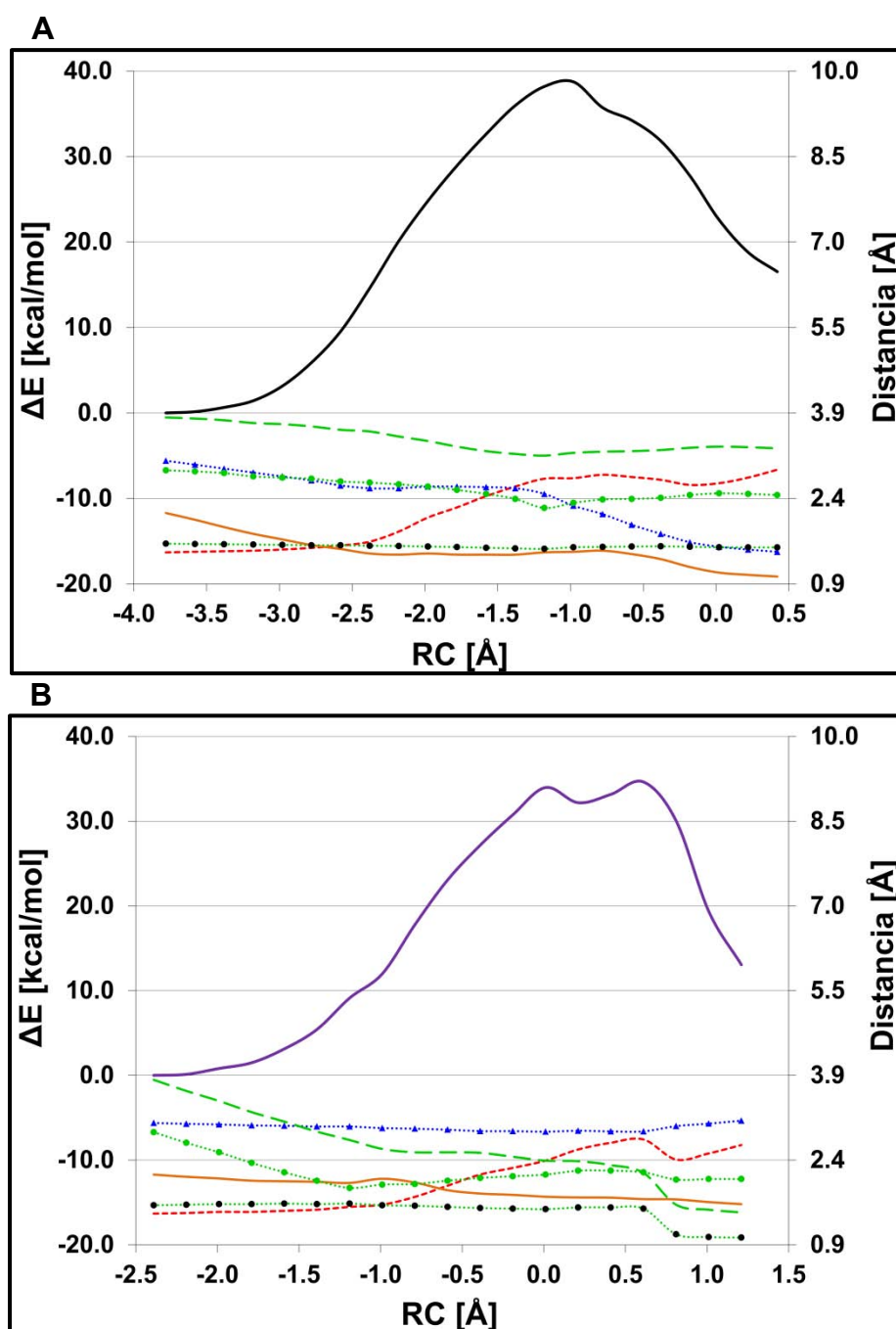
Varias estructuras de la simulación de dinámica molecular fueron seleccionadas para evaluar la reactividad. El criterio de selección fue establecido en base a la distancia del nucleófilo al carbono anomérico ( $C1_{Glc}-O3_{PGA} \leq 3.5$  Å), de modo que todos los complejos seleccionados tuvieran una distancia que pudiera ser adecuada para la reacción. Tanto la posibilidad de un mecanismo de ataque frontal, como la del ataque de la molécula de agua ( $W_2$ ) por la cara  $\beta$  de la Glc fueron investigados mediante optimizaciones restringidas a lo largo de una coordenada de reacción adecuada (los detalles metodológicos de estos cálculos se dan en el siguiente apartado 4.2.3). En todos los casos, las barreras de energía potencial calculadas a nivel QM(BP86/SVP)/MM(CHARMM) eran superiores a 34 kcal/mol, un valor bastante más alto que el derivado experimentalmente a partir de una  $k_{cat}$  de  $800 \text{ min}^{-1}$  (310 K), que corresponde a un valor de energía libre de activación fenomenológica de 17.5 kcal/mol. Estos resultados sugieren que nuestro modelo inicial no es apto para la catálisis pues no estabiliza suficientemente el estado de transición y los productos. Como muestra de estos resultados, a continuación se dan los correspondientes a uno de los complejos seleccionados, que denominaremos Modelo 0 o modelo no catalítico.



**Figura 22.** Representación del Modelo 0 o modelo no catalítico obtenido a partir de la estructura 0 (Figura 19). Se señalan las distancias inter-sustratos más importantes en Å, así como aquellas que involucran a  $W_2$ , Leu209 y Thr187. Adicionalmente, se dibuja una flecha apuntando hacia arriba para recalcar la orientación de  $C3_{PGA}$ . Los hidrógenos apolares no se muestran.

La disposición del grupo fosfato de PGA en el Modelo 0, es similar a la que se observa en el complejo de Michaelis 4Y6N (Figura 25). Sin embargo, a partir de  $C3_{PGA}$  y continuando con el hidroxilo nucleofílico y el  $COO^-_{PGA}$ , aparecen diferencias en la conformación de PGA que se pueden reflejar a través de  $\epsilon$ . Este diedro tiene un valor de  $-0.4^\circ$  en el Modelo 0 y  $24.0^\circ$  en el complejo de Michaelis 4Y6N. Por otra parte, la distancia  $C1_{Glc}-O3_{PGA}$  en el Modelo 0 es de 3.09 Å, un valor semejante al que se encuentra en el complejo de pre-Michaelis 4Y6U (3.25 Å), pero más largo que el encontrado en el complejo de Michaelis 4Y6N, que es de apenas 2.65 Å (Figura 20). La  $W_2$  se encuentra a 3.86 Å del carbono anomérico y formando un fuerte enlace de hidrógeno con  $O1_{PGA}$  a una distancia de 1.61 Å. Además, interacciona fuertemente con  $W_4$  (1.75 Å) y  $W_{353}$  (1.76 Å) pero no con la  $W_3$ , la cual se mantiene interaccionando con  $O1_{PGA}$  a una distancia de 1.87 Å (Figura 22). A continuación, se presentan las curvas de energía potencial a nivel QM(BP86/SVP)/MM(CHARMM) que se han obtenido para el mecanismo de ataque frontal y ataque de  $W_2$ , usando como base catalítica el  $O1_{PGA}$ . Las coordenadas de reacción que se utilizaron para simular estas reacciones químicas se dan en el siguiente apartado 4.2.3.





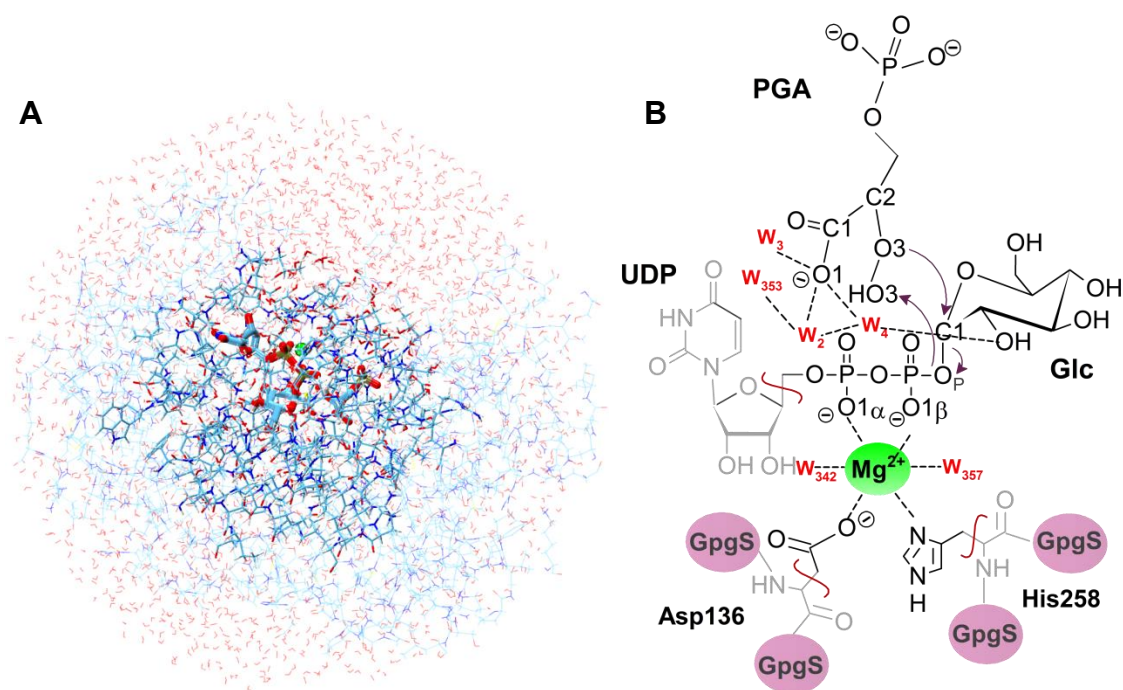
**Figura 23.** Curvas de energía potencial a nivel QM(BP86/SVP)/MM(CHARMM) para A: el mecanismo de ataque frontal (línea continua negra) y B: ataque de la molécula de agua  $W_2$  con  $O_{1PGA}$  ejerciendo el rol de base catalítica y activando a la  $W_2$  (línea morada). También se señalan algunas distancias importantes como:  $C_{1Glc}-O_P$  en línea discontinua roja,  $C_{1Glc}-O_{3PGA}$  en línea punteada azul,  $HO_{3PGA}-O_P$  en línea naranja,  $H_{1Glc}-O_{W2}$  en línea y círculo verde,  $C_{1Glc}-O_{W2}$  en línea discontinua verde y  $O_{1PGA}-H_{W2}$  en línea verde y círculo negro. Las coordenadas de reacción que se utilizaron para simular estas reacciones se dan en la siguiente sección 4.2.3.

El mecanismo de ataque frontal ocurre con una barrera de 38.7 kcal/mol y el producto está 16 kcal/mol sobre el reactivo. En el punto de más alta energía, el  $O_{W2}$  llega a posicionarse a 3.18 Å del carbono anomérico mientras  $O_{3PGA}$  está a 2.50 Å de éste; la distancia  $C_{1Glc}-O_P$

tiene un valor de 2.77 Å, una distancia relativamente corta que llega a alargarse hasta 2.84 Å cuando el ataque de PGA tiene lugar. La conformación de PGA en el Modelo 0, que hemos definido a través del ángulo diedro  $\epsilon$ , tiene un valor  $-0.4^\circ$  y el C3 apuntando hacia arriba (Figura 22). En cuanto al ataque de  $W_2$ , la barrera de energía es un poco más baja 34 kcal/mol y el producto queda en aproximadamente 13 kcal/mol (Glucosa). Cabe destacar que para simular este proceso se definió una coordenada de reacción doble que no incluía la abstracción del protón del agua por parte de PGA ( $O1_{PGA}-H_{W2}$ ). Como se comentará en el apartado 4.2.4, la coordenada de reacción doble no funciona para todos los modelos pero en este caso sí lo ha hecho debido a la orientación relativa que tiene  $W_2$  con respecto a la glucosa.

### 4.2.3 Simulaciones QM/MM del Mecanismo de Ataque Frontal

A continuación se describe de forma genérica cómo se han realizado los cálculos QM/MM para el estudio del mecanismo de ataque frontal. Primeramente, la caja de agua fue reducida a una esfera que incluyó todas las moléculas de agua dentro de un radio de 30 Å alrededor del carbono anomérico del UDP-Glc del monómero B, aunque conservando los residuos de ambos monómeros (12911 átomos) (Figura 24A). A continuación se definió el grupo de átomos activos, esto es, todos los átomos que se moverán durante las optimizaciones o dinámicas realizadas a este nivel de teoría. El grupo de átomos activos incluyó la totalidad de los residuos que tienen algún átomo dentro de una radio de 15 Å del carbono anomérico ( $C1_{Glc}$ ) del UDP-Glc (monómero B), con un total de 2093 átomos activos, mientras que las coordenadas de los átomos fuera de la esfera de 15 Å mantuvieron fijas para aumentar la eficiencia computacional de los cálculos. La región QM quedó conformada por 85 átomos que incluyeron al sustrato PGA, el monosacárido Glc y el grupo difosfato de UDP (hasta el grupo  $CH_2$ ), el cofactor  $Mg^{2+}$  y su primera esfera de coordinación formada por dos moléculas de agua ( $W_{357}$  y  $W_{342}$ ), y los residuos His258 y Asp136; además, para el Modelo 1 y aquellos en los que las aguas están presentes, también se incluyeron en la región QM las 4 moléculas de agua  $W_2$ ,  $W_3$ ,  $W_4$  y  $W_{353}$ , las cuales forman una red de enlaces de hidrógeno alrededor del grupo carboxilato de PGA (Figura 24.B). También fueron necesarios tres átomos de hidrógeno (*link atoms*)<sup>241,242</sup> para saturar la valencia de los átomos QM en los que se definió la partición QM/MM. Para reasignar la carga de los átomos frontera se utilizó el esquema *charge-shift*.<sup>179,243</sup>



**Figura 24.** A: Representación de la esfera seleccionada a partir del sistema solvatado en una caja de aguas y previamente minimizado y equilibrado a través de MD. La capa exterior de la esfera representada con líneas muy finas y suaves permanece fija a lo largo de las optimizaciones y la capa más interna, representada por líneas gruesas, se optimiza libremente, los sustratos se dibujan en barras más gruesas. B: Ilustración de la región QM usada en los cálculos QMMM, los átomos MM han sido dibujados en color gris; las líneas curvas rojas señalan el sitio de corte entre región QM y MM. Las aguas  $W_{342}$ ,  $W_{357}$ , Asp136, His258 y los oxígenos  $O1\beta$  y  $O1\alpha$  de los fosfatos  $\beta$  y  $\alpha$ , respectivamente, conforman la primera esfera de coordinación de  $Mg^{2+}$ . En el caso del sustrato dador PGA y la glucosa transferible, no se realizaron particiones quedando descritas en su totalidad a nivel cuántico; las aguas  $W_2$ ,  $W_3$ ,  $W_4$ , y  $W_{353}$  son incluidas en la región QM también en aquellos modelos en los que están presentes. Por último, se señala con flechas de color lila la dirección en la que se moverán los electrones durante la reacción química y también se señalan las interacciones de las aguas en líneas negras discontinuas.

El mecanismo de reacción fue simulado a través de una coordenada de reacción triple o RC (Reaction Coordinate), definida como  $RC = [d(O_P-C1_{Glc}) - d(O3_{PGA}-C1_{Glc}) - d(HO3_{PGA}-O_P)]$ , (Figura 24.B) que permitió conducir el sistema desde reactivos a productos. La exploración del camino que conecta estas dos especies se llevó a cabo en pasos de  $0.2 \text{ \AA}$  y para la descripción cuántica de los átomos se usó el funcional de densidad BP86<sup>152,154,244-246</sup> junto la base SVP.<sup>247</sup> Los átomos MM quedaron descritos con el campo de fuerzas CHARMM, este nivel de teoría se denota como QM(BP86/SVP)/MM(CHARMM). Después, utilizando estas geometrías, se recalculó la energía para cada una de las estructuras del perfil empleando el nivel QM(M05-2X<sup>166</sup>/TZVP<sup>248</sup>)/MM(CHARMM). El método M05-2X ha sido recomendado por otros grupos para el estudio de carbohidratos y la disociación del enlace UDP-galactosa.<sup>249</sup> La combinación de métodos QM de distinto nivel es una práctica habitual que se utiliza con el objetivo de ahorrar tiempo de cálculo. Además, para el Modelo 1 se probaron optimizaciones utilizando el funcional de densidad PBE.<sup>163</sup>

Los cálculos QMMM fueron hechos con el programa modular ChemShell,<sup>181,250</sup> empleando el programa DL\_POLY<sup>251</sup> para el cálculo de gradientes y energías al nivel MM. Para obtener energías y gradientes al nivel QM se utilizaron TURBOMOLE,<sup>252</sup> GAUSSIAN09<sup>253</sup> o MNDO<sup>254</sup> según se usará el funcional BP86 (o PBE), M05-2X o la aproximación SCC-DFTB,<sup>172,173</sup> respectivamente. En el cálculo de las interacciones no covalentes, ningún cutoff fue usado a nivel MM o QMMM. Para este último, se aplicó el esquema de *acoplamiento*

*electrostático*.<sup>243</sup> Las optimizaciones libres o restringidas por un potencial en el caso de que fuera necesario, fueron llevadas a cabo definiendo un conjunto de átomos activos, los cuales se subdividen en una región QM y MM. La optimización o búsqueda de estados de transición utiliza el esquema microiterativo, en el que además de los grupos de átomos ya definidos, se debe también establecer un núcleo o conjunto de átomos que describan el proceso químico que ocurre.<sup>255,256</sup> Después de que un paso de optimización es completado para el núcleo en el que se está rompiendo o formando un enlace (ciclo macroiterativo), el resto del sistema es relajado (ciclo microiterativo). El algoritmo de minimización empleado es el L-BFGS (Limited memory Broyden-Fletcher-Goldfarb-Shanno);<sup>257,258</sup> y en el caso de la búsqueda de estados de transición, se utilizó un método que combina el algoritmo L-BFGS con el algoritmo P-RFO (Partitioned-Rational Function Optimization).<sup>259,260</sup> Ambos algoritmos están implementados en el módulo HDLCopt<sup>255</sup> de ChemShell.

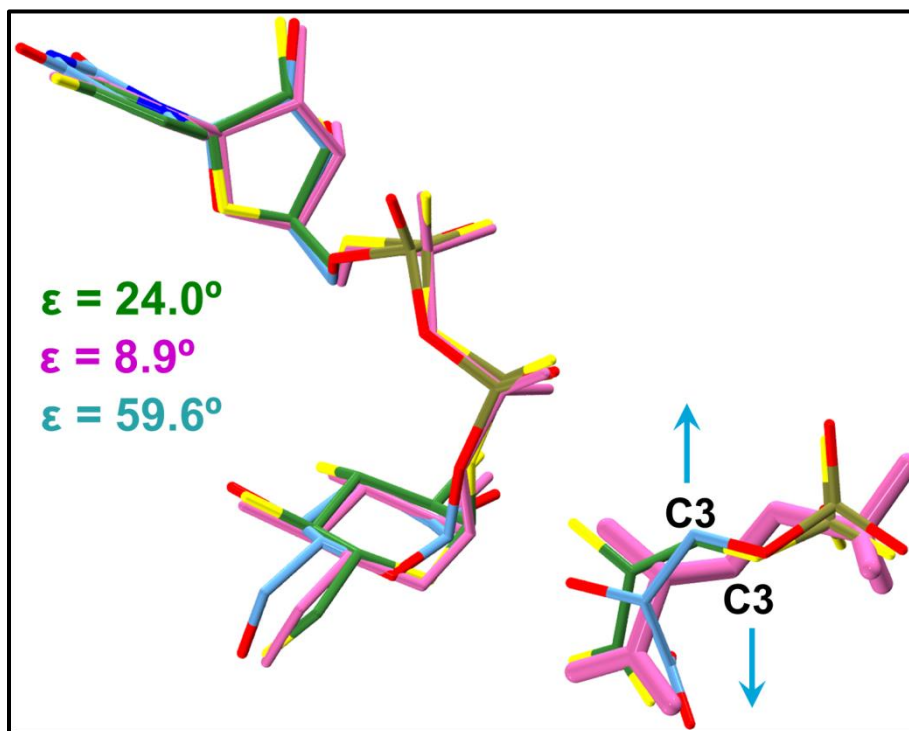
#### 4.2.4 Cálculos QM/MM para la Hidrólisis de UDP-Glc por la $W_2$

Debido a la proximidad de la  $W_2$  al  $C1_{Glc}$ , en aquellos modelos en los que esta molécula de agua está presente, se simuló la hidrólisis para formar glucosa usando como base catalítica el grupo carboxilato de PGA ( $COO^-_{PGA}$ ), específicamente el  $O1$  que forma un enlace de hidrógeno con esta agua. De forma general, el cálculo se realizó igual que en el caso del mecanismo de ataque frontal, pero definiendo otra coordenada de reacción. En el caso del complejo no catalítico (Modelo 0, apartado 4.2.2) y el Híbrido 1.A (apartado 4.3.5), una coordenada de reacción doble fue suficiente para simular el ataque de la molécula de agua  $W_2$ , esto es,  $RC = [d(O_P-C1_{Glc}) - d(O_{W2}-C1_{Glc})]$ , verificándose que uno de los hidrógenos de  $W_2$  se transfiere hasta el  $O1_{PGA}$  sin necesidad de incluirlo en la RC. Contrariamente, en el caso del complejo catalítico competente (Modelo 1, apartado 4.3.5) y el Híbrido 1.B (apartado 4.3.5), la RC doble no dio resultados y se probó una RC triple,  $RC = [d(O_P-C1_{Glc}) - d(O_{W2}-C1_{Glc}) - d(H_{W2}-O1_{PGA})]$ , la cual tampoco funcionó. El análisis de las estructuras de la exploración muestra que el  $H1_{Glc}$  comienza a orientarse directamente hacia el  $O_{W2}$  durante el camino de reacción. En esta nueva configuración, generada cuando el enlace UDP-Glc ya se ha roto, la molécula de agua parece quedar bloqueada por el  $H1_{Glc}$  y el ataque nucleofílico no procede. Todos los detalles de estos estudios se darán en la sección 4.3 cuando se discuta cada modelo.

#### 4.2.5 Refinamiento del Modelo: Obtención del Modelo 1

Con el objetivo de mejorar el modelo de cara a la catálisis, se procedió a un refinamiento que en otras ocasiones ha resultado efectivo. Este consiste en realizar una simulación adicional de MD de 50 ps, esta vez usando la aproximación QM(SCC-DFTB)<sup>172,173</sup>/MM(CHARMM), con el objetivo de conseguir un producto más estabilizado. Para hacer esto, se utilizan los complejos enzima-producto (enzima-[PGA-Glc]) obtenidos en las exploraciones del mecanismo realizadas con el Modelo 0. Después de una minuciosa inspección de la simulación, fueron seleccionadas tres estructuras que se usaron como puntos de partida para conducir las de nuevo hasta reactivos. Para esto, se llevaron a cabo

optimizaciones de geometría restringidas a lo largo de una coordenada de reacción inversa, obteniendo de esta manera un nuevo conjunto de complejos de Michaelis. El nuevo conjunto de complejos así obtenidos (por ejemplo el Modelo 1), comparte ciertas características que lo hacen más parecido al complejo de Michaelis cristalográfico que finalmente se publicó (GpgS·Mn<sup>2+</sup>·UDP-Glc·PGA-1, código PDB 4Y6N), que fue obtenido cuando el primer estudio computacional ya estaba terminado.



**Figura 25.** Superposición de complejos ternarios nativos cristalográficos junto al Modelo 1. El complejo ternario nativo inicial de GpgS (GpgS·Mn<sup>2+</sup>·UDP-Glc-PGA, estructura 0), se dibuja en color rosa; el complejo de Michaelis 4Y6N (GpgS·Mn<sup>2+</sup>·UDP-Glc-PGA-1) se representa a través de carbonos verdes; y al Modelo 1 se le asignan carbonos celestes. Además, se indica la orientación de C3<sub>PGA</sub> y el diedro  $\epsilon$  en cada complejo.

Como puede ser observado en la Figura 25, el proceso de modelización llevó a un cambio en la conformación de PGA en la que el grupo fosfato adopta una orientación bastante similar a la que se encuentra en el complejo de Michaelis 4Y6N, con C3<sub>PGA</sub> apuntando hacia arriba en ambos casos. Curiosamente, la disposición del hidroxilo nucleofílico con respecto a COO<sup>-</sup><sub>PGA</sub> en el Modelo 1 (Modelo catalítico) con  $\epsilon=59.6^\circ$ , no muestra demasiadas similitudes con ninguna de las dos estructuras cristalográficas. Esto es porque O3<sub>PGA</sub> y O1<sub>PGA</sub> definen planos diferentes; mientras que en el complejo de Michaelis 4Y6N el valor de  $\epsilon$  (24.0°) es más pequeño que en el Modelo 1, en el caso de la estructura 0  $\epsilon$  tiene un valor de 8.9°. A pesar de esto y debido a la notable semejanza en la orientación del grupo fosfato y la distancia C1<sub>Glc</sub>-O3<sub>PGA</sub> < 3.0 Å, consideramos que los resultados obtenidos con el Modelo 1 eran suficientemente fiables.

El refinamiento del Modelo 0 recién descrito, nos permitió conseguir complejos en los que la distancia C1<sub>Glc</sub>-O3<sub>PGA</sub> < 3.0 Å. Además se observa que el agua W<sub>2</sub> se ha desplazado ~ 1 Å, quedando a unos 5 Å del carbono anomérico e interactuando de cerca con el COO<sup>-</sup><sub>PGA</sub>. Las barreras de energía potencial a nivel QM(BP86/SVP)/MM(CHARMM) correspondientes

a los cuatro complejos que se seleccionaron son de 20.8, 20.9, 28.6 y 15.9 kcal/mol, dando un promedio exponencial de 16.7 kcal/mol, que ya es más cercano al valor derivado experimentalmente. Cabe mencionar que para llevar a cabo el cálculo del promedio exponencial de barreras de energía, se asume que cada uno de los complejos E·S visitados durante la trayectoria de MD tienen la misma probabilidad de ocurrencia. Además, se asume que cada estructura seleccionada reacciona con una constante de velocidad local que cumple con la Teoría del Estado de Transición Convencional (Conventional Transition State Theory, CTST), y que la barrera de energía libre es semejante a los valores energía potencial.<sup>176,261</sup> Así, la barrera de energía potencial promedio que correspondería a la constante de velocidad global de la reacción es.<sup>262,263</sup>

$$\Delta E_{ave}^{\ddagger} = -RT \ln \left\{ \frac{1}{n} \sum_{i=1}^n \exp \left( \frac{-\Delta E_i^{\ddagger}}{RT} \right) \right\} \quad [4.1]$$

en donde  $\Delta E_{ave}^{\ddagger}$  y  $\Delta E_i^{\ddagger}$  corresponde a la barrera de energía promedio y la barrera de energía calculada para el camino de reacción  $i$ , respectivamente.  $T$  es la temperatura,  $R$  es la constante de los gases ideales y  $n$  es el número de caminos de reacción calculados. Al aplicar la ecuación 4.1, sólo consideramos un número pequeño de estructuras a partir de las cuales se generaron los caminos de reacción, en este caso  $n=4$ , las cuales mostraban una distancia  $C1_{Glc}-O3_{PGA} \leq 3.5 \text{ \AA}$ . Una selección aleatoria de un número mucho más grande de estructuras iniciales para el cálculo de caminos de reacción no podría llevarse a cabo pues sobrepasaría las capacidades computacionales de las se dispone. La ecuación 4.1 también será usada en los capítulos 5 y 6.

La comparación entre las estructuras indica que estas diferencias energéticas provienen de diferencias estructurales sutiles pero, como vemos, con importantes efectos. Se trata por ejemplo, de la orientación tomada por los átomos de hidrógeno de la molécula  $W_{342}$  o por el grupo  $O6_{Glc}$ . El complejo que presentaba las mejores interacciones en relación a la catálisis fue seleccionado para proseguir con el estudio. A este complejo lo llamamos Modelo 1 (barrera de 15.9 kcal/mol). Con él se estudió el mecanismo de ataque frontal y el de ataque de la molécula de agua calculando los perfiles de energía potencial y localizando los puntos estacionarios, se hicieron análisis de cargas, de interacciones electrostáticas y se estimó la importancia de los términos entrópicos y térmicos a la barrera.

#### 4.2.6 Modelo 2: Modelo sin Aguas en el Centro Activo

Como se ha comentado en la introducción de este capítulo, la resolución a la que se obtuvo la estructura cristalográfica de estos complejos ternarios nativos de GpgS (2.35 Å en el caso del complejo de Michaelis), no permite concluir la presencia de las moléculas de agua  $W_2$ ,  $W_3$  ni  $W_{353}$ , de ahí que en las estructuras cristalográficas publicadas (4Y6N, 4Y6U y 4Y9X) no se den coordenadas para dichas moléculas. Consistentemente con esta observación experimental, y partiendo del Modelo 1, se generó un modelo *in silico* que fuera coherente con este hecho, al que llamamos Modelo 2. La comparación de ambos modelos nos

permitirá analizar mejor el papel de estas moléculas de agua en caso de estar presentes, así como ver si esto afecta las conclusiones mecanísticas sobre GpgS.

#### **4.2.7 Simulaciones Adicionales del Mecanismo de Reacción: Híbrido 1.A e Híbrido 1.B**

Con el objetivo de investigar más a fondo el rol de las moléculas de aguas que se encuentran en la vecindad del centro anomérico, así como también la influencia de la conformación del PGA en las barreras energéticas del mecanismo, se han construido modelos híbridos *in silico*. Los modelos *in silico* se construyeron combinando coordenadas del complejo de Michaelis catalítico (Modelo 1), con las del Modelo 0 y la estructura cristalográfica 4Y6N, Híbrido 1.A e Híbrido 1.B, respectivamente. En el Modelo 1 se reemplazaron las coordenadas de PGA,  $W_2$ ,  $W_3$  y  $W_4$ , por las coordenadas de PGA,  $W_2$ ,  $W_3$  y  $W_4$  correspondientes a dos estructuras: el Modelo 0 ó complejo no catalítico (Híbrido 1.A o H 1.A); y el complejo de Michaelis cristalográfico GpgS·Mn<sup>2+</sup>·UDP-Glc·PGA-1, código PDB 4Y6N (Híbrido 1.B o H 1.B). En este último caso, las coordenadas de las aguas corresponden a las que se obtendría si se asignara la densidad electrónica en cuestión. Hecho esto, los cálculos de optimización y coordenada de reacción se llevaron a cabo tal como se describió previamente en el apartado 4.2.3.

#### **4.2.8 Cálculo de Contribuciones Electroestáticas**

Con el objetivo de cuantificar el rol de diferentes residuos enzimáticos, así como el de las moléculas de agua en la vecindad del azúcar, en los modelos en los que están incluidas, se llevan a cabo cálculos puntuales de energía al nivel QM(M05-2X/TZVP)/MM(CHARMM) a lo largo del camino de reacción. Para calcular la contribución de aquellos residuos que forman parte del subsistema clásico (residuos MM), la carga del residuo de interés se hace cero y se calcula la energía del sistema [término 1]. Después se extrae el término energético QM/MM<sub>electr</sub> asociado al residuo con su carga MM [término 2] y se le resta el término [1]. La sustracción de estos dos términos da una estimación de la interacción electrostática de dicho residuo con la zona QM, a la que nos referimos como contribución electrostática del residuo. La contribución electrostática la definimos arbitrariamente de la siguiente manera: un término negativo indica que hay una contribución favorable, es decir, que el residuo mejora la o las interacciones que establece con la región QM a lo largo de la reacción; por el contrario, un término positivo indica que el residuo deteriora sus interacciones con la región QM. En ambos casos se utiliza como referencia el término electrostático residuo-región QM calculado para el reactivo y por consiguiente, las interacciones se ven favorecidas o desfavorecidas con respecto a éste.

Cuando se quiere calcular la contribución electrostática de residuos que forman parte de la región QM, como pasa con COO<sup>-</sup><sub>PGA</sub> y las moléculas de agua  $W_2$ ,  $W_3$ ,  $W_{353}$  y  $W_4$ , el procedimiento incluye un paso adicional. Este paso adicional consiste en calcular el término 2, con el residuo en cuestión puesto en el subsistema MM. Después se le resta el término [1], al igual como se explicó previamente.

#### 4.2.9 Cálculo de Cargas Naturales e Interacciones de Orbitales Naturales

Para llevar a cabo el análisis de cargas y de orbitales de enlace naturales (NPA<sup>264</sup> y NBO,<sup>265,266</sup> respectivamente, por su sigla en inglés), se ha utilizado el programa NBO v3.1<sup>267</sup> incluido en el programa Gaussian09.<sup>253</sup> El nivel de teoría que se emplea para realizar este análisis corresponde al QM(M05-2X/TZVP)/MM(CHARMM), y se utilizan las estructuras de mínimos (R, IP y P), de puntos de silla (TS) o de estructuras que se han optimizado con algoritmos de búsqueda de estados de transición, pero para los cuales no se ha encontrado una frecuencia imaginaria (oxobarbenio, OC). Todas estas estructuras se han obtenido al nivel de teoría QM(BP86/SVP)/MM(CHARMM). Como podrá verificarse más adelante, se seleccionan las cargas de los átomos C1, O5 y H1 de la glucosa, las cuales se suman y relativizan con respecto al reactivo, por lo que la carga de esta especie será cero. En el caso de las interacciones entre orbitales moleculares, el análisis se hizo para todas las especies UDP----OC pero sólo se mostrará el relacionado con el IP. El IP es una estructura representativa de las especies UDP----OC y ha sido inequívocamente caracterizado como mínimo.

#### 4.2.10 Simulación de Umbrella Sampling para el Mecanismo de Ataque Frontal

Se llevaron a cabo simulaciones de MD empleando la técnica de *Umbrella Sampling*<sup>202-204</sup> (US por su sigla en inglés) y el nivel QM(SCC-DFTB)/MM(CHARMM) con el objetivo de obtener una estimación de la importancia de los efectos térmicos y entrópicos en la barrera de energía potencial del mecanismo de ataque frontal para los Modelos 1 y 2. A pesar de que este método de cálculo sobreestima las barreras de energía relacionadas con reacciones de transferencia de azúcares, tendiendo a dar como resultado perfiles más bien concertados, su naturaleza semiempírica permite evaluar el perfil de energía libre en un tiempo razonable.

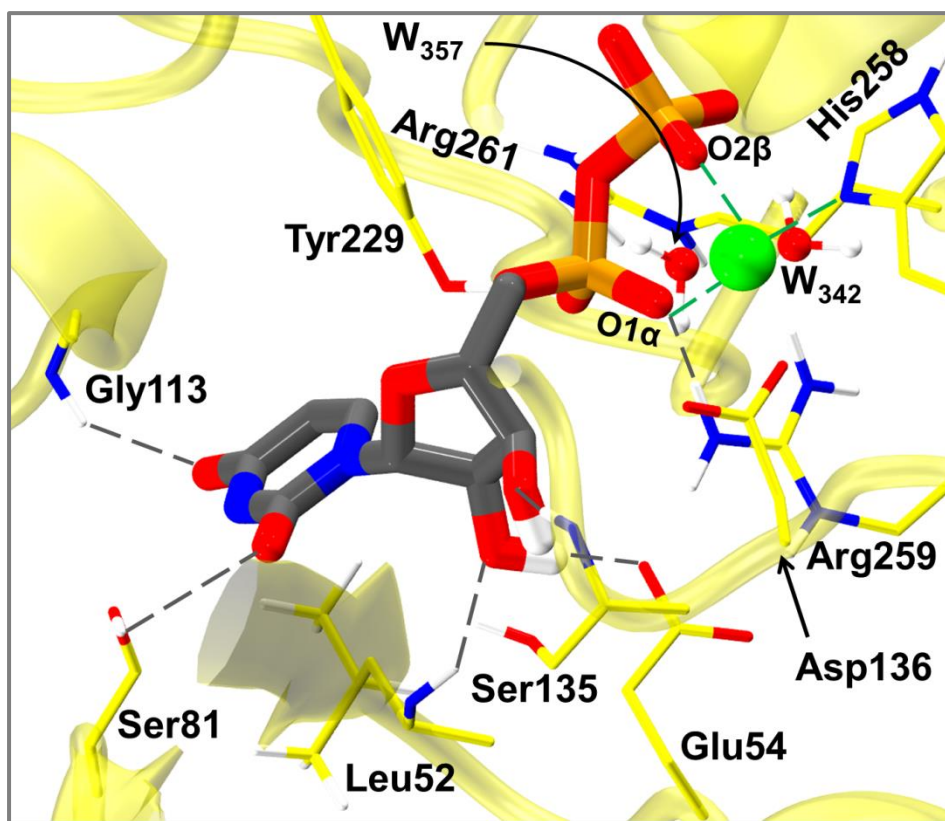
En este caso, el camino de reacción fue explorado en pasos de 0.1 Å y 80 estructuras resultaron del cálculo de este camino; a cada una de éstas se le denomina ventana dentro del marco US. Durante las simulaciones de MD realizadas para cada una de las ventanas, se aplicó una constante de fuerza de  $59 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$  sobre la RC =  $[d(O_P-C1_{Glc}) - d(O3_{PGA^-}-C1_{Glc}) - d(HO3_{PGA^-}-O_P)]$ , de modo de prevenir que la estructura se aleje demasiado de la estructura de referencia dentro de cada ventana, pero también permitiendo solapamiento entre ventanas sucesivas. Para la simulación de MD se especificó un colectivo NVT y se usó la región activa anteriormente definida en el apartado 4.2.3. La temperatura se ajustó aplicando los termostatos de Nosé-Hoover<sup>196,197</sup> y Berendsen,<sup>198</sup> y el algoritmo SHAKE se aplicó a todas las distancias O-H. La temperatura inicial del sistema fue establecida en 10 K y luego incrementada en pasos de 5 K. Al inicio de cada nueva definición de temperatura el



sistema fue sometido a cortas equilibraciones de 0.1 ps hasta alcanzar los 300 K. En esta etapa, el sistema fue equilibrado durante 10 ps, seguido de una MD de producción de 50 ps más a partir de la cual se realizó el muestreo para el cálculo del perfil de energía libre. El análisis de los datos generados se llevó a cabo con el método *Umbrella Integration* (UI)<sup>206,207</sup> y *Weighted Histogram Analysis Method*<sup>205</sup> (WHAM), aunque sólo el primero se consiguió un buen ajuste de los datos. Para las simulaciones de US se utilizó el módulo *Dynamics* implementado en el programa ChemShell.

### 4.3 Resultados y Discusión

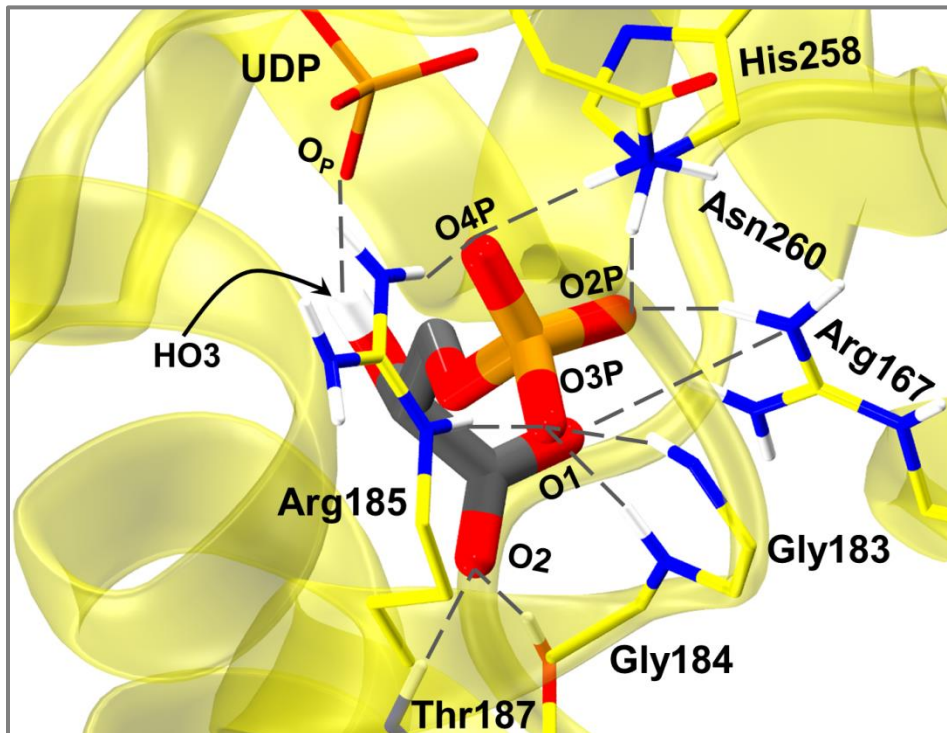
Para iniciar la descripción de este estudio, comentaremos sobre los residuos que interaccionan con el sustrato dador UDP-Glc y el sustrato aceptor PGA. Estos residuos se subdividen en aquellos que interaccionan con UDP y luego con Glucosa. En el caso de las moléculas de agua situadas en la cercanía del carbono anomérico o del grupo carboxilato de PGA, se describirán en la sección correspondiente de cada modelo.



**Figura 26.** Residuos en el sitio de asociación de UDP. Las distancias de interacción se indican a través de líneas discontinuas grises y verdes, en el caso de los residuos que forman la primera esfera de coordinación del  $Mg^{2+}$  (esfera verde). Los residuos que se encuentran en la cavidad son; Gly113 y Ser81 en el caso de Uracilo; Glu54, Ser135 y Leu52 en el caso de 4-deoxiribosa; Tyr229, Arg261, Arg259 en el caso del grupo difosfato; y el cofactor metálico, cuya esfera de coordinación la conforman dos moléculas de agua,  $W_{342}$  y  $W_{357}$ , el Asp136, la His258 y los oxígenos  $O1\alpha$  y  $O2\beta$  del grupo difosfato.

Como puede apreciarse en la Figura 26, UDP interacciona con diferentes residuos: su grupo Uridina está rodeado por Ser81, Leu52, Ser135, Gly113 y Glu54. El grupo difosfato interacciona con: Tyr229 (O2 $\alpha$ ), Arg259 (O1 $\alpha$ ) y Arg261 (O2 $\alpha$ ). Por supuesto, la asociación de este grupo está fuertemente vinculada al catión metálico o cofactor, Mg<sup>2+</sup>. La primera esfera de coordinación del catión incluye a His258, Asp136, O1 $\alpha$ , O2 $\beta$ , W<sub>342</sub> y W<sub>357</sub>.

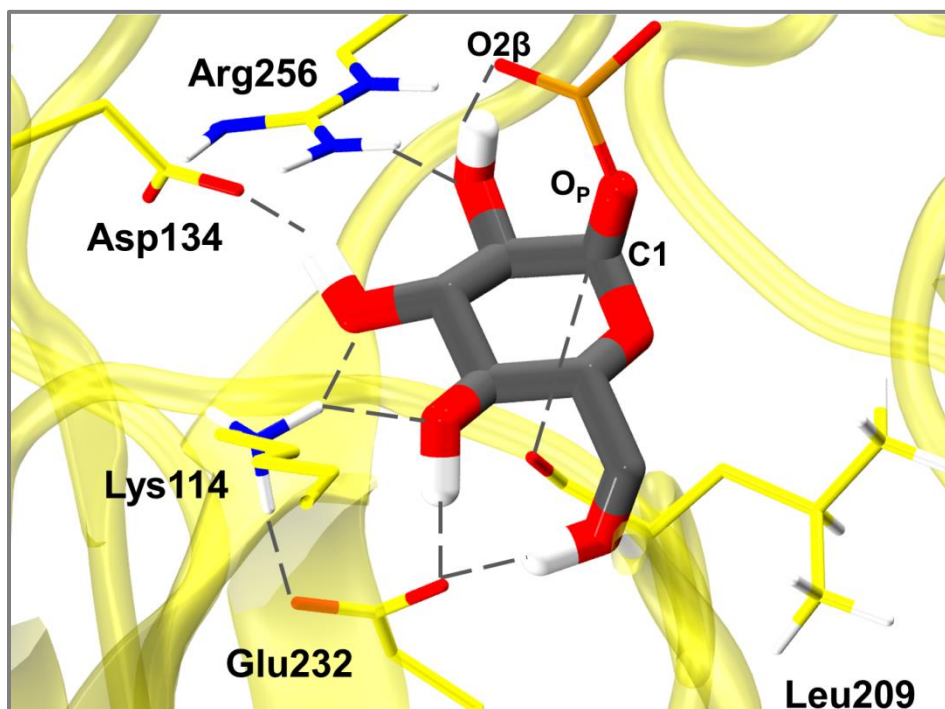
El sustrato aceptor PGA, es una molécula con carga neta negativa igual a -3 y es de esperar que sus grupos cargados, COO<sup>-</sup> y PO<sub>3</sub><sup>2-</sup>, se alejen lo más posible el uno del otro.



**Figura 27.** Residuos en el sitio de unión de PGA, las distancias de interacción se indican a través de líneas discontinuas grises. Los residuos que se encuentran en esta cavidad de unión del sustrato aceptor son; Arg185, Thr187, Gly184, Gly183, Arg167, Asn260 e His258. Además, el PGA orienta su hidroxilo nucleofílico hacia O<sub>P</sub>, una interacción relevante en la catálisis.

Se observa que el grupo fosfato de PGA forma varias interacciones fuertes con Asn260, Arg167, Gly183, Gly184, Arg185 e His258, el último residuo también forma parte de la esfera de coordinación del Mg<sup>2+</sup>. Así, 6 residuos, dos de ellos con carga neta positiva, están interaccionando con el fosfato de PGA. En el caso del grupo carboxilato, éste forma dos interacciones con un único residuo, la Thr187. Por último, el PGA establece una importante interacción el grupo fosfato  $\beta$  (HO3<sub>PGA</sub>-O<sub>P</sub>), que como se ha comentado tiene una gran relevancia en la catálisis debido a que O<sub>P</sub> abstraerá ese protón para activar el PGA para el ataque nucleofílico, formándose de esta manera el nuevo enlace glicosídico.

Finalmente, analizamos los residuos que interaccionan con la Glucosa. La Glucosa se une a la enzima formando parte del sustrato dador UDP-Glc, así que en la Figura 28 se incluye el grupo fosfato  $\beta$  de UDP.



**Figura 28.** Residuos en el sitio de unión de Glc, las distancias de interacción se indican nuevamente a través de líneas discontinuas grises. Los residuos que se encuentran en este sitio de unión son; Arg256, Asp134, Lys114, Glu232 y Leu209. Adicionalmente, se incluye el fosfato  $\beta$  de UDP para señalar la interacción intra-sustrato dador  $O2\beta$ - $HO2_{Glc}$ .

La Arg256 interacciona con Glucosa a través de su  $O2$  mientras que Asp134, que es parte del motivo Asp134-Ser135-Asp136, interacciona de cerca con  $HO3$ . Lys114 y Glu232 interaccionan fuertemente entre sí y corresponden a la pareja de residuos conservados que se encuentran en las GTs-ret (GT-A) y pueden también aparecer como Arg o Asp (ver apartado 1.3.3.2). Cada uno de ellos establece dos interacciones con la Glucosa: Lys114 con  $O3/O4$  y Glu232 con  $HO4/HO6$ . Además, se incluye a la Leu209 que se ubica en la cara  $\beta$  de la Glucosa, a  $\sim 4.2$  Å del carbono anomérico.

De acuerdo a las discusiones que se harán más adelante, es conveniente tener en mente los residuos que definen la cavidad del azúcar transferible, en este caso Glucosa; y los residuos de la cara  $\beta$  del azúcar, que en esta enzima corresponde a una Leu209. De hecho, de cara al estudio mecanístico, sólo el carbonilo de la Leu209 cuyo oxígeno está a una distancia que oscila entre  $4.1 - 4.3$  Å del carbono anomérico  $C1_{Glc}$ , está situado en la cara  $\beta$  del azúcar en los Modelos que presentaremos a lo largo de este capítulo. En principio, es una posición que parece indicada para el ataque nucleofílico, sin embargo este grupo no es buen nucleófilo tal como está y sería necesaria una activación. Debido a esto, el mecanismo de doble desplazamiento involucrando a un residuo de la enzima no fue simulado para esta enzima.

A continuación, se describirán en detalle los resultados y conclusiones derivados de los diferentes experimentos computacionales que se realizaron en cada uno de los modelos antes mencionados (excepto para el Modelo 0, cuyos resultados ya se han comentado en el apartado 4.2.2). Para ser fieles a la cronología de nuestro trabajo, se explicarán los resultados conseguidos a través del Modelo 1 (con aguas). Después se discutirán los

resultados obtenidos en base al Modelo 2 (sin aguas). Los resultados publicados corresponden al Modelo 2 pues en éste no se ha considerado el efecto de las posibles moléculas de aguas y por consiguiente, representa un modelo más acomodado a la información experimental que surgió después de haber finalizado el estudio del Modelo 1.

### 4.3.1 Modelo 1 o Complejo Catalítico: Mecanismo de Ataque Frontal

Como se indicó más arriba, se seleccionó uno de los complejos para los que se había calculado el perfil de energía potencial, con el objetivo de llevar a cabo un análisis más completo de la reacción. A continuación se presentan y discuten los resultados para este modelo. Para guiar el sistema desde reactivos a productos se empleó una coordenada de reacción triple  $RC = [d(O_P-C1_{Glc}) - d(O3_{PGA}-C1_{Glc}) - d(HO3_{PGA}-O_P)]$ . Inicialmente, se usaron dos niveles de teorías para estudiar el mecanismo de ataque frontal, el método QM(BP86/SVP)/MM(CHARMM) y el QM(PBE/SVP)/MM(CHARMM). En ambos casos se recalcularon las energías de las estructuras del camino de reacción al nivel QM(M05-2X/TZVP)/MM(CHARMM). Los caminos de reacción obtenidos resultaron ser muy semejantes tanto a nivel estructural como energético, y los valores de la coordenada de reacción a lo largo del camino para los puntos de más alta energía ( $\Delta E^{\ddagger}_1$  y  $\Delta E^{\ddagger}_2$ ), y el intermedio ( $\Delta E^{IP}$ ) son análogos, diferenciándose en apenas centésimas de Å. Como muestra de estas similitudes, en la Tabla 7 se indican las energías más altas a lo largo del camino de reacción (posibles TS1 y TS2), así como el posible IP. En este último caso, también se dan las distancias del punto asociado al IP.

**Tabla 7.** Barreras de Energía Potencial ( $\Delta E^{\ddagger}$ ) y energías de reacción ( $\Delta E$ ) en kcal/mol, junto a algunas distancias relevantes para el mecanismo de ataque frontal. Los valores energéticos y distancias fueron obtenidos a dos niveles de teoría diferentes, QM(BP86/SVP)/MM(CHARMM) y QM(PBE/SVP)/MM(CHARMM). En cada caso se realizaron cálculos de energía puntuales al nivel QM(M05-2X/TZVP)/MM(CHARMM). Los reactivos y sus distancias corresponden a estructuras optimizadas sin restricciones. Por el contrario, las barreras de energía y las distancias del IP corresponden a puntos del camino de reacción con valores de coordenada de reacción análogos (-1.29 y -1.31 Å a nivel QM(BP86/SVP) y QM(PBE/SVP), respectivamente).

| Tratamiento QM         | $\Delta E^{\ddagger}_1$ | $\Delta E^{IP}$ | $\Delta E^{\ddagger}_2$ | $\Delta E$ | Reactivo (IP) [Å]                 |                                      |                                    |                                    |
|------------------------|-------------------------|-----------------|-------------------------|------------|-----------------------------------|--------------------------------------|------------------------------------|------------------------------------|
|                        |                         |                 |                         |            | C1 <sub>Glc</sub> -O <sub>P</sub> | C1 <sub>Glc</sub> -O3 <sub>PGA</sub> | C1 <sub>Glc</sub> -O <sub>W2</sub> | O <sub>P</sub> -HO3 <sub>PGA</sub> |
| BP86/SVP               | 15.9                    | 12.4            | 11.8                    | -2.7       | 1.50(3.00)                        | 2.91(2.64)                           | 5.10(3.21)                         | 2.49(1.66)                         |
| M05-2X/TZVP// BP86/SVP | 18.4                    | 14.9            | 18.9                    | 0.0        |                                   |                                      |                                    |                                    |
| PBE/SVP                | 15.3                    | 14.6            | 12.4                    | -1.3       | 1.50(2.99)                        | 2.91(2.64)                           | 5.02(3.24)                         | 2.49(1.66)                         |
| M05-2X/TZVP// PBE/SVP  | 18.6                    | 15.2            | 19.3                    | -1.3       |                                   |                                      |                                    |                                    |

Como puede observarse en la Tabla 7, tanto a nivel energético y estructural los resultados BP86 y PBE con la base SVP son muy parecidos. Esto se mantiene cuando las energías son recalculadas a nivel M05-2X, que es el recomendado para la correcta representación del rompimiento del enlace glicosídico y que, como se observa, incrementa el valor de las barreras energéticas (especialmente la asociada con al ataque nucleofílico); también incrementa la energía del intermedio con respecto a los reactivos. Una vez observados estos comportamientos, decidimos quedarnos con el doble nivel que el grupo había estado utilizando en otros sistemas.

Partiendo del camino optimizado a nivel QM(BP86/SVP), diferentes puntos del camino de reacción fueron usados como aproximaciones iniciales de las estructuras de estados de transición o mínimos. Estos puntos fueron optimizados sin ninguna restricción y caracterizados con cálculos de frecuencias y cargas, indicándose también las conformaciones que adquiere el azúcar representadas por los parámetros Cremer-Pople (Tabla 8). La siguiente descripción involucra a los puntos estacionarios obtenidos: los reactivos, un intermedio de par iónico UDP-oxocarbenio, los productos y los correspondientes puntos silla (TS1 y TS2) que los conectan (Tabla 8).

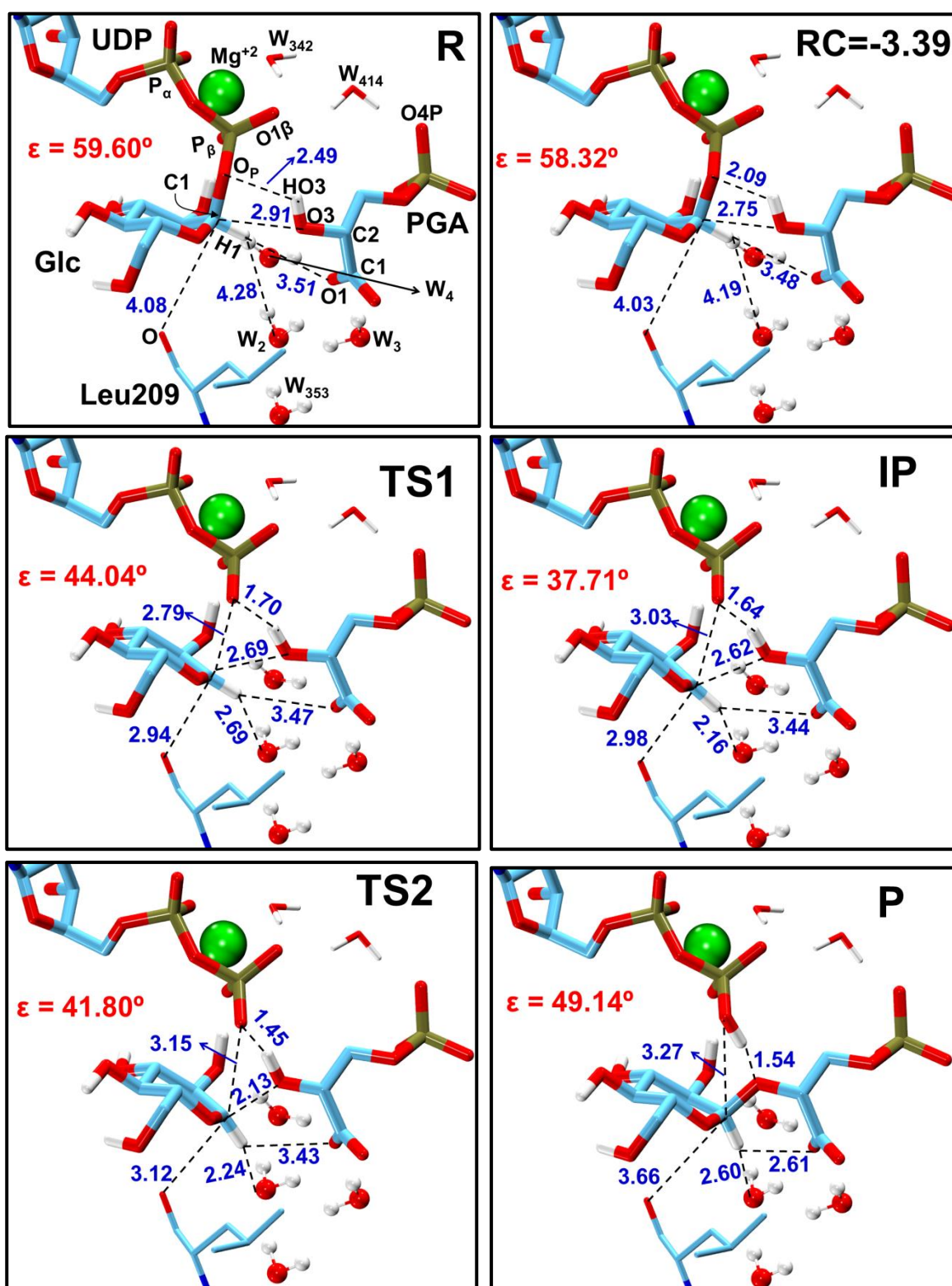
En el Reactivo de este Modelo 1, el  $O_{3_{PGA}}$  está a 2.91 Å del  $C_{1_{Glc}}$  mientras que el  $HO_{3_{PGA}}$  se orienta hacia  $O_P$ , ubicándose a una distancia de 2.49 Å. Además, vale la pena mencionar la distancia entre el grupo carboxilato de PGA y el hidrógeno enlazado al carbono anomérico de la glucosa ( $O_{1_{PGA}}-H_{1_{Glc}}$ ), que tiene un valor de 3.51 Å. Esta distancia es importante debido a que el grupo carboxilato podría facilitar la ruptura UDP-Glc a través de la estabilización de las especies oxocarbenio (OC) que se generan durante este proceso. Curiosamente, la molécula de agua cristalográfica  $W_2$  tiene su oxígeno a 4.28 Å del  $H_{1_{Glc}}$  y a la vez interacciona con  $O_{1_{PGA}}$  a 1.91 Å (ver Tabla 8). La conformación adoptada por el hidroxilo nucleofílico con respecto al carboxilato de PGA es un parámetro interesante de seguir a lo largo de la reacción de transferencia y para hacer esto, se define el ángulo de torsión formado por  $(O_1-C_1-C_2-O_3)_{PGA}$ , el cual llamamos  $\epsilon$  y tiene un valor de 59.60° en el Reactivo.

Es interesante comentar cómo es en este Modelo 1 la red de enlaces de hidrógeno formada por las aguas  $W_2$ ,  $W_3$ ,  $W_4$  y  $W_{353}$  (Figura 29). Las aguas  $W_2$ ,  $W_3$  y  $W_4$  forman enlaces de hidrógeno con el  $O_{1_{PGA}}$ ;  $W_4$  también establece interacción con  $O_{2_{Glc}}$ ;  $W_3$  interacciona con  $W_2$  y con la Arg167 y  $W_2$  forma enlaces de hidrógeno con  $W_4$  y  $W_{353}$ . Esta red de interacciones es el resultado de la combinación de simulaciones clásicas (MD) y QMMM (caminos de reacción y MD). En el Modelo 1, la molécula de agua más cercana al  $C_{1_{Glc}}$  se posiciona a 5.10 Å y a su vez, a 1.91 Å del  $O_{1_{PGA}}$  ( $W_2$ ). Continuando con la descripción mecanística y las interacciones que podrían facilitar la disociación de UDP-Glc, es importante indicar que en el punto del perfil RC= -3.39 Å, la distancia  $O_P-HO_{3_{PGA}}$  se ha reducido a 2.09 Å, 0.40 Å más pequeña que el reactivo; y el oxígeno nucleofílico  $O_{3_{PGA}}$  está a 2.75 Å del  $C_{1_{Glc}}$ . La presencia de la interacción  $O_P-HO_{3_{PGA}}$  en el reactivo o su formación al inicio de la reacción se ha encontrado en todas las GTs-ret estudiadas hasta la fecha y forma parte de la estrategia de catálisis asistida por sustratos para facilitar la salida del grupo saliente. Interacciones adicionales que podrían participar de la catálisis asistida por sustratos son por ejemplo las interacciones:  $C_{1_{Glc}}-O_{1_{PGA}}$  y  $HO_{2_{Glc}}-O_{2\beta}$ .

**Tabla 8.** Características estructurales y electrónicas de los puntos estacionarios. Las energías relativas ( $\Delta E(QM/MM)$ ) se dan en kcal/mol y corresponden al nivel QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM). Las unidades de medida para las distancias, frecuencias, cargas y los parámetros de Cremer-Pople ( $\varphi$  y  $\theta$ ), corresponden a Å,  $\text{cm}^{-1}$ , unidades atómicas (u.a) y grados ( $^\circ$ ), respectivamente. Adicionalmente, se señalan en azul las interacciones entre PGA y  $W_2$  con Glc.

| Interacción                  | Reactivo  | TS1         | IP             | TS2             | Producto          |                   |
|------------------------------|---|-------------|----------------|-----------------|-------------------|-------------------|
|                              | $C1_{\text{Glc}}-O_{\text{P}}$                                  | 1.50        | 2.79           | 3.03            | 3.15              | 3.27              |
|                              | $C1_{\text{Glc}}-O3_{\text{PGA}}$                               | 2.91        | 2.69           | 2.62            | 2.13              | 1.46              |
|                              | $C1_{\text{Glc}}-O5_{\text{Glc}}$                               | 1.38        | 1.27           | 1.27            | 1.29              | 1.40              |
|                              | $O_{\text{P}}-HO3_{\text{PGA}}$                                 | 2.49        | 1.70           | 1.64            | 1.45              | 1.03              |
|                              | $O3_{\text{PGA}}-HO3_{\text{PGA}}$                              | 0.98        | 1.01           | 1.02            | 1.06              | 1.54              |
| Glc --- PGA                  | $C1_{\text{Glc}}-O1_{\text{PGA}}$                               | 4.60        | 4.51           | 4.39            | 3.91              | 3.33              |
| Glc --- $W_2$                | $C1_{\text{Glc}}-O_{W2}$  | 5.10        | 3.66           | 3.20            | 3.29              | 3.63              |
| Glc --- PGA                  | $H1_{\text{Glc}}-O1_{\text{PGA}}$                               | 3.51        | 3.47           | 3.44            | 3.43              | 2.61              |
| Glc --- $W_2$                | $H1_{\text{Glc}}-O_{W2}$  | 4.28        | 2.69           | 2.16            | 2.24              | 2.60              |
| $H_{\text{enlaces}} W_2$     | $H2_{W3}-O_{W2}$  | 1.95        | 2.65           | 3.12            | 2.78              | 2.30              |
| $H_{\text{enlaces}} W_2$     | $H2_{W353}-O_{W2}$  | 1.82        | 1.82           | 1.85            | 1.84              | 1.85              |
| $H_{\text{enlaces}} W_2$     | $O_{W4}-H2_{W2}$  | 1.80        | 1.79           | 1.78            | 1.79              | 1.85              |
| Ión-dipolo $W_2$             | $O1_{\text{PGA}}-H1_{W2}$                                       | 1.91        | 1.83           | 1.80            | 1.79              | 1.80              |
| Dipolo-ión $O1_{\text{PGA}}$ | $H1_{W3}-O1_{\text{PGA}}$                                       | 1.93        | 1.75           | 1.77            | 1.81              | 1.88              |
| Dipolo-ión $O1_{\text{PGA}}$ | $H2_{W4}-O1_{\text{PGA}}$                                       | 1.93        | 1.89           | 1.83            | 1.79              | 1.78              |
|                              | $H1_{W4}-O2_{\text{Glc}}$                                       | 1.83        | 1.87           | 1.83            | 1.75              | 1.71              |
|                              | $HO2_{\text{Glc}}-O2_{\beta_{\text{UDP}}}$                      | 1.51        | 1.40           | 1.42            | 1.47              | 1.64              |
|                              | Torsión $\epsilon$  | 59.60       | 44.04          | 37.71           | 41.80             | 49.14             |
|                              | $C1_{\text{Glc}}-O_{\text{Leu209}}$                             | 4.08        | 2.94           | 2.98            | 3.12              | 3.66              |
|                              | $O1_{\beta}-H2_{W342}$  | 1.91        | 1.64           | 1.63            | 1.61              | 1.64              |
|                              | $O1_{\beta}-H1_{W414}$  | 2.04        | 1.96           | 1.92            | 1.85              | 1.81              |
|                              | $O4_{\text{PGA}}-H2_{W414}$                                     | 1.71        | 1.71           | 1.71            | 1.69              | 1.67              |
|                              | $q(C1_{\text{Glc}})$  | 0.35        | 0.57           | 0.56            | 0.52              | 0.34              |
|                              | $q(H1_{\text{Glc}})$  | 0.24        | 0.28           | 0.28            | 0.25              | 0.20              |
|                              | $q(O5_{\text{Glc}})$  | -0.48       | -0.36          | -0.36           | -0.40             | -0.51             |
|                              | $\Delta q(C1_{\text{Glc}} + H1_{\text{Glc}} + O5_{\text{Glc}})$ | 0.00        | 0.38           | 0.37            | 0.26              | -0.08             |
|                              | $\varphi$   | 146.38      | 246.81         | 251.39          | 257.98            | 258.48            |
|                              | $\theta$  | 5.91        | 44.15          | 48.78           | 44.67             | 24.72             |
|                              | Conformación  | ${}^4C_1$   | ${}^4E$        | ${}^4E-{}^4H_5$ | $\approx {}^4H_5$ | $\approx {}^4C_1$ |
|                              | $\Delta E$  | <b>0.00</b> | <b>14.99</b>   | <b>14.81</b>    | <b>18.83</b>      | <b>0.36</b>       |
|                              | Frecuencias   |             | 42.82 <i>i</i> | Reales          | 42.35 <i>i</i>    |                   |

Al alcanzar el TS1 la distancia  $C1_{\text{Glc}}-O_{\text{P}}$  tiene un valor de 2.79 Å; el  $HO3_{\text{PGA}}$  y  $O3_{\text{PGA}}$  están a 1.70 Å del  $O_{\text{P}}$  y 2.69 Å del  $C1_{\text{Glc}}$ , respectivamente. Estas reorganizaciones ocurren al mismo tiempo que la distorsión de Glc, que alcanza una conformación  ${}^4E$  en la que  $O5-C1-C2$  son coplanares. La disminución de la distancia  $C1_{\text{Glc}}-O5_{\text{Glc}}$  desde 1.38 Å hasta 1.27 Å da cuenta del fuerte carácter de doble enlace ( $C1_{\text{Glc}} sp_3 \rightarrow sp_2$ ) que estabiliza parcialmente el desarrollo de carga positiva (0.38 u.a.) en el centro anomérico. Las distancias  $C1_{\text{Glc}}-O1_{\text{PGA}}$  y  $H1_{\text{Glc}}-O_{W2}$  se han reducido a 4.60 y 2.69 Å, respectivamente. El coste energético para llegar a TS1 es de 14.99 kcal/mol al nivel QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM). Luego, el sistema evoluciona hasta un IP (UDP----OC) en el que la energía se ha reducido en 0.18 kcal/mol y las distancias  $C1_{\text{Glc}}-O_{\text{P}}$ ,  $C1_{\text{Glc}}-O3_{\text{PGA}}$  y  $O_{\text{P}}-HO3_{\text{PGA}}$  son 3.03, 2.62 y 1.64 Å, respectivamente.



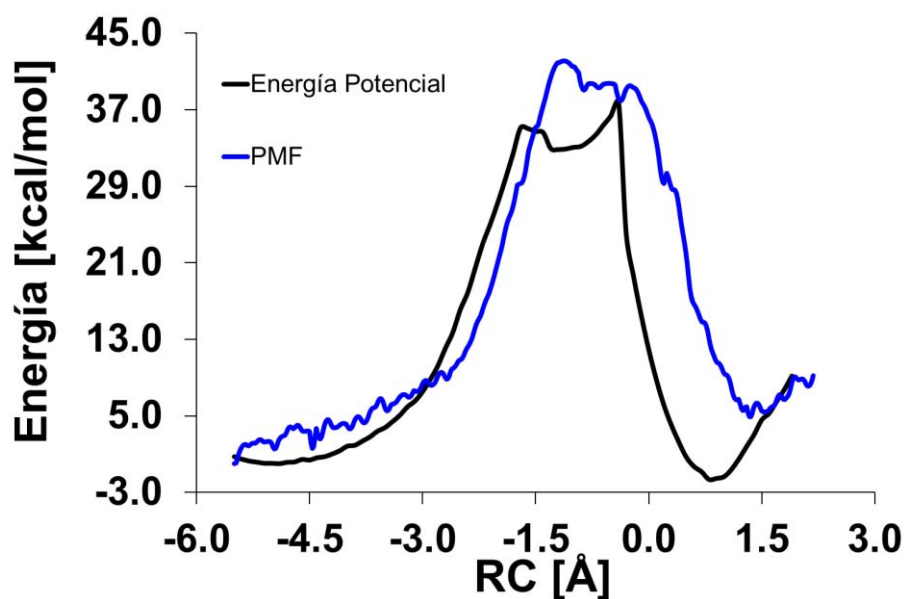
**Figura 29.** Representación de las reorganizaciones atómicas claves a lo largo del camino de reacción. Cada estructura corresponde a un punto optimizado en el orden: Reactivo (R), otra configuración del R que corresponde al punto -3.39 Å de la RC, el estado de transición 1 (TS1), el intermedio IP, el estado de transición 2 (TS2) y el producto (P). Las distancias se dan en Å.



El IP corresponde a una estructura altamente disociada en el que la carga en el centro anomérico es de 0.37 u.a. y las distancias  $C1_{Glc}-O3_{PGA}$  y  $O_P-HO3_{PGA}$  han disminuido discretamente hasta 2.62 y 1.64 Å, respectivamente. Curiosamente, la distancia  $H1_{Glc}-O_{W2}$  y el diedro  $\epsilon$  han disminuido hasta 2.16 Å y 37.71°, respectivamente. La reorientación de  $O_{W2}$  hacia  $H1_{Glc}$  ocurre conservando su interacción con el carboxilato de PGA ( $O1_{PGA}-H_{W2}= 1.80$  Å en el IP y 1.91 Å en el R). Para que la reorientación de  $W_2$  sea posible, ésta debe romper un enlace de hidrógeno con el agua  $W_3$  (Tabla 8). A partir de aquí, el aceptor se aproxima a Glc formando un TS2 con distancias  $C1_{Glc}-O3_{PGA}$  y  $O_P-HO3_{PGA}$  de 2.13Å y 1.45Å, respectivamente, mientras que  $C1_{Glc}-O_P$  tiene un valor de 3.15 Å. La formación del producto ocurre con una barrera global de 18.83 kcal/mol. La nomenclatura IUPAC para el mecanismo sería  $D_{N_{SS}}^*A_N^{\ddagger}D_hA_{xh}$ , lo que quiere decir que se trata de un mecanismo en dos etapas, aunque la estabilización del IP con respecto al TS1 es muy pequeña. En el IP, el grado de disociación del UDP-Glc es alto y muy sensible a pequeñas variaciones de los grupos que rodean al centro anomérico en el sitio activo; además, el IP es la estructura que establece el escenario para que PGA consiga aproximarse sin que eso suponga un gran impedimento estérico para la catálisis. La compleja naturaleza de la especie UDP----OC se refleja en una hipersuperficie muy plana que dificulta extremadamente su caracterización como mínimo (IP) o como punto de silla (TS). A pesar de esto, se trata de un mecanismo de ataque frontal con la formación de un IP de alta energía y probablemente de muy corta vida, que consideramos la especie precursora del nucleófilo activado.

Los cálculos señalan que la barrera de energía limitante de la reacción es 18.8 kcal/mol, la cual se acerca razonablemente a la barrera de activación fenomenológica de 17.5 kcal/mol calculada a partir de una  $k_{cat}$  de 800 min<sup>-1</sup> a 37°C.

Con la intención de estimar el efecto de las contribuciones entrópicas y térmicas en el mecanismo de reacción, se llevaron a cabo simulaciones de US al nivel QM(SCC-DFTB)/MM(CHARMM). A pesar de las limitaciones que ya se han comentado para este método, su naturaleza semiempírica permite evaluar el perfil de energía libre en un tiempo razonable.



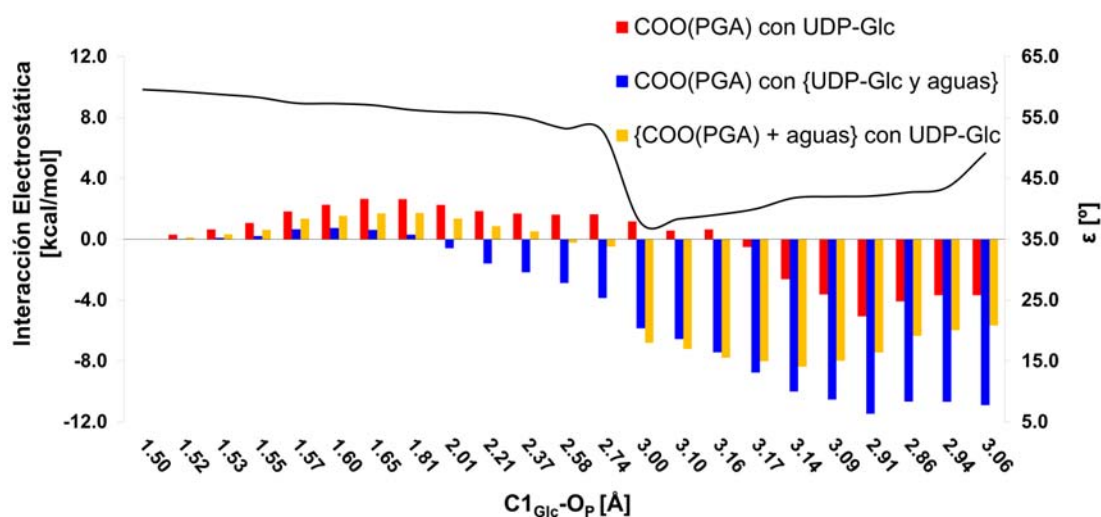
**Figura 30.** Perfil de energía potencial (línea negra) y el Potencial de Fuerza Media (PMF por su sigla en inglés, línea azul) para el mecanismo de ataque frontal en GpgS en base al Modelo 1.

Se observa que los valores de energía libre y potencial comienzan a diferenciarse pronto en la reacción. Al analizar las estructuras se observa que el problema central de esta simulación, es que al optimizar el Reactivo a nivel QM(SCC-DFTB)/MM(CHARMM) se pierde la interacción  $\text{HO3}_{\text{PGA}}\text{-O}_{\text{P}}$ . Al aplicar la coordenada de reacción triple, el  $\text{O3}_{\text{PGA}}$  se comienza a acercar hasta  $\text{C1}_{\text{Glc}}$  demasiado pronto, es decir, cuando el enlace  $\text{C1}_{\text{Glc}}\text{-O}_{\text{P}}$  aún no ha comenzado a disociarse. La disociación estaría más favorecida, en el sentido de que podría ocurrir más pronto, si la interacción  $\text{HO3}_{\text{PGA}}\text{-O}_{\text{P}}$  estuviera presente desde el principio. A consecuencia de esto, PGA y UDP se alejan repentinamente del carbono anomérico a un valor de RC  $-1.3 \text{ \AA}$ , justamente en la región en la que aparece una discreta bajada de energía (Figura 30, línea negra). Después, la energía continúa subiendo hasta que PGA se acerca bruscamente al  $\text{C1}_{\text{Glc}}$  (RC  $-0.3 \text{ \AA}$ ). El valor más alto de energía potencial es de 38 kcal/mol y corresponde a la estructura que antecede el acercamiento repentino entre PGA y  $\text{C1}_{\text{Glc}}$ . En este punto de más alta energía, las distancias  $\text{C1}_{\text{Glc}}\text{-O}_{\text{P}}$ ,  $\text{C1}_{\text{Glc}}\text{-O3}_{\text{PGA}}$  y  $\text{HO3}_{\text{PGA}}\text{-O}_{\text{P}}$  toman valores de  $3.74 \text{ \AA}$ ,  $2.51 \text{ \AA}$  y  $1.63 \text{ \AA}$ , respectivamente.

El perfil de energía libre muestra zonas en las que la energía diverge bastante de los valores de energía potencial, apareciendo diferencias considerables a partir del inicio de la disociación UDP-Glc y en toda la zona de coexistencia de las especies UDP---OC. La barrera de energía libre calculada tiene un valor de 42 kcal/mol, 4 kcal/mol más alta que la obtenida a nivel de energía potencial. Por todo ello, los cálculos SCC-DFTB aunque no sean muy fiables muestran que las contribuciones térmicas y entrópicas aumentan la barrera un máximo de un 10%. Así extrapolando al nivel M05-2X/TZVP, esto implicaría un razonable aumento de barrera de menos de 2 kcal/mol.



aproximado un poco más a ésta. Adicionalmente, el cambio de  $\epsilon$  [(O1-C1-C2-O3)<sub>PGA</sub>] parece estar relacionado no sólo con la disociación UDP-Glc sino también con la apropiada orientación de PGA con respecto a UDP a través de la interacción O<sub>P</sub>-HO3<sub>PGA</sub>. Como se ha destacado anteriormente, la interacción O<sub>P</sub>-HO3<sub>PGA</sub> es fundamental en esta reacción pues facilita el alejamiento del grupo saliente y, además, promueve la activación del nucleófilo para el ataque.

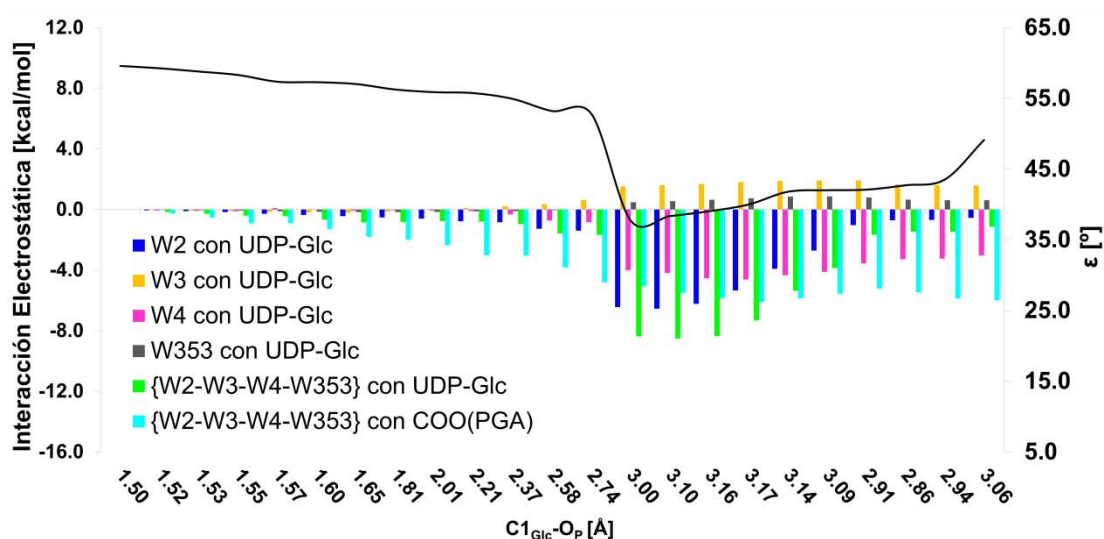


**Figura 32.** Contribuciones electrostáticas de COO<sup>-</sup><sub>PGA</sub> con el sustrato dador UDP-Glc a lo largo de la reacción de transferencia de glucosa. En un caso se calcula el término de interacción de COO<sup>-</sup><sub>PGA</sub> sólo con UDP-Glc (barras rojas), y luego se añade el término de interacción con las moléculas de agua (barras azules). En el último término (barras amarillas), se evalúa el efecto de COO<sup>-</sup><sub>PGA</sub> y las aguas como un conjunto. En el eje vertical secundario (derecha), se señalan los valores de  $\epsilon$  ((O1-C1-C2-O3)<sub>PGA</sub>), y la línea negra representa su variación a lo largo de la reacción. Esto nos permitirá observar su relación con el grado de disociación del sustrato dador UDP-Glc. Los valores negativos indican que el residuo o fragmento mejora sus interacciones con la región QM con respecto al reactivo, o dicho de otra manera, estabiliza durante la reacción. Los valores positivos indican lo contrario.

Cuando el término de interacción con las aguas es considerado, ya sea en la interacción de COO<sup>-</sup><sub>PGA</sub> con el conjunto {UDP-Glc y aguas} (azul), o en la interacción del conjunto {COO<sup>-</sup><sub>PGA</sub> y aguas} con UDP-Glc (amarillo), se observa que en ambos casos la interacción es más estabilizadora, lo cual sugiere que esta especie de esfera de solvatación que forman las aguas alrededor de COO<sup>-</sup><sub>PGA</sub> podrían facilitar la formación de la especie UDP----OC. Curiosamente, la estabilización sigue siendo importante aun cuando la transferencia de glucosa ya ha sido completada y muestra valores especialmente altos cuando el diedro  $\epsilon$  ya ha alcanzado un valor de 40.0°.

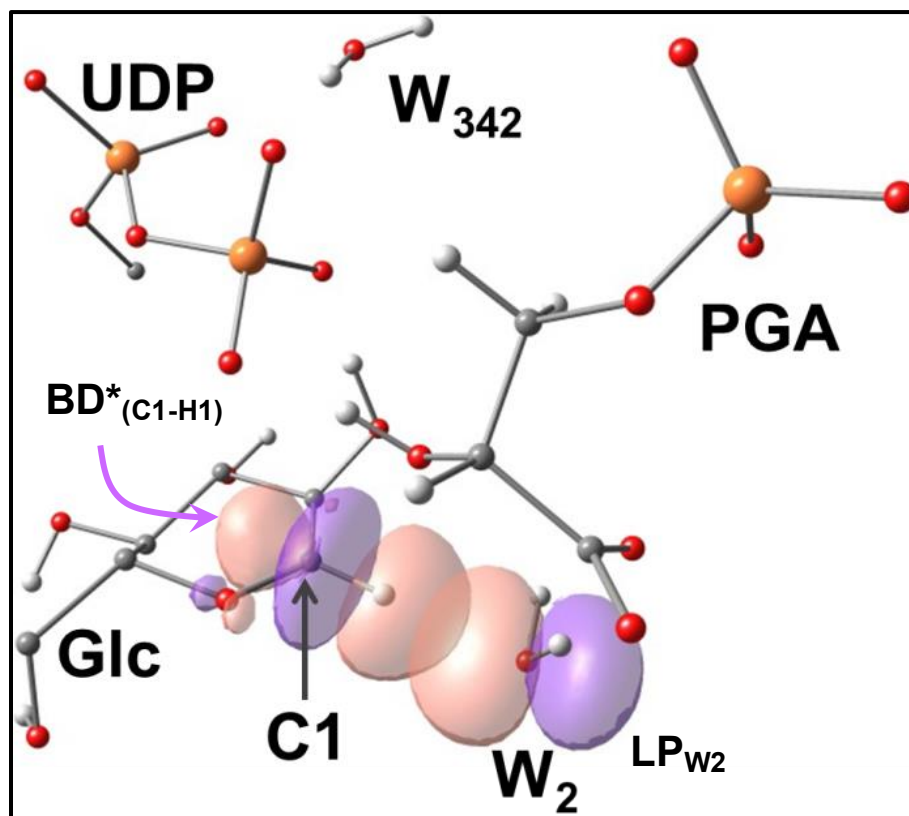
Este primer análisis parece indicar que las moléculas de agua podrían brindar un buen ambiente electrostático que mejoraría las interacciones entre los dos sustratos. La pregunta que surge en este punto es ¿se necesita que algo mejore la habilidad intrínseca de un grupo carboxilato para interactuar favorablemente con una carga neta positiva en la especie UDP----OC? Pues en este caso pareciera ser que sí. Si bien la aproximación de COO<sup>-</sup><sub>PGA</sub> sería favorable para la especie UDP----OC, lo prioritario es el acercamiento del hidroxilo nucleofílico y ambos forman parte de la misma molécula. Bajo este escenario, la proximidad de COO<sup>-</sup><sub>PGA</sub> podría requerir de una regulación adicional en orden a no desfavorecer el ataque del nucleófilo. Para intentar desentrañar en alguna medida el papel

que juegan las moléculas de agua, también se han calculado las interacciones electrostáticas de éstas con UDP-Glc y además, se ha estimado el valor de la interacción entre ellas y  $\text{COO}^-_{\text{PGA}}$ .



**Figura 33.** Contribuciones electrostáticas de cada molécula de agua y las aguas en conjunto con UDP-Glc:  $W_2$  barras azules,  $W_3$  barras amarillas,  $W_4$  barras rosas y  $W_{353}$  barras grises. La interacción de  $\{W_2-W_3-W_4-W_{353}\}$  o simplemente aguas con UDP-Glc, es representada por las barras verdes. Además se incluye el término de interacción de las aguas con  $\text{COO}^-_{\text{PGA}}$  en barras celestes. La línea negra muestra el comportamiento de  $\epsilon$  a lo largo de la reacción.

Cuando se analiza la interacción con UDP-Glc ( $\text{COO}^-_{\text{PGA}}$  está en la región MM), los valores obtenidos sugieren que las aguas en conjunto comienzan a influir cuando el enlace UDP-Glc ya está altamente disociado (barras verdes); además, la estabilización se hace considerablemente más eficiente justo cuando la conformación de PGA ( $\epsilon$ ) cambia (Figura 33). En la misma Figura 33, si miramos la interacción de las aguas con  $\text{COO}^-_{\text{PGA}}$  (barras celestes), se ve que el término electrostático aparece pronto, cuando la disociación UDP-Glc es aún pequeña y mejora cuando  $\epsilon$  alcanza un valor  $\sim 40.0^\circ$ . También se puede observar que las aguas  $W_2$  y  $W_4$ , que son las que más efecto tienen sobre el término de interacción electrostática de las aguas, generan una estabilización considerable. Esto se debe a que  $W_2$  está interaccionando de cerca con el  $\text{COO}^-_{\text{PGA}}$  desde principio a fin, y a la vez,  $O_{W_2}$  se acerca a  $(\text{C1-H1})_{\text{Glc}}$  durante la reacción (centro anomérico donde formalmente aparece la carga positiva). En el IP, la distancia  $O_{W_2}\text{-H1}_{\text{Glc}}$  es de 2.16 Å, un valor pequeño que sugiere la formación de una interacción de naturaleza estereoelectrónica. Un análisis de interacciones de orbitales moleculares, encontró ésta interacción entre el orbital del par de electrones solitarios del  $O_{W_2}$  (Lone Pair o LP en inglés) y el orbital antienlazante del enlace  $(\text{C1-H1})_{\text{Glc}}$  (Bond Donor o BD\* en inglés).



**Figura 34.** Representación de la interacción entre un orbital natural antienlazante correspondiente a  $(C1-H1)_{Glc}$  ( $BD^*_{(C1-H1)}$ ) y el LP de  $O_{W2}$  ( $LP_{W2}$ )

Como se puede observar en la Figura 34, existe un solapamiento entre los lóbulos de  $BD^*_{(C1-H1)}$  y  $LP_{W2}$ , que se representa a través de las densidades rosas. Este cálculo indica que existe una interacción que no es puramente dipolar sino que implicaría interacción de orbitales. Así, es plausible que la reorientación de esta molécula de agua sea impulsada por efectos estereoelectrónicos que podrían ayudar a estabilizar la especie  $UDP\cdots OC$ .

Por otra parte,  $W_4$  también hace de puente entre los dos sustratos, en este caso mediante cercanas interacciones con  $COO^-_{PGA}$  y un enlace de hidrógeno con  $O2_{Glc}$ . El efecto de  $W_{353}$  es pequeño, lo cual resulta normal pues  $W_{353}$  sólo forma un enlace de hidrógeno con  $W_2$ , el cual se mantiene a lo largo de todo el camino de reacción; la  $W_3$  muestra una discreta pero desfavorable interacción justo después de que  $PGA$  ha cambiado su conformación (ver  $\epsilon$ , Figura 33).

Luego, al evaluar las interacciones de las aguas con ambos sustratos (Figura 35), observamos la misma tendencia, las aguas actúan en conjunto más eficientemente cuando la disociación  $UDP-Glc$  es alta pero sin desmedro de lo anterior, también hay una moderada estabilización previa, la cual se debe principalmente a  $W_2$  y  $W_4$ , las cuales continúan siendo las dos moléculas de agua que más estabilizan en las especies oxocarbenio.





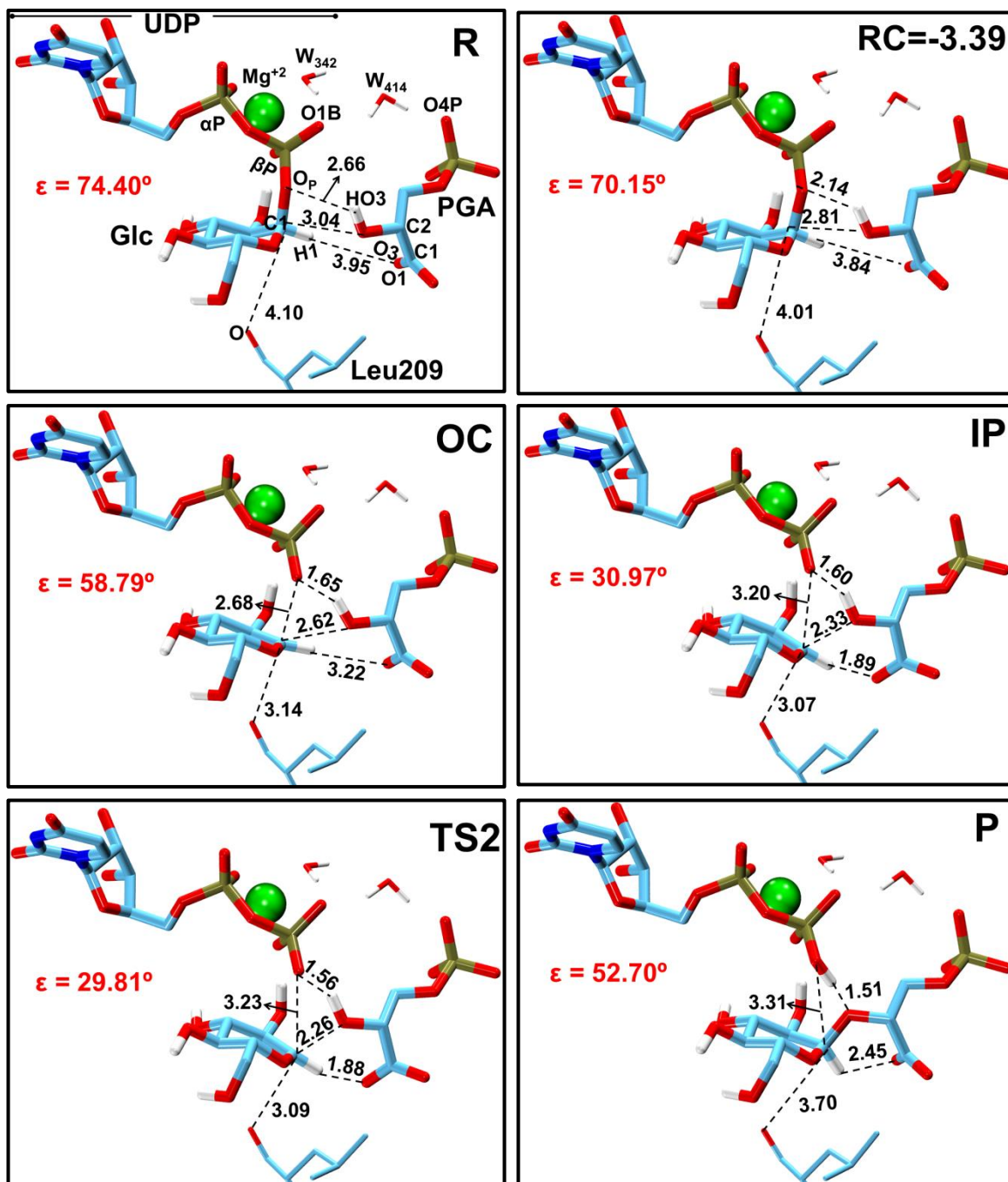


desde 3.03 Å en el reactivo hasta 1.93 Å en el intermedio. Las contribuciones tienen que ver con la estabilización del grupo saliente UDP (Arg261 y Arg259), con PGA (Arg185) y también con la estabilización del oxocarbenio (Asp134, Glu232 y Leu209). La Leu209 en la cara  $\beta$  del azúcar, da una discreta estabilización de la especie oxocarbenio a través de su grupo carbonilo.

#### **4.3.2 Modelo 2 o Complejo Catalítico sin Aguas en el Centro Activo: Mecanismo de Ataque Frontal**

Como se indicó en la sección 4.2 (apartado 4.2.6), el Modelo 2 es análogo al complejo de Michaelis publicado bajo el código PDB 4Y6N, por lo que no contiene moléculas de agua en las cercanías del carbono anomérico ni de PGA. Igual que para el Modelo 1, se utilizó una coordenada de reacción triple para conducir el sistema desde reactivos a productos, esto es,  $RC = [d(O_P-C1_{Glc}) - d(O3_{PGA}-C1_{Glc}) - d(HO3_{PGA}-O_P)]$ . Las geometrías de mínimos y estados de transición han sido calculadas al nivel QM(BP86/SVP)/MM(CHARMM), luego se han calculado las energías de cada una de las estructuras del perfil empleando el nivel QM(M05-2X/TZVP)/MM(CHARMM). Los estados de transición y mínimos optimizados han sido caracterizados con cálculos de frecuencias y cargas, indicándose en cada caso, la conformación que adquiere el azúcar a través de los parámetros Cremer-Pople. (Tabla 9)

En el Reactivo del Modelo 2, el  $O3_{PGA}$  está a 3.04 Å del  $C1_{Glc}$  y  $HO3_{PGA}$  se orienta hacia  $O_P$ , quedando a una distancia de 2.66 Å. Además vale la pena mencionar la distancia entre el grupo carboxilato de PGA y el hidrógeno enlazado al carbono anomérico de la glucosa ( $O1_{PGA}-H1_{Glc}$ ), que tiene un valor de 3.95 Å. La distancia es aún grande en el reactivo (incluso mayor que en el Modelo 1), lo cual tiene relación con la conformación adoptada por el carboxilato respecto al resto del aceptor PGA, mostrando un ángulo de torsión  $\epsilon$  igual a 74.40° (casi 15° más grande que en el Modelo 1)



**Figura 38.** Representación de las reorganizaciones atómicas claves a lo largo del camino de reacción. Cada estructura corresponde a un punto optimizado en el orden: Reactivo (R), otra configuración del R que corresponde al punto -3.39 Å de la RC, la especie oxocarbonio de más alta energía (OC), el intermedio IP oxocarbonación, el estado de transición 2 (TS2) y el producto (P).

Continuando con la descripción mecanística, se observa que la distancia  $C1_{\text{Glc}}-O_{\text{P}}$  se va alargando mientras las otras distancias involucradas en la RC,  $O3_{\text{PGA}}-C1_{\text{Glc}}$  y  $HO3_{\text{PGA}}-O_{\text{P}}$ , van disminuyendo. De hecho en el punto RC= -3.39 Å, que corresponde aún a una estructura de reactivo ( $C1_{\text{Glc}}-O_{\text{P}} = 1.55$  Å), la distancia  $HO3_{\text{PGA}}-O_{\text{P}}$  se ha reducido en 0.52 Å

consiguiendo quedar en 2.14 Å; como ya se ha comentado para el Modelo 1, esta interacción entre sustrato aceptor y dador se ha encontrado en todas las GTs-ret estudiadas hasta la fecha, y resulta ser fundamental para que la ruptura nucleótido-azúcar ocurra eficientemente (catálisis asistida por sustratos).<sup>111,112,116,124,125,131,268</sup> Otras interacciones que participan de la catálisis asistida por sustratos son por ejemplo: C1<sub>Glc</sub>-O1<sub>PGA</sub> y HO2<sub>Glc</sub>-O2B<sub>UDP</sub>.

**Tabla 9.** Varias características importantes de los puntos estacionarios se resumen en esta tabla, incluyendo las energías relativas ( $\Delta E$ ) en kcal/mol y calculadas al nivel QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM). Las unidades de medida para las distancias, frecuencias, cargas y los parámetros de Cremer-Pople ( $\varphi$  y  $\theta$ ), corresponden a Å, cm<sup>-1</sup>, unidades atómicas (u.a) y grados (°), respectivamente.

|   | Reactivo                    | OC             | IP  | TS2                         | Producto                    |
|---|-----------------------------|----------------|---|-----------------------------|-----------------------------|
| C1 <sub>Glc</sub> -O <sub>P</sub>                                       | 1.50                        | 2.68           | 3.20  | 3.23                        | 3.31                        |
| C1 <sub>Glc</sub> -O3 <sub>PGA</sub>                                    | 3.04                        | 2.62           | 2.33  | 2.26                        | 1.45                        |
| C1 <sub>Glc</sub> -O5 <sub>Glc</sub>                                    | 1.38                        | 1.28           | 1.28  | 1.29                        | 1.41                        |
| C1 <sub>Glc</sub> -O1 <sub>PGA</sub>                                    | 5.04                        | 4.33           | 2.98  | 2.96                        | 3.19                        |
| H1 <sub>Glc</sub> -O1 <sub>PGA</sub>                                    | 3.95                        | 3.22           | 1.89  | 1.88                        | 2.45                        |
| HO2 <sub>Glc</sub> -O2β <sub>UDP</sub>                                  | 1.54                        | 1.48           | 1.53  | 1.55                        | 1.74                        |
| Torsión $\epsilon$  | 74.40                       | 58.79          | 30.97                                       | 29.81                       | 52.70                       |
| C1 <sub>Glc</sub> -O <sub>Leu209</sub>                                  | 4.10                        | 3.14           | 3.07  | 3.09                        | 3.70                        |
| O <sub>P</sub> -HO3 <sub>PGA</sub>                                      | 2.66                        | 1.65           | 1.60  | 1.56                        | 1.03                        |
| O3 <sub>PGA</sub> -HO3 <sub>PGA</sub>                                   | 0.98                        | 1.01           | 1.02  | 1.02                        | 1.51                        |
| O1β-H2 <sub>W342</sub>  | 1.87                        | 1.59           | 1.57  | 1.57                        | 1.62                        |
| O1β-H1 <sub>W414</sub>  | 2.10                        | 1.90           | 1.81  | 1.80                        | 1.80                        |
| O4P <sub>PGA</sub> -H2 <sub>W414</sub>                                  | 1.70                        | 1.70           | 1.69  | 1.68                        | 1.67                        |
| q(C1 <sub>Glc</sub> )   | 0.35                        | 0.56           | 0.54  | 0.54                        | 0.34                        |
| q(H1 <sub>Glc</sub> )   | 0.25                        | 0.30           | 0.30  | 0.30                        | 0.21                        |
| q(O5 <sub>Glc</sub> )   | -0.48                       | -0.37          | -0.38                                       | -0.39                       | -0.51                       |
| $\Delta q$ (C1 <sub>Glc</sub> + H1 <sub>Glc</sub> + O5 <sub>Glc</sub> ) | 0.00                        | 0.37           | 0.34  | 0.33                        | -0.08                       |
| $\varphi$   | 143.17                      | 238.574        | 259.87                                      | 261.85                      | 258.37                      |
| $\theta$  | 6.01                        | 35.85          | 43.16                                       | 42.74                       | 21.97                       |
| Conformación  | <sup>4</sup> C <sub>1</sub> | <sup>4</sup> E | <sup>4</sup> E- <sup>4</sup> H <sub>5</sub> | <sup>4</sup> H <sub>5</sub> | <sup>4</sup> C <sub>1</sub> |
| $\Delta E$  | <b>0.00</b>                 | <b>21.91</b>   | <b>21.32</b>                                | <b>21.83</b>                | <b>1.96</b>                 |
| Frecuencias   |                             |                | Reales                                      | 42.35 <i>i</i>              |                             |

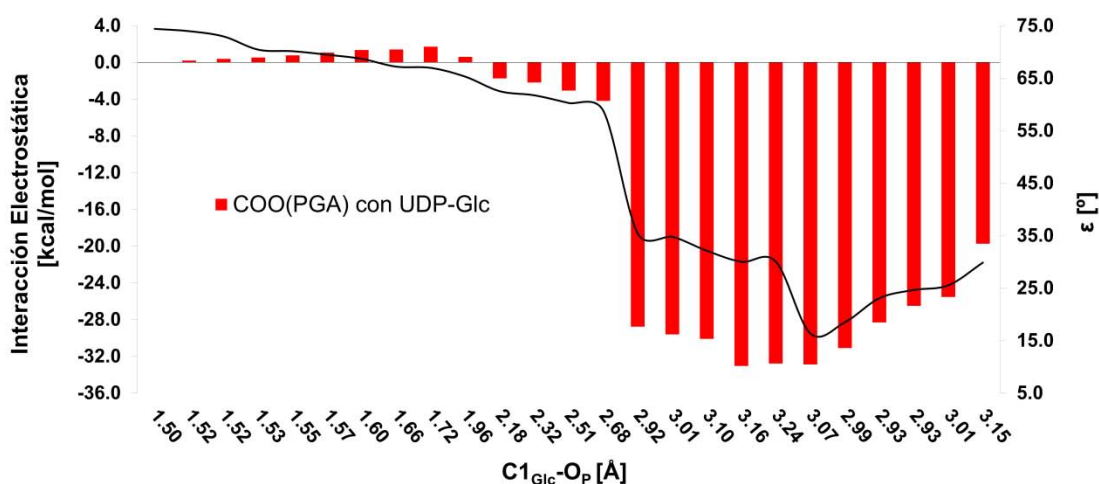
El proceso inicial implica un incremento continuo y gradual de energía hasta alcanzar el OC con una energía de 21.91 kcal/mol, aunque no se ha conseguido caracterizarlo con el cálculo de frecuencias como un punto de silla (Tabla 9). En esta especie el enlace UDP-Glc ya se ha roto (2.68 Å) mientras que O3<sub>PGA</sub>-C1<sub>Glc</sub> y HO3<sub>PGA</sub>-O<sub>P</sub> se encuentran a 2.62 Å y 1.65 Å, respectivamente. Adicionalmente, la distancia (O5-C1)<sub>Glc</sub> es de 1.28 Å y su conformación corresponde a una de tipo <sup>4</sup>E, por lo que nos encontramos ante una especie oxocarbenio que exhibe un incremento de carga ( $\Delta q$ ) de 0.37 u.a. (Figura 38). Sobrepasando este punto, el sistema evoluciona hacia un intermedio IP de alta energía, un mínimo que tiene una energía de 21.32 kcal/mol, apenas 0.6 kcal/mol más baja que la especie OC. En esta estructura es apropiado recalcar que las distancias involucradas en la coordenada de reacción, C1<sub>Glc</sub>-O<sub>P</sub>, O3<sub>PGA</sub>-C1<sub>Glc</sub> y HO3<sub>PGA</sub>-O<sub>P</sub>, se han modificado de manera gradual hasta 3.20 Å, 2.33 Å y 1.60 Å, respectivamente, mientras que H1<sub>Glc</sub>-O1<sub>PGA</sub> ha disminuido en 1.33

Å y  $\epsilon$  en  $27.82^\circ$  respecto al OC (Tabla 9 y Figura 38). Este es un IP altamente disociado, previo al paso de ataque nucleofílico, que se corresponde con un primera etapa de tipo  $D_{N_{SS}}^*A_N^\ddagger$ . La disminución abrupta de la distancia  $H1_{Glc}-O1_{PGA}$  en el IP, denota la necesidad de estabilización de la especie oxocarbenio cuyo incremento de carga positiva en el centro anomérico es de 0.34 u.a, 0.03 u.a más bajo que en la especie OC previa. Además, la disminución de esta distancia ocurre de manera simultánea con el alargamiento de la distancia  $C1_{Glc}-O_P$  hasta 3.20 Å, y el acercamiento de PGA al carbono anomérico. Así, se pasa de una especie UDP----OC menos disociada (OC,  $C1_{Glc}-O_P = 2.68$  Å), a una especie UDP----OC altamente disociada (IP,  $C1_{Glc}-O_P = 3.20$  Å). Por último, con la excepción del cambio en el  $\Delta q$ , el resto de características asociadas a la especie oxocarbenio, como la distancia  $(O5-C1)_{Glc}$  o los parámetros de Cremer-Pople permanecen prácticamente invariables.

Luego del intermedio IP, la coordenada de reacción nos guía hasta un segundo estado de transición (TS2), que tiene una energía de 21.83 kcal/mol. La nomenclatura IUPAC para este tipo de mecanismo en dos etapas es  $D_{N_{SS}}^*A_N^\ddagger D_hA_{xh}$ , el mismo tipo de mecanismo caracterizado para el Modelo 1.

#### 4.3.2.1 Interacciones Favorables para la Catálisis en el Modelo 2.

Tal como se hizo antes, ahora se analizan las contribuciones electrostáticas de  $COO^-_{PGA}$  (Figura 39), y las de los residuos enzimáticos (Figura 40) a lo largo del camino de reacción.

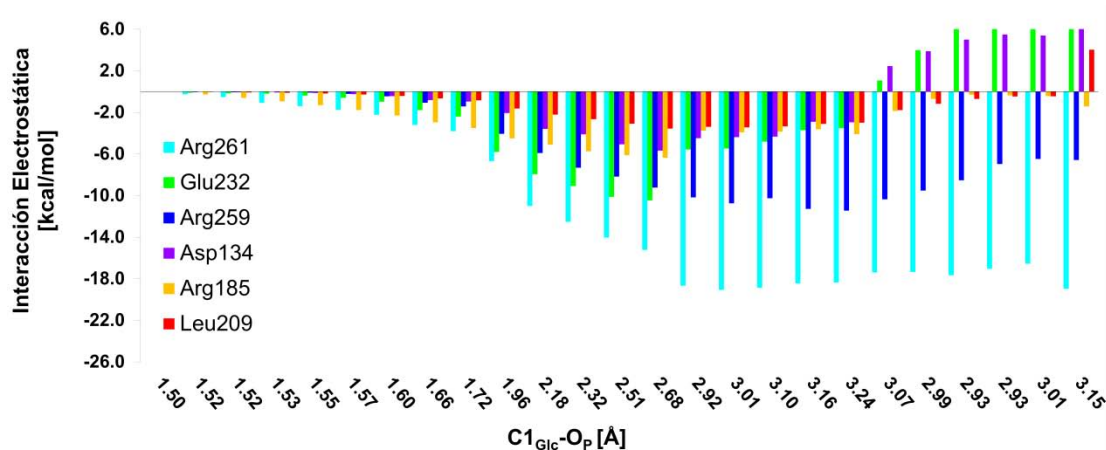


**Figura 39.** Contribuciones electrostáticas de  $COO^-_{PGA}$  con el sustrato dador UDP-Glc a lo largo de la reacción de transferencia de glucosa. En el eje vertical secundario (derecha), se señalan los valores de  $\epsilon$  y la línea negra representa su variación a lo largo de la reacción.

Se observa que la variación de  $\epsilon$ , así como el efecto electrostático del  $COO^-_{PGA}$  es diferente de lo obtenido en el Modelo 1, aunque en términos cualitativos ambos análisis indican que

el  $\text{COO}^-_{\text{PGA}}$  mejora sus interacciones con UDP-Glc cuando su conformación cambia. De hecho, en este modelo sin aguas el cambio de conformación de PGA al alcanzar la especie OC ( $\text{C1}_{\text{Glc}}-\text{O}_P$  2.68 Å), indica un valor de  $\epsilon = 58.79^\circ$ . Esto es prácticamente igual al valor que tiene  $\epsilon$  en el Reactivo del Modelo 1 ( $59.60^\circ$ ). Esto sugiere que para que PGA y UDP-Glc se dispongan en una orientación relativa competente para la catálisis, PGA debe alterar su conformación interna. Además, es lógico que un grupo de esta naturaleza pueda estabilizar en alguna medida las especies oxocarbenio y observando la Figura 39, la habilidad estabilizadora se relaciona directamente con su conformación interna.

A continuación, se dan las contribuciones electrostáticas de diferentes residuos.



**Figura 40.** Contribuciones electrostáticas de algunos residuos que interaccionan directamente con los sustratos a lo largo de la reacción de transferencia de Glucosa.

Al igual que lo observado para el Modelo 1, los residuos que más contribuyen a lo largo de la reacción de transferencia de Glucosa son la Arg261 y la Arg259, las cuales interaccionan con el UDP. El Asp134, el Glu232 y la Leu209 son capaces de estabilizar las especies OC discretamente durante la etapa de disociación del sustrato dador, y la Arg185 hace lo mismo con el sustrato aceptor PGA.



La comparación entre el perfil de energía libre y el de energía potencial calculados a este nivel (Figura 41), indica que la etapa de rotura del enlace UDP-Glc, aproximadamente hasta la RC  $\approx -1.60 \text{ \AA}$ , no es sustancialmente afectada. En la zona en la que coexisten las especies UDP---OC, hasta RC  $\approx -1.00 \text{ \AA}$ , se observa que el perfil de energía libre se hace ligeramente más ancho. Esto último sugiere un mecanismo en dos etapas con un intermedio de reacción, un resultado semejante a lo que se obtiene utilizando métodos de cálculo más sofisticados como DFT para la región QM. Por otra parte, la región en la que se explora el ataque de PGA no queda bien descrita pues UDP se acerca nuevamente al carbono anomérico y por consiguiente se aproxima también a PGA. En términos de energía potencial estas estructuras no son las de más alta energía ya que al estarse formando el producto, la energía comienza a decrecer. Además de esto, la transferencia protónica ocurre sincrónicamente con el ataque nucleofílico, lo cual parece ser un proceso energéticamente favorable a este nivel de teoría. En términos de energía libre, la exploración de esta región es más compleja que las anteriores y también de más alta energía; esto es razonable debido al impedimento estérico inherente al tipo de estructura generada. Aun así, la región previa al ataque nucleofílico muestra que el perfil de energía libre y energía potencial coinciden aceptablemente.

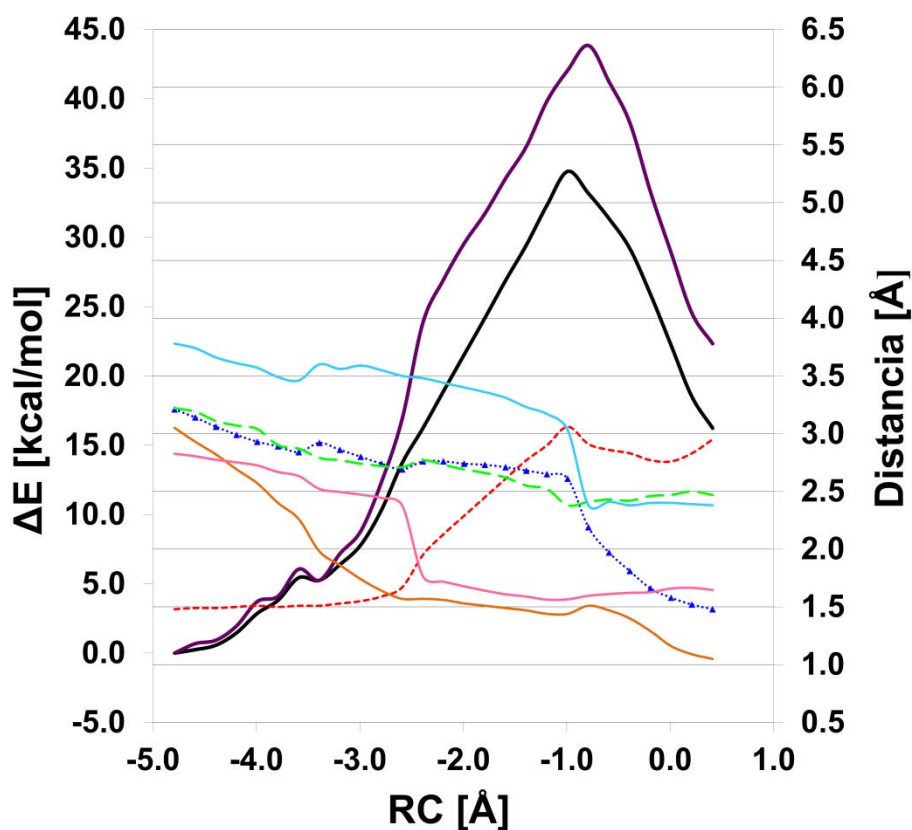
Finalmente y retomando los cálculos QM(DFT)/MM(CHARMM), la barrera de energía total obtenida es de 21.91 kcal/mol, un valor semejante a la barrera de energía libre de activación fenomenológica de 17.5 kcal/mol que ha sido calculada con un  $k_{\text{cat}}$  de  $800 \text{ min}^{-1}$ . De esta manera, concluimos que Gpgs sigue un mecanismo de ataque frontal con formación de un IP de alta energía y altamente disociado.

### 4.3.3 Híbrido 1.A: Mecanismo de Ataque Frontal

Para construir este modelo se han tomado las coordenadas del complejo catalítico competente (Modelo 1), con la excepción de las coordenadas que corresponden al PGA y a las aguas  $W_2$ ,  $W_3$  y  $W_4$ , las cuales se han reemplazado por las coordenadas que estas moléculas tienen en el Modelo 0. De este modo, se puede analizar el efecto de modificar únicamente las coordenadas de ese conjunto de átomos, dejando el resto del sistema como en el Modelo 1. Para llevar a cabo el estudio mecanístico, se realizó el mismo procedimiento ya descrito en el apartado 4.2.3. En la Figura 42 se muestra el Reactivo del Híbrido 1.A ya minimizado a nivel QM(BP86/SVP)/MM(CHARMM).







**Figura 43.** Curvas de energía potencial para el mecanismo de ataque frontal en el Híbrido 1.A a nivel QM(BP86/SVP)/MM(CHARMM) (línea negra) y recalculada a nivel QM(M05-2X/TZVP)/MM(CHARMM) (línea morada). También se señalan algunas distancias importantes como: C1<sub>Glc</sub>-O<sub>P</sub> en línea discontinua roja, C1<sub>Glc</sub>-O3<sub>PGA</sub> en línea punteada azul, HO3<sub>PGA</sub>-O<sub>P</sub> en línea naranja, C1<sub>Glc</sub>-O<sub>W2</sub> en línea discontinua verde, O1<sub>PGA</sub>-H1<sub>Glc</sub> en línea celeste y H<sub>W342</sub>-O1<sub>β</sub> en línea rosa. La barrera de energía para este proceso es de 34.8 kcal/mol a nivel QM(BP86/SVP)/MM(CHARMM) y 43.8 kcal/mol a nivel QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM).

Como puede observarse en la Figura 43, el impacto en las barreras energéticas es enorme. Debido a las diferentes características de este complejo, el primer proceso que se describe es la reorientación de HO3<sub>PGA</sub> hacia O<sub>P</sub>, lo cual requiere de aproximadamente 4 kcal/mol (RC = -3.8 Å). Sin embargo, el coste energético para alcanzar el punto en que la distancia C1<sub>Glc</sub>-O<sub>P</sub> comienza a elongarse (RC = -2.4 Å) es de casi 24 kcal/mol (energías QM(M05-2X/TZVP)). Este punto coincide con la formación de la interacción anión-dipolo O1<sub>B<sub>UDP</sub></sub><sup>-</sup> (H2)<sub>W342</sub>, que parece ser clave para la estabilización de UDP. La formación de dicha interacción no queda reflejada en una bajada de energía en el perfil, aunque si ya hubiese estado formada en el reactivo, el inicio de la rotura del enlace nucleótido-azúcar podría haberse producido más fácilmente. Es interesante comentar que en la construcción inicial del Híbrido 1.A la interacción O1<sub>β</sub>-H2<sub>W342</sub> está presente, puesto que lo estaba en el Modelo 1, pero en el proceso de optimización se pierde a consecuencia de pequeños reajustes atómicos. En particular, los mayores cambios se observan en los átomos que forman el enlace glicosídico UDP-Glc, el fosfato β y los hidrógenos de la agua W<sub>342</sub>. Otra característica curiosa de esta exploración tiene relación con el grado de disociación que alcanza UDP-Glc (C1-O<sub>P</sub> < 3.00 Å), menor grado de disociación que el observado en los modelos previos. Si el grado de disociación es menor, cuando se aproxima el ataque del nucleófilo los impedimentos estéricos pueden hacerse mayores y tener un impacto negativo en el perfil, tal como se observa en la Figura 43.



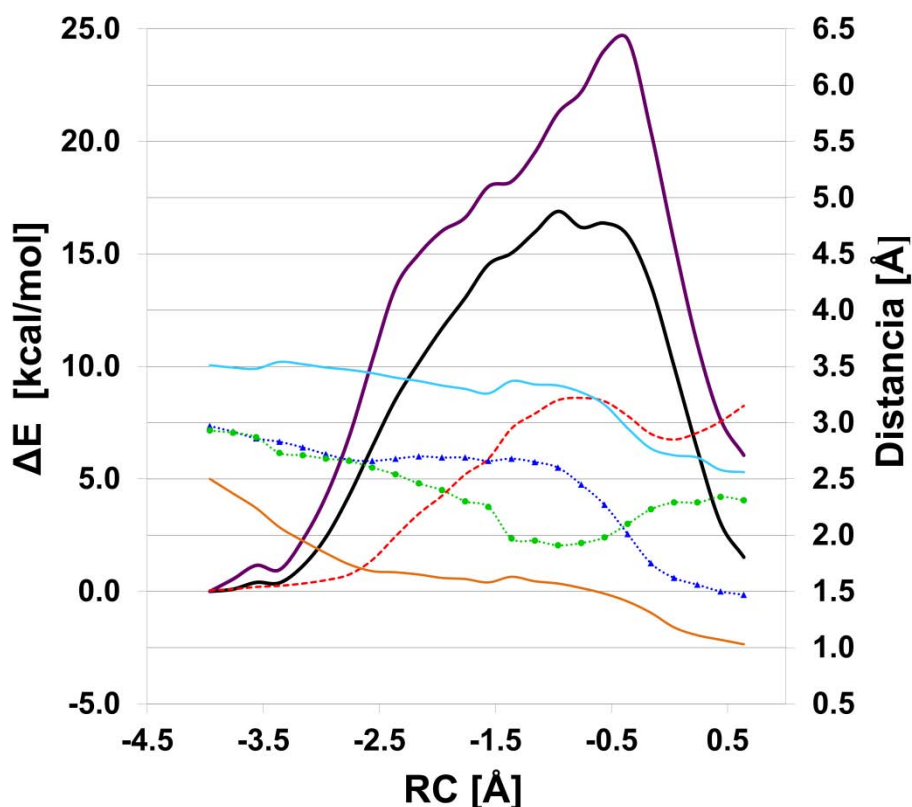
**Figura 45.** Contribución electrostática de cada una de las aguas y las aguas en conjunto con {UDP-Glc+PGA+aguas}, esto es, considerando la interacción entre aguas (barras rojas), en cada uno de los siguientes términos:  $W_2$  barras azules,  $W_3$  barras amarillas,  $W_4$  barras rosas y  $W_{353}$  barras grises. La interacción de las aguas { $W_2$ - $W_3$ - $W_4$ - $W_{353}$ } con {UDP-Glc+PGA+aguas} es representada por las barras verdes. La línea negra describe el comportamiento de  $\epsilon$ .

A partir de la gráfica se puede concluir que en este modelo el efecto de las aguas en conjunto y por separado, no es favorable a lo largo de la transferencia de glucosa. Esto puede ser una consecuencia de que ninguna de las aguas orienta su oxígeno al centro anomérico y por lo tanto, ningún tipo de interacción puede ser establecido entre ellos. Además, tampoco hay ninguna agua puente que interaccione con ambos sustratos (agua  $W_4$  en el Modelo 1). Curiosamente, los valores que toma el diedro  $\epsilon$  en este híbrido no se parecen en nada a los observados en el Modelo 1. Este notable cambio ha de tener relación con la pérdida de la orientación óptima entre PGA y UDP, viéndose afectada una de las más importantes interacciones entre ellos:  $O_P$ - $HO3_{PGA}$ . El patrón de interacciones observado ahora incluye a  $O1\beta$  posicionado a 1.89 Å de  $HO3_{PGA}$  y sin interaccionar con el agua de coordinación  $W_{342}$ . Estos cambios desfavorables de las interacciones inter-sustratos, el nuevo ordenamiento de la red de moléculas alrededor de  $COO^-_{PGA}$  y la conformación misma de PGA, confirman que se trata de una conformación no catalítica (o pre-catalítica como la bautizó el grupo de cristalografía) del complejo ternario, en el sentido que no es la óptima para que la reacción tenga lugar.

#### 4.3.4 Híbrido 1.B: Mecanismo de Ataque Frontal

Para construir este modelo se ha considerado la configuración del complejo catalítico competente (Modelo 1) y se han reemplazado las coordenadas del PGA por las que se encuentran en la estructura cristalográfica del denominado, complejo de Michaelis ( $GpgS \cdot Mn^{2+} \cdot UDP-Glc \cdot PGA-1$ , código PDB 4Y6N). También se han utilizado las coordenadas de los oxígenos de las aguas  $W_2$ ,  $W_3$  y  $W_4$ , que formarían parte de la estructura 4Y6N si la densidad electrónica hubiese asignada. El reactivo obtenido presenta características ligeramente diferentes en cuanto a distancias inter-sustratos pero, en el caso de la red de aguas, la orientación relativa entre ellas se ve algo más alterada, aunque las coordenadas de sus átomos de oxígeno son relativamente parecidas (Figura 46).

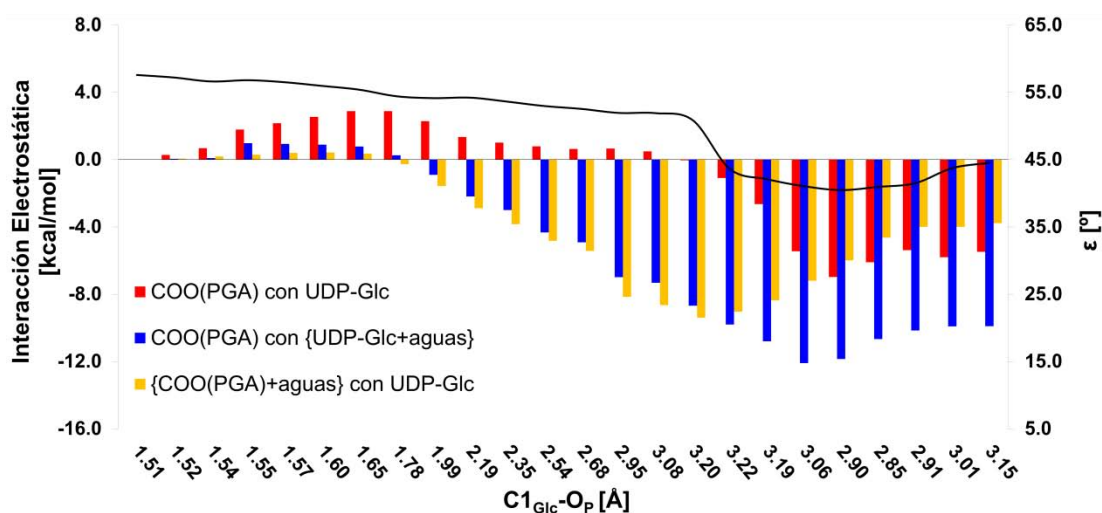




**Figura 47.** Curvas de energía potencial para el mecanismo de ataque frontal en el Híbrido 1.B a nivel QM(BP86/SVP)/MM(CHARMM) (línea negra) y recalculada a nivel QM(M05-2X/TZVP)/MM(CHARMM) (línea morada). También se señalan algunas distancias importantes como: C1<sub>Glc</sub>-O<sub>P</sub> en línea discontinua roja, C1<sub>Glc</sub>-O3<sub>PGA</sub> en línea punteada azul, HO3<sub>PGA</sub>-O<sub>P</sub> en línea naranja, H1<sub>Glc</sub>-O<sub>W2</sub> en línea y círculo verde y O1<sub>PGA</sub>-H1<sub>Glc</sub> en línea celeste. La barrera de energía potencial a nivel QM(BP86/SVP)/MM(CHARMM) es de 16.89 kcal/mol y a nivel QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM) es de 24.53 kcal/mol.

A pesar de que las interacciones inter-sustratos son muy semejantes a las ya descritas para el Modelo 1, una vez que se forma al especie UDP---OC altamente disociada (RC = -1.0 Å) y PGA comienza a acercarse, la energía sube gradualmente hasta alcanzar un valor máximo de casi 25 kcal/mol (línea morada). Este valor es ~ 6 kcal/mol mayor que en el TS2 del Modelo 1 (18.8 kcal/mol), pero mucho menor que el Híbrido 1.A (43.8 kcal/mol). La mayor cantidad de energía requerida para el ataque nucleofílico, podría deberse en parte a que W<sub>2</sub> está demasiado cerca del carbono anomérico cuando el PGA debe aproximarse. A la vez, esto también podría relacionarse con el hecho de que en esta simulación no hay una sola estructura en la que el diedro ε llegue a tomar valores < 40.00°, a diferencia de lo observado en el Modelo 1. La estructura del IP en el Modelo 1, es la que muestra el valor más pequeño de ε (37.71°), el cual alcanza un valor de 41.80° en el TS2. La estructura que correspondería al TS2 en este modelo híbrido muestra un valor de ε muy semejante de 41.05°, pero no se ha muestreado una estructura con un diedro menor a ese. Estos resultados sostienen las hipótesis previamente establecidas: la conformación de PGA es importante en orden llevar a cabo del ataque nucleofílico eficientemente y las aguas pueden modular ese cambio de conformación necesario para la catálisis.

Debido a esto, se decide hacer la estimación de contribuciones electrostáticas nuevamente, enfocándonos en  $\text{COO}^-_{\text{PGA}}$  y las aguas.



**Figura 48.** Contribución electrostática de  $\text{COO}^-_{\text{PGA}}$  con el sustrato dador y aguas {UDP-Glc+aguas} y de la interacción de { $\text{COO}^-_{\text{PGA}}$ +aguas} con UDP-Glc a lo largo de la reacción de transferencia de glucosa. En el primer caso, el término de interacción no incluye a las moléculas de agua (barras rojas) y en el segundo caso, sí se incluye (barras azules). La interacción { $\text{COO}^-_{\text{PGA}}$ +aguas} con UDP-Glc, representa a  $\text{COO}^-_{\text{PGA}}$  y aguas actuando como conjunto (barras amarillas). La línea negra representa la variación de  $\epsilon$  a lo largo de la reacción.

Puede observarse que el gráfico obtenido es muy parecido al del Modelo 1. Los valores apuntan a que { $\text{COO}^-_{\text{PGA}}$ +aguas} actuando como un conjunto y también sólo  $\text{COO}^-_{\text{PGA}}$  son capaces de mejorar sus interacciones con UDP-Glc y {UDP-Glc+aguas}, respectivamente. Nuevamente es posible concluir que el papel de las aguas parece estar asociado al cambio conformacional de PGA pues los valores siempre son mejores cuando el diedro se ha hecho más pequeño. Probablemente parte del papel de las moléculas de agua sea relajar este cambio de conformación de PGA (representado por  $\epsilon$ ) durante la reacción de transferencia a través de una solvatación de  $\text{COO}^-_{\text{PGA}}$  que acompaña sus movimientos. Analizando las interacciones que establece  $\text{COO}^-_{\text{PGA}}$  con la proteína esta idea se configura como bastante coherente pues tal como se indicó antes, el único residuo que interacciona con este fragmente cargado de PGA es la Thr187.



### 4.3.5 Modelo 1, Híbrido 1.A y 1.B: Mecanismo de Ataque del agua $W_2$ o Hidrólisis de UDP-Glc

Para llevar a cabo esta exploración, se usó la coordenada de reacción  $RC = [d(O_P-C1_{Glc}) - d(O_{W_2}-C1_{Glc}) - d(H_{W_2}-O1_{PGA})]$ . La simulación del ataque de  $W_2$  se exploró usando todos los modelos adecuados sin embargo, sólo se consiguió en el caso del Híbrido 1.A y del complejo no catalítico (ver Tabla 11).

**Tabla 11.** Barreras de energía potencial para el ataque de  $W_2$  y el mecanismo de ataque frontal obtenidas en base a los diferentes modelos estructurales investigados. Las energías QM(BP86/SVP)/MM(CHARMM) se dan en kcal/mol y corresponden a estructuras del camino de reacción.

| Modelo      | Ataque de $W_2$     |            | Ataque Frontal        |                 |                       |            | Reactivo<br>$C1_{Glc}-O_{W_2}$ [Å] |
|-------------|---------------------|------------|-----------------------|-----------------|-----------------------|------------|------------------------------------|
|             | $\Delta E^\ddagger$ | $\Delta E$ | $\Delta E^\ddagger_1$ | $\Delta E^{IP}$ | $\Delta E^\ddagger_2$ | $\Delta E$ |                                    |
| Modelo 0    | 34.6                | 13.0       | 38.7                  | -               | -                     | -3.7       | 3.86                               |
| Modelo 1    | N/A*                | N/A        | 15.9                  | 12.4            | 11.8                  | -2.7       | 5.10                               |
| Híbrido 1.A | 32.6                | 8.9        | -                     | -               | 34.8                  | 16.2       | 3.96                               |
| Híbrido 1.B | N/A**               | N/A        | 16.9                  | 16.2            | 16.4                  | 1.5        | 3.71                               |

\* El proceso comienza con la disociación del sustrato dador UDP-Glc. A medida que  $W_2$  se aproxima a  $C1_{Glc}$ , el  $H1_{Glc}$  va orientándose hacia  $O_{W_2}$  de modo que el ataque del  $W_2$  sobre el  $C1_{Glc}$  no logra proceder.

\*\* Tal como ocurrió en el Modelo 1, el azúcar adopta una configuración en la  $W_2$  no consigue acercarse al  $C1_{Glc}$  por la cara  $\beta$  de Glucosa debido a que  $H1_{Glc}$  apunta directamente a  $O_{W_2}$ . Sin embargo, la  $W_2$  logra ponerse por arriba del  $H1_{Glc}$  y consigue proceder con el ataque nucleofílico por la cara  $\alpha$  de Glucosa, aunque con una barrera 60 kcal/mol.

En el Modelo 0 y el Híbrido 1.A el ataque nucleofílico del agua se consigue simular, pero con una barrera > 30 kcal/mol. En vista de tales valores que indican que la hidrólisis no podría ser catalizada por GpgS, no se calculan las usuales correcciones energéticas al nivel de teoría QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM). La situación observada en el Modelo 1 y el Híbrido 1.B es diferente pues una vez formada la especie UPD----OC altamente disociada, el oxígeno de  $W_2$  se orienta hacia el  $H1_{Glc}$  (Figura 50), semejante a como se ve en el IP del Modelo 1.





## 4.4 Conclusiones

La discusión mecanística que existe respecto a la síntesis del enlace glicosídico con retención de la configuración del carbono anomérico, se desarrolla en torno a la posibilidad de un ataque frontal del aceptor, o por la misma cara por la que UDP sale, y el mecanismo de doble desplazamiento. En el último, es fundamental que un residuo enzimático adecuadamente ubicado en la cara  $\beta$  del azúcar pueda actuar como nucleófilo formando el intermedio covalente CGE.

El mecanismo de reacción a través del cual GpgS cataliza la formación del enlace glicosídico Glc $\alpha$ 1-3PGA con retención de la configuración del carbono de anomérico de la Glucosa, ha sido investigado en este trabajo. Nuestros resultados muestran la existencia de un mecanismo de ataque de frontal con formación de un intermedio caracterizado como mínimo en la superficie de energía potencial, el IP. Este intermedio es un complejo UDP----oxocarbenio, a partir del cual se pueden inferir las características electrónicas y estructurales que son necesarias para que el controvertido ataque frontal proceda eficientemente. La reacción de transferencia comienza con la disociación del sustrato dador UDP-Glc, el cual dará lugar al IP. Durante esta disociación se generan especies cargadas (UDP $^{\delta-}$  y Glc $^{\delta+}$ ) que deben ser estabilizadas por la enzima. Entre los residuos que más contribuyen a estabilizar estas especies se pueden mencionar por ejemplo; a la Arg261 y a la Arg259, que interactúan con UDP; al Glu232, la Leu209 y el COO $^-$ <sub>PGA</sub> que interactúan con Glc; y a la Arg185 que interactúa con PGA. En GpgS, la cara  $\beta$  de la Glucosa es ocupada por la Leu209, que tiene su oxígeno carbonílico apuntando al centro del anillo pero cuya habilidad nucleofílica es bastante escasa sin una apropiada activación. En la vecindad de la cara  $\beta$  del azúcar, en una zona muy próxima al grupo carboxilato de PGA (COO $^-$ <sub>PGA</sub>), se observaba experimentalmente una densidad electrónica atribuible a moléculas de agua, sin embargo la resolución a la que se obtuvieron las estructuras (2.4 Å) no permitía asignar de forma inequívoca esta densidad, por lo que las estructuras publicadas no la contienen. A pesar de esto, y tal como se explicó en la Introducción, nuestra investigación contempla la presencia de estas aguas (Modelos 0 y 1, junto a los Híbridos 1.A y 1.B) y no lo hace en el Modelo 2. Como una de estas aguas, W<sub>2</sub>, se ubicaba muy cerca del carbono anomérico de la Glucosa (4.16  $\pm$  0.47 Å en los modelos que la poseen), también se exploró la hidrólisis de UDP-Glc usando W<sub>2</sub> como nucleófilo, comprobando que la hidrólisis no sería posible debido a las altas barreras de energía obtenidas (> 30 kcal/mol). Esto también está en acuerdo con los ensayos experimentales que no detectaron actividad hidrolítica para GpgS. De esta manera, el mecanismo de ataque frontal sería el camino preferido por esta enzima para catalizar la reacción de transferencia de Glucosa.

El estudio interdisciplinario de las GTs-ret en los últimos años, ha estado aportando importante información respecto a la importancia de las interacciones **inter-sustratos**, que en estas enzimas resulta ser clave no sólo en la unión del sustrato aceptor sino que también propiamente en la catálisis. En este contexto, nuestro estudio también encuentra que la interacción entre el O<sub>P</sub> y el hidroxilo nucleofílico (catálisis asistida por sustratos) es clave en el mecanismo pues tiene un gran impacto en las barreras energéticas del proceso (ver Modelo 0 e Híbrido 1.A). Las interacciones que establece el UDP con la esfera de

coordinación del  $Mg^{2+}$  resultan ser también muy importantes dado que afectan la orientación relativa entre UDP y Glucosa. Adicionalmente, nuestros resultados indican que la conformación del sustrato aceptor PGA ( $\epsilon$ ) tiene un notable impacto en el mecanismo, especialmente en la etapa de ataque nucleofílico. La comparación de los Modelos 1 y 2, así como los análisis electrostáticos que se han llevado a cabo, permiten concluir que las aguas podrían estar ayudando a estabilizar las especies oxocarbenio (interacción  $O_{W2}-H1_{Glc}$ ) y además, facilitando las interacciones entre UDP y PGA. Este último rol lo atribuimos a que las aguas forman una pseudo-esfera de solvatación que acompaña los movimientos de  $COO^-_{PGA}$  y que al estar interaccionando al mismo tiempo con la Glucosa, ayudan a que los sustratos se mantengan cerca durante el camino de reacción.

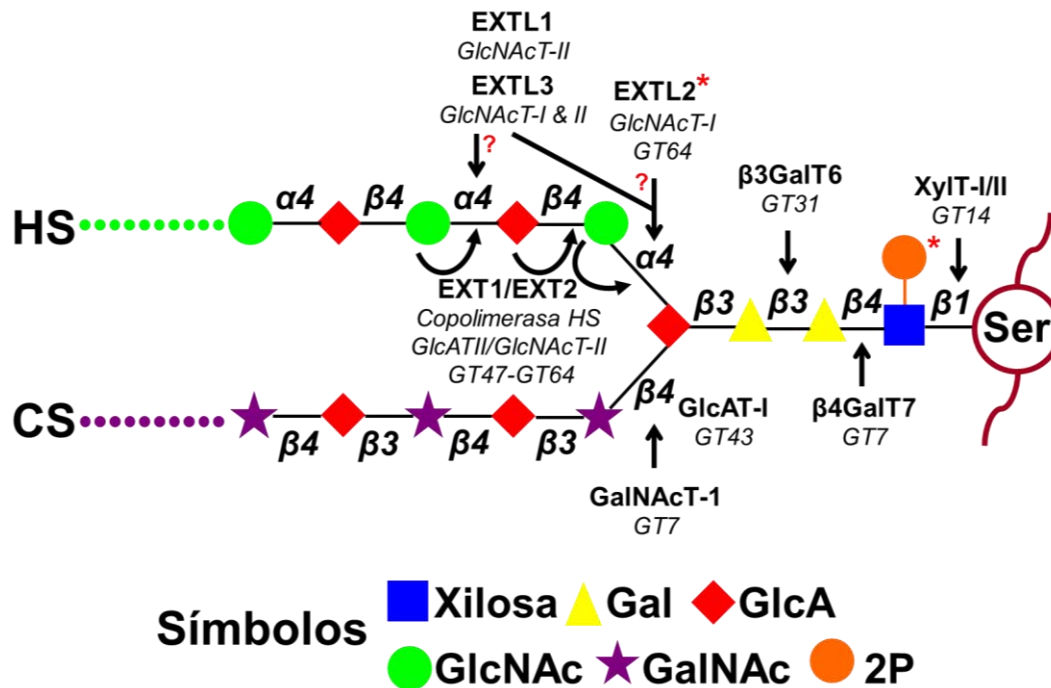
A la luz de nuestros resultados teóricos, el mecanismo sin la presencia de las moléculas de agua parece factible, aunque con moléculas de agua también y, de hecho, implicando una menor barrera energética. Dichas aguas, si están realmente presentes, deben ser muy móviles, por lo que no es de extrañar que la resolución de las estructuras cristalográficas no haya sido hasta ahora suficiente para probar la existencia de dichas aguas mediante la identificación de la densidad electrónica de las mismas. Por otra parte, la entrada de algunas aguas en el centro activo parece una situación muy probable debido a la carga de PGA. De esta manera, la presencia o no de las aguas en la vecindad del carbono anomérico y  $COO^-_{PGA}$  no alteran las conclusiones mecanísticas sobre GpgS. Aunque en el caso de no haber aguas, la disociación UDP-Glc sería un proceso energéticamente más costoso que implicaría un acercamiento mucho más marcado  $COO^-_{PGA}$  hasta la Glucosa para conseguir una mayor estabilización de las especies oxocarbenio ( $TS2$ ,  $H1_{Glc}-O1_{PGA}=1.88 \text{ \AA}$  en el Modelo 2, y  $3.43 \text{ \AA}$  en el Modelo 1). Así GpgS cataliza un mecanismo de ataque frontal en el que una vez formado el IP, el UDP actuará como base catalítica formándose de esta manera en el nuevo enlace glicosídico.

# 5 $\alpha$ 1,4-N-Acetilhexosaminiltransferasa: EXTL2-GalNAc

## 5.1 Aspectos Generales

EXTL2 está relacionada con la biosíntesis del Heparán Sulfato (HS),<sup>269</sup> un glicosaminoglicano (GAG) que forma parte de todos los tejidos animales y se encuentra en la superficie celular o matriz extracelular como proteoglicano (PG).<sup>270</sup> Como tal, participa en una amplia variedad de procesos biológicos que incluyen la angiogénesis (proceso de formación de vasos sanguíneos), coagulación sanguínea, proliferación y diferenciación celular o la infección bacterial y viral.<sup>271,272</sup> Las interacciones que HS pueda establecer con diferentes proteínas, factores de crecimiento, morfógenos y proteínas extracelulares sientan las bases de los fenómenos biológicos relacionados con HS.<sup>273</sup> En este contexto, se ha reportado que la expresión de HS puede ser reducida en un ~ 96 % usando 4-deoxi-GlcNAc (GlcNAc se refiere a N-acetil-glucosamina), un análogo del sustrato natural GlcNAc que es necesario en la iniciación y la elongación de HS.<sup>274</sup>

Por otra parte, diferentes estudios han mostrado que mutaciones en los genes que codifican a EXT1 y EXT2, dos GTs que polimerizan las cadenas de HS, han sido identificadas como las responsables de la Exostosis Múltiple Hereditaria (HME, derivado del inglés Hereditary Multiple Exostoses).<sup>274</sup> HME es un desorden genético que se caracteriza por la formación de múltiples tumores cartilagosos que pueden llegar a ser muy dolorosos. Inicialmente, tres diferentes loci fueron relacionados con el desorden HME a través de análisis génicos: EXT1, EXT2 y EXT3, los cuales conforman la familia de genes EXT también conocida como familia supresora de tumores.<sup>275</sup> En EXT2 se detectó actividad transferasa para dos azúcares involucrados en la elongación de HS: ácido glucurónico (GlcA) y GlcNAc (Figura 51). En el caso de EXT1, se sabe que es capaz de elevar los niveles de GlcA y GlcNAc en enzimas mutantes y por consiguiente, corregir defectos en la biosíntesis de HS.<sup>276</sup>



**Figura 51.** Esquema histórico de la biosíntesis de HS en el que también se señala la enzima que inicia la ruta de biosíntesis de CS. Se llama histórico debido a que hace referencia al primer rol que se postuló para EXT2, en el que ésta iniciaba la biosíntesis de HS separándola de la ruta de CS.<sup>269</sup> Para completar el esquema, también se indica el nuevo rol postulado para EXT2 (\*), en el que \* representa el escenario en el que EXT2 añade GlcNAc al núcleo común de unión fosforilado en xilosa, lo que significa el término de la elongación de la cadena.<sup>277</sup> Los monosacáridos son representados por diferentes figuras geométricas dibujadas en distintos colores. Se utiliza un círculo naranja para indicar que la xilosa está fosforilada en la posición 2.

El núcleo natural de unión GAG-proteína, que es común para GAGs similares, es **GlcAβ1-3Galβ1-3Galβ1-4Xylβ1-O-Ser**: en el que GlcA representa el ácido glucurónico, Gal se refiere a la galactosa, Xyl es la xilosa y Ser es un residuo proteico, la serina. Si se sintetiza la fracción **GlcNAcα1-4GlcAβ1-3Galβ1-3Galβ1-4Xylβ1-O-Ser**, podría formarse HS o Heparina (H), siendo la H un GAG más sulfatado que HS y por ende de mayor peso molecular; si se forma el pentasacárido **GalNAcβ1-4GlcAβ1-3Galβ1-3Galβ1-4Xylβ1-O-Ser**, se obtendría Condroitín Sulfato (CS) o Dermatan Sulfato (DS), los cuales se diferencian entre otras cosas, por la presencia de L-glucoronato y L-iduronato, respectivamente. Aunque las GTs involucradas en la polimerización de HS se han identificado como EXT1 y EXT2, la iniciación y diferenciación de HS permanece aún con varios interrogantes. Los miembros posteriores de la familia de genes EXT, el gen EXT2 (otros genes EXT-like también forman parte de la familia EXT: EXT1 y EXT3),<sup>275</sup> codifica la enzima α1,4-N-Acetilhexosaminiltransferasa que, *in vitro*, transfiere GlcNAc y GalNAc hasta un análogo del núcleo de unión común GAG-proteína, **GlcAβ1-3Galβ1-3Galβ1-4Xylβ1-O-naftalenmetanol** y hasta el núcleo natural, respectivamente. Se trata de la GT-ret EXT2, que lleva el mismo nombre del gen que la codifica,<sup>269</sup> y aunque aún no existe conexión entre los genes EXT-like

y el desorden HME, análisis genéticos de diversos tipos de cáncer sugieren que estos genes también podrían codificar enzimas supresoras de tumores.<sup>273</sup>

La enzima EXTL2 posee una actividad transferasa del tipo GlcNAcT-I, término acuñado para indicar que la reacción de transferencia de GlcNAc ocurre si se utiliza como aceptor el núcleo de unión GAG-proteína  $\text{Glc}\alpha\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl}\beta 1\text{-O-Ser}$  o análogos de éste, pero no si se utiliza un oligosacárido más largo aunque contenga el final no reductor GlcA (actividad transferasa del tipo GlcNAcT-II).<sup>278</sup> La actividad transferasa GlcNAcT-I llevó a pensar que EXTL2 podría ser la enzima clave en la biosíntesis de HS, excluyendo CS/DS de la ruta sintética.<sup>269</sup> Más tarde se informó que los genes EXTL1 y EXTL3 codifican las enzimas del mismo nombre que transfieren GlcNAc. Mientras EXTL1 posee únicamente actividad transferasa GlcNAcT-II, EXTL3 muestra actividad transferasa GlcNAcT-I y II, siendo pues capaz de transferir GlcNAc hasta un oligosacárido que representa una cadena creciente de HS (GlcNAcT-II), y hasta un análogo del núcleo común de unión de GAGs (GlcNAcT-I). Esto también llevó a pensar que EXTL3 era capaz de participar en la iniciación y elongación de HS mientras que EXTL1 posiblemente estaba involucrada en la elongación de la cadena (Figura 51).<sup>279</sup> Debido a que la actividad transferasa de las enzimas EXTL no ha sido detectada usando el sustrato común de GAGs como aceptor, no se ha podido confirmar inequívocamente que EXTL2 o EXTL3 sea la enzima que inicia la biosíntesis de HS. Igualmente, tampoco se sabe si EXTL1 participa de la etapa de polimerización junto EXT1/EXT2. De hecho, durante años el o los roles específicos de EXTL2 en la biosíntesis de GAGs han sido elusivos a las investigaciones. Existen estudios que han aportado piezas de información que, complementadas por estudios más recientes, están siendo útiles para descifrar el rol de EXTL2. Parte de esta nueva información se basa en estudios realizados sobre un pentasacárido sintetizado naturalmente en el hígado de ratón,  $\text{GalNAc}\alpha 1\text{-4Glc}\alpha\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl}\beta 1\text{-O-Ser}$ , que es el producto de reacción de  $\alpha\text{-GalNAcT}$  y no funciona como sustrato aceptor para la polimerización de HS o CS.<sup>277</sup> Esto llevó a pensar que la adición de GalNAc podría ser una señal de término de la elongación del núcleo GAG-proteína de estos PGs, lo cual es bastante novedoso dado que antes no había sido encontrada una estructura que terminara con  $\alpha\text{GalNAc}$  en GAGs naturalmente sintetizados.<sup>269,280</sup> Como EXTL2 posee la actividad transferasa  $\alpha 4\text{-GalNAcT}$ ,<sup>269</sup> es posible que la falta o escasez de EXTL2 incremente las cantidades de HS y CS. Esta idea ha sido demostrada utilizando células de hígado de ratón, encontrándose que EXTL2 transfiere específicamente GlcNAc hasta  $\text{Glc}\alpha\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl(2-O-fosfato)}$  pero no hasta  $\text{Glc}\alpha\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl}$ . Adicionalmente, utilizando el complejo polimerasa EXT1/EXT2 se encontró que  $\text{GlcNAc}\alpha 1\text{-4Glc}\alpha\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl(2-O-fosfato)}$  no funciona como aceptor para la reacción de polimerización, mientras que el pentasacárido no fosforilado ( $\text{GlcNAc}\alpha 1\text{-4Glc}\alpha\beta 1\text{-3Gal}\beta 1\text{-3Gal}\beta 1\text{-4Xyl}$ ), actúa como un buen aceptor.<sup>277</sup> La enzima que añade fosfato a la xilosa ayudándose de ATP ha sido previamente identificada como la xilosa quinasa (XK, del inglés Xylose Kinase) FAM20B.<sup>281</sup> Si FAM20B está expresada, el número de cadenas de GAGs aumenta pero si EXTL2 y FAM20B están expresadas, la síntesis de GAGs disminuye. Es curioso que EXTL2 sea capaz de detener la elongación de HS o CS cuando el tetrasacárido está fosforilado, pues se cree que la fosforilación temporal de xilosa mejora la formación del núcleo común de unión en GAGs.<sup>277,281</sup> Así, dentro de este laberinto de datos, se sugiere que EXTL2 podría controlar la biosíntesis de GAGs de una manera concertada con FAM20B y previamente a la desfosforilación de xilosa, pues la polimerización de HS o CS ocurre a partir del tetrasacárido no fosforilado. Contrariamente a lo que se creyó al principio respecto a EXTL2, ésta podría funcionar suprimiendo la

biosíntesis de GAGs que es mejorada por FAM20B, estableciendo un mecanismo de regulación para la biosíntesis de PGs (Figura 51\*).<sup>277</sup> Esta información se suma a un estudio en el que se indicó que a pesar de que el incremento de HS/H no tiene un impacto detectable en la fisiología normal del animal, perjudica la regeneración del hígado.<sup>282</sup> De este modo, si EXTL2 detiene la elongación de HS cuando el núcleo de unión está fosfatado o si reemplaza a EXT1/EXTL2 en ausencia de estos, permanecen aún como interrogantes y podrían indicar una regulación celular específica para la biosíntesis del HS.<sup>283</sup>

Aunque el rol de EXTL2 no está del todo desvelado, se ha visto que está relacionado con la regulación y biosíntesis de HS de una manera clave, y que ambos sustratos dadores, UDP-GalNAc y UDP-GlcNAc son necesarios para llevar a cabo esas funciones. Debido a lo anterior, esta enzima se configura como una GT-ret muy importante cuyos detalles atómicos y mecánicos son abordados computacionalmente por primera vez en esta tesis.

### 5.1.1 EXTL2 y Estructuras Cristalográficas

EXTL2 es codificada por el gen que lleva el mismo nombre y que es miembro de la familia de genes EXT. Como ya se mencionó, EXTL2 podría terminar la polimerización de GAGs transfiriendo GalNAc al tetrasacárido fosforilado en xilosa, y que una escasez de EXTL2 se asocia a la acumulación de HS o CS.<sup>277</sup> EXTL2 es la proteína más pequeña entre las proteínas EXTL (330 a.a) y además, posee una alta homología de secuencia con la región C-terminal de EXT1 y EXT2, la cual presumiblemente constituye el dominio catalítico  $\alpha$ 1,4-N-acetilglucosiltransferasa. Gracias a lo anterior, EXTL2 es un excelente modelo para investigar el mecanismo de transferencia de GlcNAc o GalNAc hasta la región de unión GAG-proteína catalizado por enzimas de la familia de genes EXT. Existen 4 estructuras cristalográficas de EXTL2: la apoenzima (código PDB 1OMX), el complejo binario EXTL2·Mn<sup>2+</sup>·UDP-GalNAc (código PDB 1OMZ), el complejo binario EXTL2·Mn<sup>2+</sup>·UDP-GlcNAc (código PDB 1ON6) y el complejo ternario EXTL2·Mn<sup>2+</sup>·UDP-[GlcA $\beta$ 1-3Gal $\beta$ 1-Onaftalenmetanol] (código PDB 1ON8), cuyo sustrato aceptor es un análogo del núcleo de unión GAG-proteína de varios PGs.<sup>70</sup> EXTL2 muestra un plegamiento GT-A y es miembro de la familia GT64; las estructuras cristalográficas revelan que el motivo DXD lo conforma la secuencia Asp151-Asp152-Asp153, siendo el último parte de la esfera de coordinación del metal Mn<sup>2+</sup>. Adicionalmente, ha sido encontrado que luego de que el sustrato aceptor se une a la enzima, ocurre un ordenamiento de los residuos que conforman un lazo que va desde el Glu275 a Glu288, el cual no es visible en ninguna de las dos estructuras que contienen al sustrato dador, 1ON6 y 1OMZ, pero sí es visible en 1ON8, que contiene UDP y un análogo del sustrato aceptor.<sup>70</sup>

EXTL2 cataliza la transferencia de GalNAc y GlcNAc hasta el extremo no reductor del análogo del sustrato aceptor, el monosacárido GlcA. Esto lo hace reteniendo la configuración del carbono anomérico y formando la unión  $\alpha$ 1-4. Como ya se sabe, dos tipos de sustituciones nucleofílicas han sido propuestas para explicar la formación del enlace glicosídico con retención de la configuración del carbono anomérico; el doble desplazamiento, que requiere de un nucleófilo bien posicionado con respecto al carbono anomérico; y la sustitución nucleofílica tipo-S<sub>N</sub>i, porque el nucleófilo se encuentra

básicamente esperando a ser activado en la misma cara por la que el grupo saliente (UDP) se aleja. Debido a lo anterior, éste es comúnmente llamado mecanismo de ataque frontal. En las estructuras 1OMZ y 1ON6 sólo la Arg293 se encuentra a una distancia  $< 4.0 \text{ \AA}$  del carbono anomérico. Esta Arg293 interacciona cercanamente con el Asp246, el cual podría actuar como nucleófilo pero está a  $> 5.0 \text{ \AA}$  del carbono anomérico. Pedersen *et. al.*<sup>70</sup> especulan que las interacciones que forma Arg293 con Asp246, Asn243 y Tyr193 podrían en alguna medida neutralizar su carga positiva, convirtiéndola incluso en una especie capaz de estabilizar la carga positiva que se desarrolla sobre el centro anomérico. Los estudios de mutagénesis sobre Arg293 y Asp246, resultaron en actividad transferasa mínima (0.2 pmol/mg de proteína/h) para Arg293Ala y actividad transferasa indetectable en el caso de Asp246Ala.<sup>70</sup> Aunque Arg293 posee un grupo guanidinio polarizable, parece inverosímil que pueda llegar a desarrollar la habilidad de estabilizar la carga que se desarrollará sobre el centro anomérico. De cualquier manera, los experimentos de mutagénesis muestran que Arg293 es importante en la catálisis y que Asp246 es clave, a pesar de encontrarse a  $5.20 \text{ \AA}$  del carbono anomérico en la estructura cristalográfica (código PDB 1OMZ).

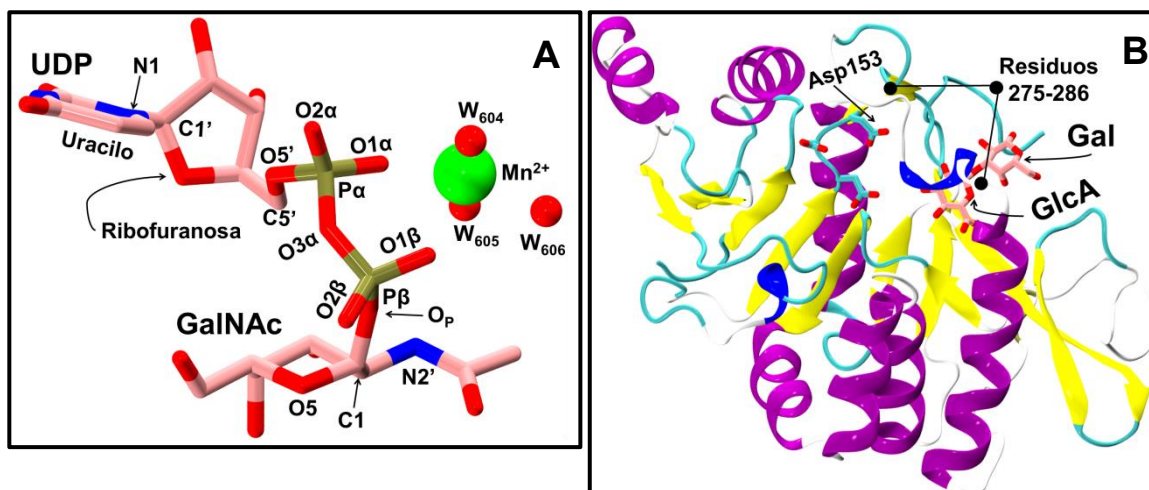
En base a esta información experimental, los objetivos de este estudio son los siguientes: caracterizar el mecanismo catalítico para la transferencia de GalNAc y GlcNAc (capítulo 6) e identificar los roles catalíticos de estos dos residuos ubicados en la cara  $\beta$  del azúcar, Asp246 y Arg293.

## 5.2 Modelos y Metodología

### 5.2.1 Construcción del Modelo Inicial para la Transferencia de GalNAc

El modelo ternario inicial, EXTL2·Mg<sup>2+</sup>·UDP-GalNAc·GlcA $\beta$ 1-3Gal, se construyó en base a dos estructuras cristalográficas de EXTL2. La estructura con código pdb 1OMZ (EXTL2·Mn<sup>2+</sup>·UDP-GalNAc), en la que no están resueltas las coordenadas de un grupo de residuos que abarcan desde el 275 al 286. Debido a esto, sólo se seleccionaron las coordenadas del sustrato dador UDP-GalNAc; las del metal Mn<sup>2+</sup>, que es coordinado por los fosfatos  $\alpha$  y  $\beta$  de UDP; y las tres moléculas de agua que también forman parte de la primera esfera de coordinación del metal. El residuo faltante en esta esfera pertenece al motivo DXD típicamente presente en las GTs con plegamiento GT-A. En EXTL2 el motivo DXD corresponde a la secuencia Asp151-Asp152-Asp153, siendo el Asp153 quien coordina al Mn<sup>2+</sup>. Las coordenadas del motivo, así como las del resto de la proteína y el sustrato aceptor análogo GlcA $\beta$ 1-3Gal, fueron tomadas del complejo ternario EXTL2·Mn<sup>2+</sup>·UDP·GlcA $\beta$ 1-3Gal $\beta$ 1-Onaftalenmetanol (código PDB 1ON8).<sup>70</sup>





**Figura 52.** Estructuras que forman parte de A: el complejo binario EXTL2·Mn<sup>2+</sup>·UDP-GalNAc (código PDB 1OMZ) del cual se toman las coordenadas del sustrato dador UDP-GalNAc, el cofactor metálico y tres moléculas de agua que forman parte de la primera esfera de coordinación y B: el complejo ternario EXTL2·Mn<sup>2+</sup>·UDP-GlcAβ1-3Galβ1-Onaftalenmetanol, del cual se usan las coordenadas de la proteína y el sustrato aceptor. En ambos casos se incluyen las etiquetas de los sustratos que participan de la reacción química y algunos átomos o residuos de interés.

La combinación de las estructuras 1OMZ y 1ON8 dejan al O4<sub>GlcA</sub> ubicado a 2.61 Å del C1<sub>GalNAc</sub> y al oxígeno del grupo N-acetilo del azúcar transferible (NAc), a tan sólo 1.65 y 2.09 Å de O4<sub>GlcA</sub> y O3<sub>GlcA</sub>, respectivamente. Aunque estas distancias son cortas y podrían generar pequeñas tensiones estéricas, se dispone de una primera aproximación del complejo ternario que será minimizado y equilibrado a través de simulaciones MM y QM/MM. Por otra parte, el Mn<sup>2+</sup> que se encuentra en la estructura cristalográfica se modelizó como Mg<sup>2+</sup> en nuestro modelo, una aproximación computacional más eficiente que ha sido validada en otros trabajos de nuestro grupo y en estudios de sistemas relacionados.<sup>125,268,284</sup> Después de reemplazar el catión metálico se añaden hidrógenos y se asignan los estados de protonación de los residuos ionizables a través del servidor PDB2PQR versión 2.0.0,<sup>225,226</sup> y el programa PROPKA3.1,<sup>222-224</sup> respectivamente. El siguiente paso consiste en asignar la topología y los parámetros adecuados al sistema y para llevar a cabo esto, se usa el campo de fuerzas CHARMM. La proteína, Mg<sup>2+</sup>, moléculas de agua y UDP se describen con la versión CHARMM22<sup>141,227,228</sup> mientras que la del enlace glicosídico y los monosacáridos están en la versión CHARMM36;<sup>229-231</sup> en lo relativo a la unión nucleótido-azúcar, se utilizaron los parámetros de anteriores trabajos del grupo. A continuación, se utiliza el *paquete solvatar* (o *solvate package* en inglés) del programa VMD<sup>240</sup> para solvatar el modelo en una caja de aguas de dimensiones 71·117·84 Å<sup>3</sup>. También con VMD, se añaden 3 cationes Na<sup>+</sup> para neutralizar la carga del sistema, el cual queda formando por un total de 50.528 átomos.

## 5.2.2 Minimizaciones y Dinámica Molecular

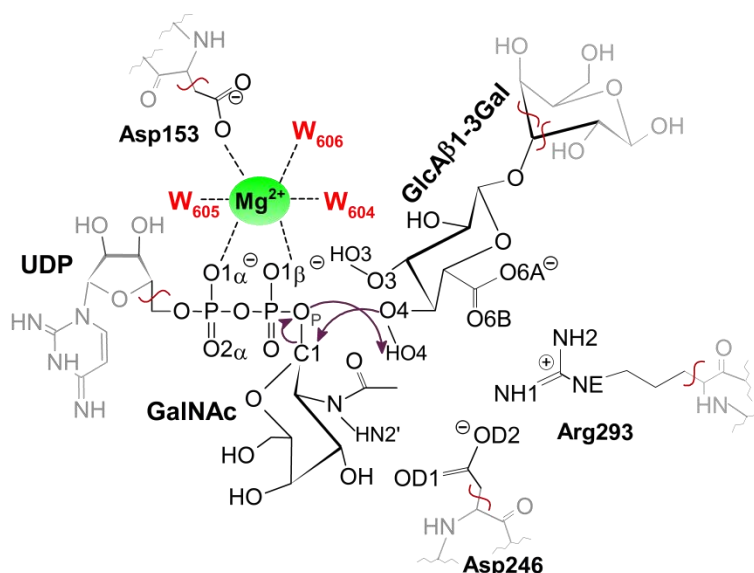
Una vez asignados los parámetros y topologías del modelo ternario, éste fue minimizado permitiendo que se optimizaran las aguas mientras la proteína y el resto del sistema se mantienen fijos. Después se minimizan la proteína junto a los cationes continuando con el UDP-GalNAc y finalmente el disacárido GlcA $\beta$ 1-3Gal. Este proceso gradual previene artefactos que puedan producirse a causa de una estructura inicial inestable. La MD del sistema se inicia con las moléculas de agua moviéndose libremente durante 400 ps, mientras el resto del sistema está fijo. Luego, se realizan simulaciones sucesivas en las que los movimientos de los distintos residuos que forman el sistema son controlados a través de restricciones armónicas. Estas restricciones son reducidas gradualmente al pasar de una simulación a otra, hasta que se quitan por completo, permitiendo la evolución libre del sistema. Este procedimiento se aplica a las cadenas laterales y al segmento base de la proteína, al metal junto a su primera esfera de coordinación y a los sustratos, UDP-GalNAc y GlcA $\beta$ 1-3Gal. En los últimos 1400 ps de equilibración, de un total de 1800 ps, el RMSD del segmento base de la proteína no superó un valor de 1.45 Å y se llevaron a cabo 5 ns más de MD. El objetivo principal de la MD es relajar la aproximación inicial de complejo ternario, eliminando tensiones estéricas que derivan del procedimiento de construcción de este modelo. Debido a esto, consideramos que la duración de la MD es suficiente para obtener un complejo ternario estable.

Las simulaciones fueron realizadas con el programa NAMD2.9,<sup>238,239</sup> estableciéndose condiciones periódicas de contorno y una temperatura y presión de 300 K y 1 bar, respectivamente. Estas variables se mantuvieron constantes utilizando dinámica de Langevin y los métodos de Nosé-Hoover<sup>200</sup> y el pistón de Langevin,<sup>199</sup> respectivamente. El algoritmo SHAKE<sup>193</sup> fue usado para mantener fijas las distancias O-H de las aguas y el paso de integración de las ecuaciones de movimiento es de 2 fs. En base a un minucioso análisis de las estructuras exploradas en la MD, se seleccionaron varios complejos de Michaelis que fueron usados como puntos iniciales para las primeras exploraciones del mecanismo de reacción mediante métodos QMMM.

Las barreras y energías de reacción que se obtienen de este conjunto de complejos seleccionados desde la MD son bastante altas, mayores que 30 y 20 kcal/mol, respectivamente, independiente de si se usa el método QM(BP86/SVP)/MM(CHARMM) o QM(SCC-DFTB)/MM(CHARMM). Esto puede estar indicando que las interacciones sustratos-enzima o sustrato-sustrato que han sido modelizadas hasta el momento no son las más adecuadas para la catálisis pues no estabilizan suficientemente el estado de transición ni el producto o, visto de otra forma, estabilizan demasiado el reactivo. Tal como se comentó en el capítulo de GpgS, una estrategia que ayuda a refinar el modelo cuando se dan estas circunstancias es llevar a cabo una MD de los productos obtenidos en esta primera exploración del mecanismo de reacción con el fin de que el sistema se relaje en este estado. El procedimiento contempla simulaciones de MD de 60 ps al nivel QM(SCC-DFTB)/MM(CHARMM), empleando los complejos enzima-producto (EXTL2·Mg<sup>2+</sup>·UDP·GalNAc $\alpha$ 1-4GlcA $\beta$ 1-3Gal) obtenidos previamente. En base a estas simulaciones, se seleccionaron varias estructuras que se usaron como puntos iniciales para simular la reacción química inversa, consiguiendo un nuevo conjunto de complejos de



En la región QM ocurre la reacción química. En ella se incluyen el sustrato dador y el monosacárido GlcA junto a los átomos de la unión glicosídica con Gal; además se incluye el  $Mg^{2+}$  y su primera esfera de coordinación que la conforman las aguas  $W_{604}$ ,  $W_{605}$  y  $W_{606}$ , el Asp153 y los fosfatos  $\alpha$  y  $\beta$ . Adicionalmente, se añaden a la región QM dos residuos que forman parte de la cara  $\beta$  de GalNAc, el Asp246 y la Arg293.



**Figura 54.** Esquema parcial del centro activo en el que se indica la partición QM para la enzima EXTL2. Las moléculas que forman parte del subsistema QM se dibujan en líneas negras mientras que las que forman parte del subsistema MM, se colorean en un tono gris. Se indican también en líneas curvadas rojas, el enlace que define la frontera entre región QM y MM. Por último, se emplean flechas moradas para señalar los enlaces que se formarán y romperán durante la reacción de transferencia.

La carga neta de la región QM es -2 y está formada por 104 átomos. Las interacciones electrostáticas entre ambas regiones se abordan con el esquema comúnmente conocido como integración electrostática (en inglés se utiliza *electronic* o *electrostatic embedding* indistintamente).<sup>243</sup> No se establece una distancia de corte para el cálculo de las interacciones no enlazantes a nivel MM o QM-MM, y el tratamiento de la interfaz QM/MM se aborda con 6 *links atoms*<sup>241,242</sup> y el método de redistribución de carga (o *charge shift* en inglés).<sup>179,243</sup> La descripción electrónica de la región QM emplea los funcionales BP86<sup>152,154,244-246</sup> y PBE<sup>163</sup> junto a los conjuntos base SVP<sup>247</sup> o TZVP,<sup>248</sup> nivel de teoría que se denota como QM(BP86/SVP), QM(BP86/TZVP), etc. La energía QMMM de las estructuras optimizadas es recalculada al nivel QM(M05-2X<sup>166</sup>/TZVP). En el caso de las dinámicas moleculares a nivel QMMM, la descripción electrónica utilizó el método semiempírico SCC-DFTB,<sup>171-173</sup> el cual en estos sistemas reduce la resolución de las integrales a costa de sobreestimar las barreras de la reacción, pero incluso así es útil en términos descriptivos. Por último, la región MM contiene todos aquellos átomos que se excluyeron de la región QM y queda descrita por el campo de fuerzas CHARMM detallado anteriormente.

### 5.2.3.1 Mecanismo de Ataque Frontal

Para conducir el sistema de reactivos a productos, se realizaron optimizaciones restringidas por una coordenada de reacción (RC por su sigla en inglés), definida como  $RC_A = [d(O_P-C1_{GalNAc}) - d(O4_{GlcA}-C1_{GalNAc}) - d(HO4_{GlcA}-O_P)]$  y sobre la cual se aplicó una constante de fuerza igual a  $527 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ . La combinación de estas distancias se varía en pasos de 0.1 Å. Esta coordenada es la que se utilizó en las exploraciones iniciales a nivel QM(BP86/SVP)/CHARMM a partir de complejos obtenidos de la MD que, como se ha comentado, dieron barreras de energía potencial muy elevadas. Una vez llevado a cabo el procedimiento de refinamiento descrito y seleccionado un complejo representativo (ver resultados), la misma coordenada se utilizó y se probaron varios métodos QM distintos (BP86 y PBE, con las bases SVP y TZVP). A partir de los perfiles de energía potencial se caracterizaron los correspondientes intermedios par iónico (IPs), cuyas geometrías y energías se presentan en la tabla siguiente junto con los valores para los reactivos y los productos. El objetivo aquí es solo presentar la comparación de métodos y comentar algunos aspectos que permiten entender la motivación detrás de la utilización de las coordenadas de reacción que se describirán a continuación.

**Tabla 12.** Estructuras optimizadas de Reactivo (R), intermedio par iónico (IP) y Producto (P) calculadas con los métodos QM BP86/SVP, PBE/SVP, BP86/TZVP y PBE/TZVP. Las distancias se dan en Å y en cada fila los valores se dan en este orden: BP86/SVP(PBE/SVP)/BP86/TZVP(PBE/TZVP). Las energías se han recalculado al nivel QM(M05-2X/TZVP//BP86/TZVP) y (M05-2X/TZVP//PBE/TZVP). Los valores de energía (kcal/mol) se dan en el mismo orden en el que se dan las distancias, esto es, BP86/SVP(PBE/SVP)/BP86/TZVP(PBE/TZVP) y a continuación, M05-2X/TZVP//BP86/TZVP(M05-2X/TZVP//PBE/TZVP).

|  | Reactivo                       | IP                             | Producto                       |
|--|--------------------------------|--------------------------------|--------------------------------|
| <b><i>d(C1<sub>GalNAc</sub>-O<sub>P</sub>)</i></b>           | <b>1.47 (1.47)/1.50 (1.49)</b> | <b>3.28 (3.26)/3.53 (3.52)</b> | <b>3.56 (3.51)/3.62 (3.60)</b> |
| <i>d(C1<sub>GalNAc</sub>-O4<sub>GlcA</sub>)</i>              | 2.97 (2.98)/3.04 (3.05)        | 2.39 (2.38)/2.43 (2.43)        | 1.50 (1.51)/1.51 (1.51)        |
| <i>d(HO4<sub>GlcA</sub>-O<sub>P</sub>)</i>                   | 3.33 (3.33)/3.27 (3.29)        | 1.52 (1.51)/1.57 (1.57)        | 1.03 (1.04)/1.02 (1.02)        |
| <b><i>d(OD1<sub>Asp246</sub>-C1<sub>GalNAc</sub>)</i></b>    | <b>5.47 (5.45)/5.43 (5.41)</b> | <b>4.29 (4.22)/3.16 (3.15)</b> | <b>4.76 (4.60)/4.41 (4.44)</b> |
| <i>d(HO4<sub>GlcA</sub>-O4<sub>GlcA</sub>)</i>               | 1.01 (1.01)/1.01 (1.00)        | 1.03 (1.03)/1.02 (1.03)        | 1.54 (1.51)/1.60 (1.60)        |
| <i>d(WAT604-O2<math>\beta</math>)</i>                        | 1.76 (1.77)/1.81 (1.82)        | 1.58 (1.59)/1.60 (1.60)        | 1.61 (1.61)/1.66 (1.69)        |
| <i>d(WAT604-O2<math>\alpha</math>)</i>                       | 1.92 (1.92)/2.19 (2.20)        | 2.71 (2.45)/4.05 (4.00)        | 4.00 (3.82)/4.31 (4.27)        |
| <i>d(HN2'<sub>GalNAc</sub>-O3<sub>GlcA</sub>)</i>            | 2.39 (2.40)/2.50 (2.51)        | 1.75 (1.76)/1.77 (1.78)        | 1.91 (1.91)/1.82 (1.95)        |
| <i>d(O1<math>\beta</math>-HO3<sub>GlcA</sub>)</i>            | 1.66 (1.66)/1.72 (1.73)        | 1.55 (1.55)/1.60 (1.61)        | 1.68 (1.69)/1.81 (1.86)        |
| <i>d(NE<sub>Arg293</sub>-O6<sub>B<sub>GlcA</sub></sub>)</i>  | 3.23 (3.22)/3.20 (3.20)        | 3.05 (2.99)/3.19 (3.11)        | 3.13 (3.14)/2.95 (3.13)        |
| <i>d(NH1<sub>Arg293</sub>-O6<sub>B<sub>GlcA</sub></sub>)</i> | 3.42 (3.41)/3.45 (3.44)        | 3.94 (3.74)/4.18 (4.05)        | 2.91 (2.85)/3.84 (3.01)        |
| <i>d(NH2<sub>Arg293</sub>-O5<sub>GalNAc</sub>)</i>           | 2.92 (2.92)/2.97 (2.96)        | 2.85 (2.86)/2.89 (2.87)        | 3.06 (3.08)/2.99 (3.04)        |
| <i>d(C1<sub>GalNAc</sub>-O5<sub>GalNAc</sub>)</i>            | 1.39 (1.39)/1.39 (1.38)        | 1.28 (1.28)/1.28 (1.28)        | 1.39 (1.38)/1.39 (1.39)        |

|   |                         |                         |                          |
|---|-------------------------|-------------------------|--------------------------|
| $d(NH2_{Arg293}-C1_{GalNAc})$             | 3.45 (3.45)/3.48 (3.48) | 3.59 (3.61)/3.61 (3.58) | 4.03 (4.08)/3.99 (4.05)  |
| <b>Método QM</b>                          |                         |                         |                          |
| BP86/SVP(PBE)/BP86/TZVP(PBE) <sup>1</sup> |                         | 14.4(14.9)/4.73(5.12)   | -3.3(-2.93)/-1.10(-4.89) |
| M05-2X/TZVP <sup>2</sup>                  |                         | 12.00(11.07)            | 0.11(-5.62)              |

<sup>1</sup> El término entre paréntesis sólo indica el cambio de funcional pues se trata del mismo conjunto base SVP y TZVP, respectivamente.

<sup>2</sup> Cálculos de energía a nivel QM(M05-2X/TZVP) sobre los puntos optimizados al nivel BP86/TZVP(PBE/TZVP).

En general, al utilizar la misma base se obtienen resultados estructurales muy parecidos a nivel BP86 y PBE (ambos funcionales basados en la aproximación GGA). Por otra parte, al comparar los conjuntos de base encontramos algunas diferencias, principalmente en el intermedio IP. A nivel de energía, con la base TZVP y los funcionales BP86<sup>152,154,244-246</sup> o PBE,<sup>163</sup> la energía es mucho más baja que con la base SVP, aunque al hacer el cálculo puntual de energía a nivel M05-2X/TZVP la energía del IP vuelve a valores más elevados. A nivel de geometrías, la diferencia principal se observa en la distancia  $OD1_{Asp246}-C1_{GalNAc}$  en el IP, cuyo valor es de 3.2 Å al utilizar el conjunto de bases TZVP y 4.2-4.3 Å cuando se utiliza el conjunto de bases SVP. Esta es una diferencia > 1 Å y, evidentemente, es importante pues se trata del residuo nucleofílico. En realidad, con la base TZVP la búsqueda del IP se realizó a partir de dos puntos distintos del perfil de energía potencial, y esto llevó a dos IPs distintos: el aquí descrito (que más adelante denominaremos IP<sub>B</sub>) y otro muy parecido al calculado con la base SVP (IP<sub>A</sub>). Con la base SVP solo se consiguió determinar un IP. Dadas estas diferencias, lo más adecuado sería utilizar el conjunto de bases más flexible y grande para la descripción de la densidad electrónica y es por ello que el estudio se continuó con el conjunto de bases TZVP. Además se escogió el funcional BP86 pues es el que nuestro grupo ha utilizado para el estudio de estas enzimas y según nuestros cálculos trabaja de la misma manera que el funcional PBE en lo que hace referencia a las estructuras.

El IP<sub>B</sub> obtenido al nivel BP86/TZVP es un punto estacionario que se aleja de la curva de energía potencial calculada con la RC<sub>A</sub>. Para comprobar su conectividad con el reactivo y el producto, el IP<sub>B</sub> se usó como estructura de partida para iniciar nuevas exploraciones a través de coordenadas de reacción que conduzcan hacia el reactivo y el producto. En el primer caso, se definió la siguiente coordenada de reacción,  $RC_B = [d(O4_{GlcA}-C1_{GalNAc}) - d(O_P-C1_{GalNAc}) - d(HO4_{GlcA}-O6B_{GlcA})]$ , mientras que en el segundo caso se utilizó la RC<sub>A</sub>. Cabe mencionar que aunque se trata de la misma coordenada de reacción, en lo sucesivo utilizaremos RC<sub>AB</sub> y no RC<sub>A</sub>, para resaltar que la exploración se inicia desde una configuración diferente de la enzima, el IP<sub>B</sub>. Una selección del nuevo conjunto de estructuras que se obtienen de estos perfiles de energía potencial se utilizan para buscar estados de transición para el proceso de disociación UDP-GalNAc y el ataque de GlcA o formación del enlace glicosídico GalNAcα1-4Gal.

Los programas utilizados para los cálculos QMMM son los mismos que los descritos en el capítulo anterior para GpgS (apartado 4.2.3).

### 5.2.3.2 Mecanismo de Ataque del Asp246

Dado el acercamiento de Asp246 a  $C1_{GalNAc}$  observado en el IP del ataque frontal (aunque no en el reactivo reoptimizado), se decidió explorar la posibilidad de que este residuo pueda efectivamente atacar al carbono anomérico. Para llevar a cabo esta exploración se utilizó una coordenada de reacción triple,  $RC_{DA} = [d(O_P-C1_{GalNAc}) - d(OD1_{Asp246}-C1_{GalNAc}) - d(HO4_{GlcA}-O_P)]$ . La distancia  $d(HO4_{GlcA}-O_P)$  es incluida en la coordenada de reacción para guiar la reorientación de  $HO4_{GlcA}$  desde  $O6B_{GlcA}$  hasta  $O_P$ , lo cual se consigue sin que ocurra la abstracción del protón por parte de UDP, que es precisamente lo que se busca. En este caso, también se buscaron intermedios (IP) y estados de transición (TS) usando varias estructuras asociadas al proceso de disociación UDP-GalNAc y al ataque de Asp246; las estructuras obtenidas se usaron para calcular los modos normales de vibración. Es importante mencionar que el CGE optimizado en esta exploración, resultaba ser estructural y energéticamente distante de la curva de energía potencial. Por este motivo, se calculó un nuevo camino que conectara este CGE con el IP a través de la coordenada de reacción  $RC_E = [d(OD1_{Asp246}-C1_{GalNAc}) - d(O_P-C1_{GalNAc})]$ . A partir de esta última exploración, fue posible caracterizar el  $TS_{CGE}$  que se muestra en la sección de Resultados y Discusión.

### 5.2.4 Cálculo de Contribuciones Electroestáticas, Cargas Naturales e Interacción de Orbitales Naturales

Con el objetivo de obtener una estimación de las interacciones electrostáticas de residuos del sitio activo, o de fragmentos moleculares que participan directamente en la reacción, se lleva a cabo un análisis electrostático al nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM análogo al descrito en el apartado 4.2.8 de GpgS, pero con los puntos estacionarios obtenidos aquí al nivel QM(BP86/TZVP)/MM(CHARMM). Estos ya han sido caracterizados como mínimos (R, IP y P), puntos de silla (TS2) o especies OC, que son especies que han sido optimizadas con el algoritmo de búsqueda de TS pero no poseen una frecuencia imaginaria.

El cálculo de cargas naturales (NPA) y el análisis de orbitales naturales de enlace (NBO) se realizó también análogamente a lo descrito en el apartado 4.2.9 de GpgS.

### 5.2.5 Cálculo de Energía Libre para el Mecanismo de Ataque Frontal

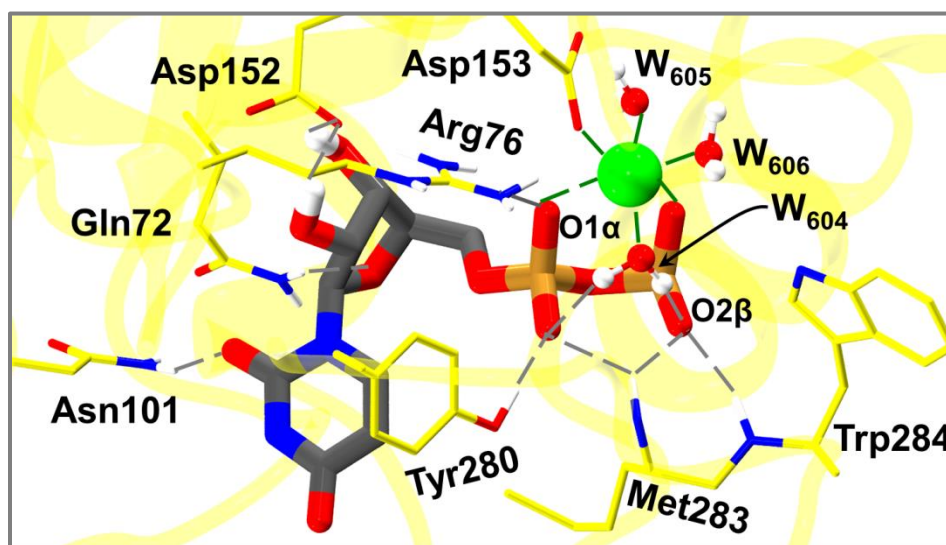
Con el fin de tener una estimación de los efectos térmicos y entrópicos en la barrera de energía de la transferencia, se efectuaron cálculos de energía libre basados en simulaciones de Umbrella Sampling (US)<sup>202-204</sup> para el mecanismo de ataque frontal. Estos cálculos se llevaron a cabo utilizando el nivel de energía QM(SCC-DFTB)/MM(CHARMM) y una zona QM más reducida que la descrita anteriormente (Asp246 y Arg293 no están en la zona QM). La primera etapa consiste en simular el camino de reacción a este nivel de teoría, para lo que se emplea la misma coordenada de reacción triple ya definida ( $RC_A$ ), con pasos de 0.1 Å. A partir del perfil de energía potencial obtenido se realizan las simulaciones de Umbrella Sampling. Cada una de las estructuras obtenidas a lo largo del camino de reacción es un punto de partida para la correspondiente ventana, que consta de una simulación de MD cuya evolución temporal estará restringida a la  $RC_A$ , a través de una constante de fuerza de  $87 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ . Las simulaciones se realizaron bajo un colectivo NVT, estableciendo una temperatura de 300 K que fue mantenida con los termostatos de Nosé-Hoover<sup>196,197</sup> y Berendsen,<sup>198</sup> las distancias O-H de todas las moléculas de aguas se mantienen fijas en cada paso a través del algoritmo SHAKE.<sup>193</sup> La simulación consistió de 10 ps de equilibración del sistema, seguida de 60 ps de muestreo conformacional en cada una de las 57 ventanas. A continuación, se prueban dos métodos para calcular la energía libre asociada a la reacción de transferencia. Uno de ellos es el método *Umbrella Integration*<sup>206,207</sup> (UI) y el otro, *Weighted Histogram Analysis Method*<sup>205</sup> (WHAM). Los resultados que se mostrarán más adelante corresponden a los obtenidos con UI. Para las simulaciones de US se utiliza el módulo *Dynamics* implementado en el programa ChemShell.<sup>181,250</sup>

Todos los programas utilizados para llevar a cabo las simulaciones y los cálculos puntuales ya han sido indicados en el apartado 4.2.3 de GpgS.



### 5.3 Resultados y Discusión

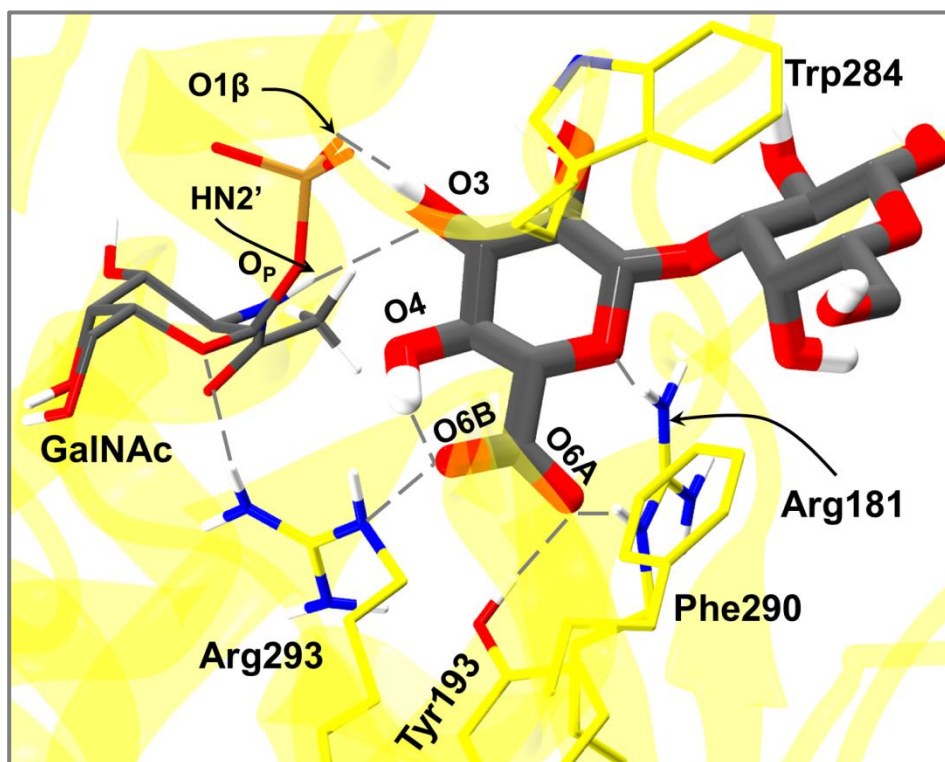
Esta sección se inicia con la descripción de las interacciones sustratos-enzima que hemos modelado combinando metodologías MM y QMMM, las cuales nos permiten obtener un complejo de Michaelis eficiente catalíticamente (el que aquí se muestra corresponde al complejo finalmente seleccionado para realizar los estudios completos de reactividad). Primero se comentarán los residuos que interaccionan con el sustrato dador UDP-GalNac, mostrando primero los residuos asociados a UDP y luego a GalNac.



**Figura 55.** Residuos que interaccionan directamente con Uridina-5'-difosfato (UDP) en el sitio activo. Los residuos enzimáticos se dibujan con carbonos amarillos mientras que los carbonos de los sustratos se dibujan en gris. La Uridina (ribofuranosa $\beta$ -1(N)uracilo) establece varios enlaces de hidrógeno con diferentes residuos entre los que cuentan: Asn101, Gln72 y Asp152. El grupo difosfato interacciona con Arg76, Tyr280, Met283, Trp284 y el metal  $Mg^{2+}$ . La primera esfera de coordinación del  $Mg^{2+}$  la conforman el Asp153, los oxígenos O1 $\alpha$  y O1 $\beta$  y tres moléculas de agua:  $W_{605}$ ,  $W_{606}$  y  $W_{604}$ . La molécula de agua  $W_{604}$  también forma dos interacciones con el grupo fosfato. Todas las interacciones se indican a través de líneas discontinuas grises, con excepción de aquellas que involucran a la primera esfera de coordinación del  $Mg^{2+}$ , las cuales se señalan en verde.

El grupo UDP del sustrato dador forma varias interacciones con residuos neutros como Asn101, Gln72, Tyr280, Met283 y Trp284, así como con dos residuos cargados, la Arg76 y el Asp152. Éste último es parte del motivo conservado DXD (aquí Asp151-Asp152-Asp153), que se encuentra casi en las GTs de plegamiento GT-A. El Asp153 forma parte de la primera esfera de coordinación del  $Mg^{2+}$ , junto al grupo difosfato (O1 $\alpha$  y O1 $\beta$ ) y tres moléculas de agua ( $W_{606}$ ,  $W_{605}$  y  $W_{604}$ ). Por último, la  $W_{604}$  orienta sus hidrógenos hacia el grupo difosfato estableciendo dos cercanas interacciones con O2 $\alpha$  y O2 $\beta$ . En este punto es importante señalar que la interacción  $W_{604}$ -O2 $\beta$ , tiene un notable efecto sobre las barreras energéticas pues todas las exploraciones del mecanismo de ataque frontal que se iniciaron desde estructuras de reactivo que no la tienen (complejos de Michaelis obtenidos desde la MD), resultaron en barreras energéticas  $> 30$  kcal/mol. Cabe decir, también, que en las simulaciones de MD realizadas con los potenciales del campo de fuerzas, esta interacción acostumbra a no estar presente, a diferencia de lo que ocurre a nivel QM(DFT)/MM.

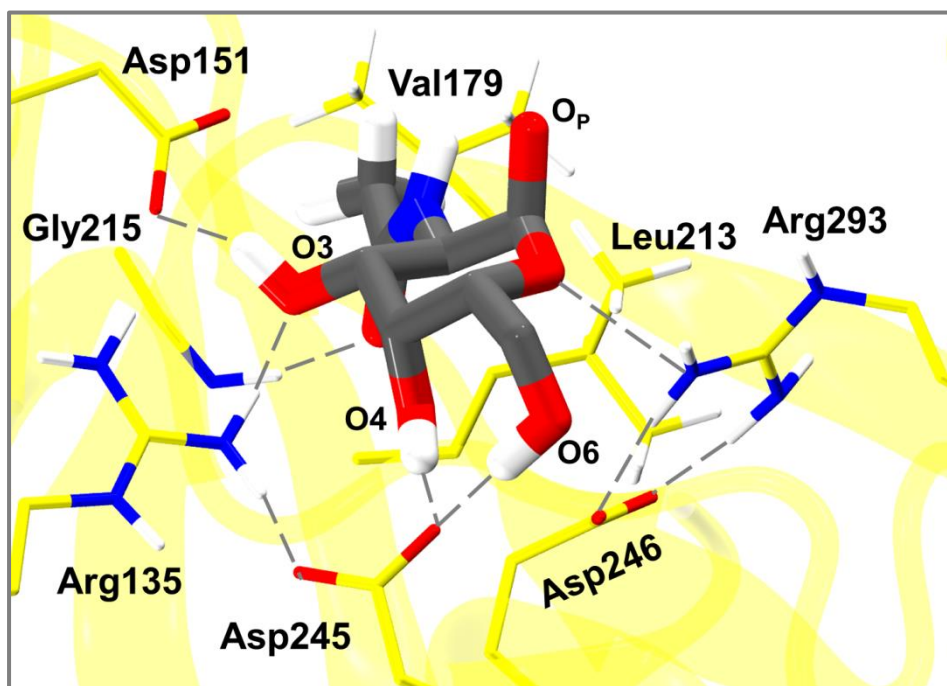
En el caso de sustrato aceptor GlcA $\beta$ 1-3Gal, se encuentra nuevamente al Trp284, pero esta vez formando una interacción de tipo CH- $\pi$  con GlcA. También se observan otros residuos aromáticos en la cavidad; la Phe290, orientando su cadena lateral hacia el sustrato aceptor; y la Tyr193, que forma un enlace de hidrógeno con el grupo carboxilato de GlcA (COO<sup>-</sup><sub>GlcA</sub>), a través del oxígeno O6A.



**Figura 56.** Se representan el sustrato aceptor GlcA $\beta$ 1-3Gal y parte del sustrato dador UDP-GalNAc (carbonos grises), y todos los residuos enzimáticos que forman parte de la cavidad de unión (carbonos amarillos). Las interacciones se señalan con líneas discontinuas grises.

Además de estos tres residuos aromáticos, también se observan dos argininas en la cavidad. La Arg181 establece dos interacciones con GlcA y la Arg293 se orienta hacia el COO<sup>-</sup><sub>GlcA</sub>. La Arg293 también aparece a distancia de interacción del O5<sub>GalNAc</sub> (3.03 Å) y se posiciona a 3.63 Å del carbono anomérico. El sustrato aceptor también establece importantes interacciones con UDP-GalNAc, por ejemplo; HO3<sub>GlcA</sub>-O1 $\beta$ , que posiblemente facilita la salida del grupo UDP; O3<sub>GlcA</sub>-HN2'<sub>GalNAc</sub>, una interacción que podría ayudar a posicionar GlcA para el ataque nucleofílico. Es importante destacar que en la mayoría de los complejos de Michaelis obtenidos desde las simulaciones de DM, esta interacción no aparecía y en lugar de ella se encontraba el oxígeno del grupo NAc formando una interacción con HO4<sub>GalNAc</sub> y HN2', interaccionando cercanamente con el Asp151 (ver la siguiente Figura para localizar el Asp151). Modelar las interacciones inter-sustratos en EXTL2, ha sido una tarea difícil y laboriosa para la que ha sido necesario combinar simulaciones MM (MD) y QMMM (exploraciones directas del mecanismo de ataque frontal

(R → P), MD QM/MM del producto, exploraciones inversas (P → R)). A nivel QM(M05-2X/TZVP//BP86/TZVP)//MM(CHARMM), la interacción (HO4-O6B)<sub>GlcA</sub>, resulta ser 4 kcal/mol más estable que HO4<sub>GlcA</sub>-O<sub>P</sub>. Sin embargo, esta última interacción se establece tempranamente en el camino de reacción, justamente antes de iniciarse la disociación UDP-GalNAc indicando, como ya se ha mencionado antes, que facilita la salida del grupo UDP y promueve el ataque del nucleófilo, puesto que O<sub>P</sub> va a abstraer el protón de O4<sub>GlcA</sub>.



**Figura 57.** Azúcar del sustrato dador, GalNAc (carbonos grises), y todos los residuos (carbonos amarillos) que forman parte del sitio de unión. Las interacciones se señalan a través de líneas discontinuas grises.

GalNAc establece variadas interacciones con diferentes tipos de residuos. El grupo NAc interacciona con residuos voluminosos como Val179 y Leu213, y también forma un enlace de hidrógeno con la Gly215. Asp151, que es parte del motivo conservado DXD, interacciona con HO3<sub>GalNAc</sub>. Además, se observa a la Arg135 interaccionando con O3<sub>GalNAc</sub> y al Asp245 formando dos enlaces de hidrógeno con GalNAc a través de HO6 y HO4. Como ya se mencionó en la Introducción (sección 1.3.3.2), la cadena lateral carboxilato de un residuo Glu/Asp en posición de interaccionar con los hidroxilos del azúcar, y con una cadena lateral con carga positiva (Arg/Lys), es la única característica conservada en la cara β del azúcar.(review2008). En EXTL2 la cadena lateral carboxilato la proporciona el Asp245 que interacciona con la Arg135, pareja Arg135-Asp245; en GpgS correspondía a la pareja Lys114-Glu232. Por último, se tiene a la Arg293 y al Asp246. Estos dos residuos interaccionan cercanamente y aunque en principio, Asp246 podría ser un buen nucleófilo, se encuentra a 5.43 Å del carbono anomérico en nuestro reactivo.

En uno de los complejos binarios (1OMZ), Asp246 está ubicado a 5.20 Å del C1<sub>GalNAc</sub> y ninguno de sus oxígenos apunta hacia el carbono anomérico, por lo que se anticipa que

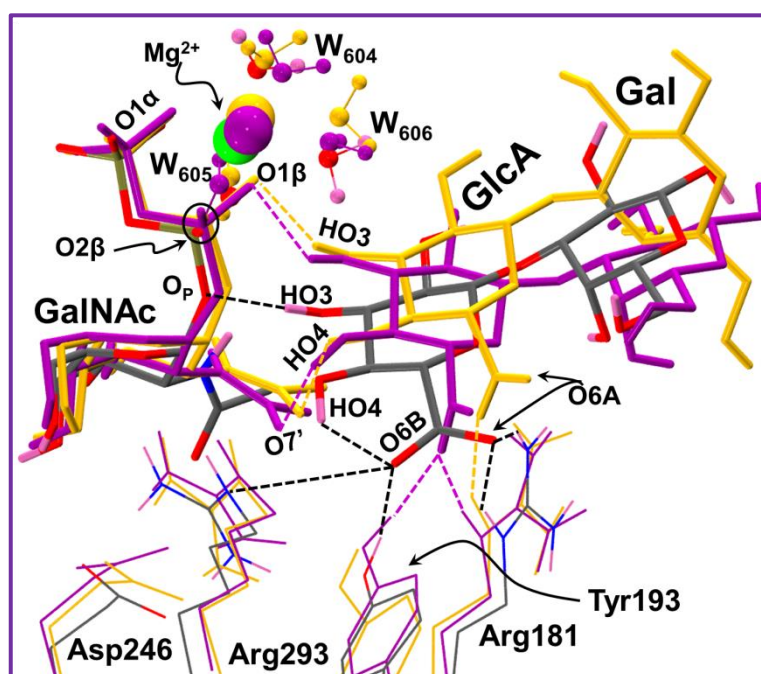
Asp246 sólo podría formar el CGE si sufriera un cambio conformacional. A lo largo de la simulación de MD esta distancia se alarga incluso más, apareciendo a un promedio de  $5.91 \pm 0.26$  Å; en los reactivos minimizados a nivel QM/MM, esta distancia toma un valor de  $5.00 \pm 0.70$  Å. Adicionalmente, en nuestro modelo el Asp246 orienta sus dos oxígenos hacia el grupo guanidinio de Arg293, lo cual probablemente reduce su habilidad nucleofílica. Esto es diferente a lo que se observa en las 3 estructuras cristalográficas que contienen sustratos (1OMZ, 1ON6 y 1ON8), en las que el grupo carboxilato de Asp246 y el grupo guanidinio de Arg293 se encuentran en planos prácticamente perpendiculares. Por la otra cara del azúcar (cara  $\alpha$ ), el  $O4_{\text{GlcA}}$  parece estar listo para atacar y basados en estas características estructurales, el mecanismo de ataque frontal parece ser el más probable.

Como se ha comentado al describir los métodos en este capítulo, durante la simulación del mecanismo de reacción a nivel QM(BP86/TZVP) se encontró que en el  $IP_B$  el Asp246 sufre un ligero cambio conformacional que resulta en una distancia  $OD1_{\text{Asp246}}-C1_{\text{GalNAc}}$  igual a 3.16 Å. La disminución de esta distancia hizo resurgir la posibilidad de la formación del CGE, de modo que el ataque nucleofílico de Asp246 también ha sido investigado. Cabe destacar que el acortamiento de esta distancia ocurre específicamente en el IP altamente disociado, es decir, una vez que ha ocurrido el rompimiento UDP-GalNAc y se ha formado la especie oxocarbenio (OC). Basados en esta información, creemos que el ligero cambio conformacional Asp246 podría facilitar la salida del grupo saliente UDP en dos maneras: estabilizando la especie UDP----OC a través de su carga negativa, y empujando a UDP lejos del carbono anomérico por la cara opuesta a la que UDP ocupa, la cara  $\beta$ . Cuando Sinnott & Jencks estudiaron la solvolisis de derivados de Glucosa, intentaron dar explicación a los resultados estereoquímicos sugiriendo por primera vez esta idea, que si se forma un estado de transición lo suficientemente “abierto” o disociado, el nucleófilo podría presionar o empujar sobre el carbono anomérico por la cara opuesta o por la misma cara del grupo saliente.<sup>117</sup> Teniendo esta idea en mente, es posible que el Asp246 facilite el mecanismo altamente disociado ( $UDP^{\delta-}----OC^{\delta+}$ ), ejerciendo un empujón nucleofílico sobre el carbono anomérico ( $C1_{\text{GalNAc}}$ ) que ayuda a expulsar al UDP, creando de esta manera, el espacio para la entrada del sustrato aceptor.

Finalmente, antes de entrar en la descripción mecanística, es importante señalar la intrigante característica de encontrar un residuo con carga neta positiva, la Arg293, en la vecindad del centro anomérico. Si bien hemos mencionado que esta Arg interacciona con el grupo  $COO^-_{\text{GlcA}}$ , no deja de ser llamativo en el contexto de la reacción de transferencia pues se espera que se desarrolle una carga formal positiva en el  $C1_{\text{GalNAc}}$  y por consiguiente, un residuo con carga positiva como la Arg no parece ser el más adecuado para ocupar esta posición. Esta Arg293 está a 3.70 Å del carbono anomérico en cualquiera de los dos monómeros contenidos en la estructura 1OMZ. A lo largo de la DM esta distancia de mantiene a  $3.70 \pm 0.87$  Å, y a  $3.48 \pm 0.70$  Å en los complejos de Michaelis optimizados a nivel QM/MM. A continuación, se describirán los resultados de las simulaciones del mecanismo de ataque frontal y el ataque de Asp246 para formar el intermedio CGE.

### 5.3.1 Mecanismo de Ataque Frontal

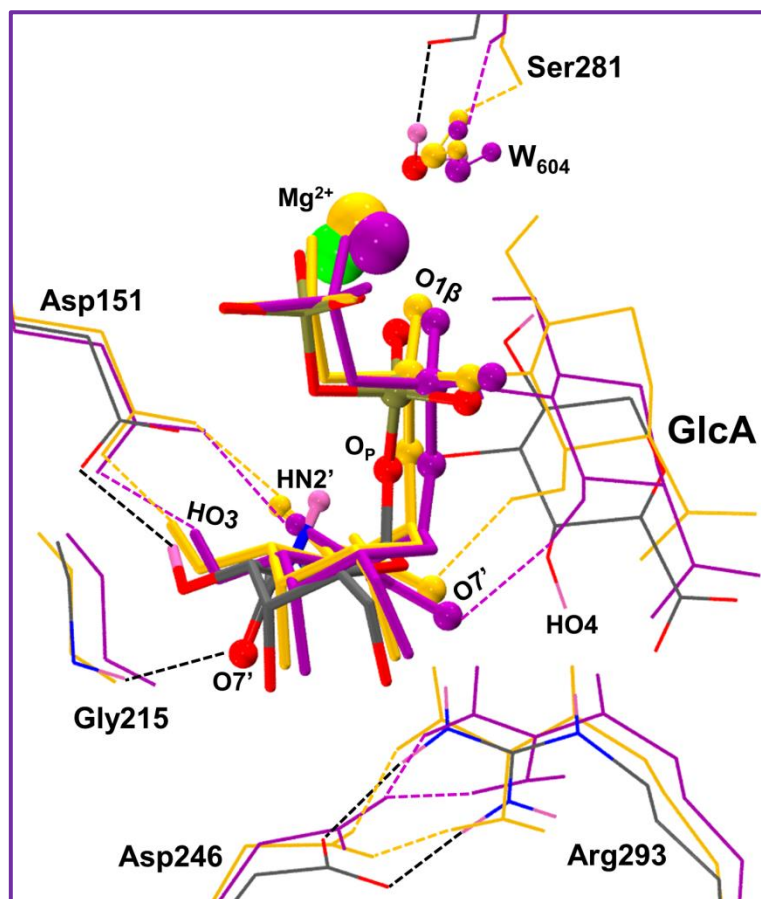
Para simular la transferencia de GalNAc hasta el GlcA del sustrato aceptor, primeramente, se seleccionaron 3 complejos de Michaelis desde la simulación de MD. Estos complejos tienen distancias  $C1_{GalNAc}-O4_{GlcA}$  en principio adecuadas para la reacción y fueron pues utilizados como estructuras de partida para las primeras exploraciones del mecanismo, las cuales se llevaron a cabo al nivel QM(BP86/SVP)/MM(CHARMM) utilizando la coordenada de reacción definida como:  $RC_A = [d(O_P-C1_{GalNAc}) - d(O4_{GlcA}-C1_{GalNAc}) - d(HO4_{GlcA}-O_P)]$ . Las barreras obtenidas de esta manera, resultan ser bastante altas ( $\geq 32$  kcal/mol). La comparación de estos reactivos nos permitió notar algunas interacciones que podrían ser claves para que nuestro modelo de EXTL2 llegue a catalizar eficientemente esta reacción. A continuación, se esquematizan las diferencias más relevantes entre estos 3 complejos de Michaelis.



**Figura 58.** Representación de 3 complejos de Michaelis seleccionados desde la MD, para los cuales se obtuvieron barreras energéticas  $\geq 32$  kcal/mol en la exploración del mecanismo de ataque frontal. Los carbonos grises se usan para representar a R-1 (barrera de energía más baja de  $\geq 32$  kcal/mol), y el color morado y naranja se usa para representar a R-2 y R-3, respectivamente. Las interacciones se indican a través de líneas punteadas y se mantiene el código de colores que se ha establecido para los reactivos: R-1 (gris), R-2 (morado) y R-3 (naranja). Los dos sustratos: UDP-GalNAc, para el cual se representan sólo el grupo difosfato junto a GalNAc, y GlcA $\beta$ 1-3Gal se dibujan en barras; los residuos enzimáticos se dibujan en líneas; para el Mg<sup>2+</sup> y las tres moléculas de agua de coordinación se utilizan esferas. Además de mostrar la orientación relativa entre los sustratos, también se señalan el conjunto de interacciones inter-sustratos y el modo de interacción de GlcA con tres residuos: Arg293, Tyr193 y Arg181.

En la Figura 58, se pueden observar varias diferencias en el modo de interacción de GlcA, tanto con los residuos Arg293, Tyr193 y Arg181, así como con UDP-GalNAc. Entre las interacciones establecidas por GlcA, destacan las que aparecen en R-1 con Tyr193 y Arg181;  $O6B_{GlcA}-HH_{Tyr193}$  y  $O6A_{GlcA}-HE/HH21_{Arg181}$ , en las que cada residuo forma una y dos

interacciones con O6B y O6A, respectivamente; además, O6B<sub>GlcA</sub> y Arg293 se encuentran a una distancia < 4 Å en R-1, por lo que es esperable que puedan interactuar electrostáticamente. Por otra parte, en R-2 y R-3 el COO<sup>-</sup><sub>GlcA</sub> adopta una conformación diferente en la que el O6A apunta hacia arriba y no interactúa con ningún residuo; en el caso de O6B, éste consigue interactuar con Tyr193 y Arg181 en R-2; aunque en R-3, el COO<sup>-</sup><sub>GlcA</sub> sólo mantiene una interacción con Arg181. Por último, en ambos complejos la distancia del grupo carboxilato hasta la Arg293 es > 6 Å. En cuanto a las interacciones inter-sustratos, se observa que HO3<sub>GlcA</sub> puede establecer dos tipos: una de ellas es de naturaleza anión-dipolo con O1β (R-2 y R-3) y en la otra, HO3<sub>GlcA</sub> interactúa con O<sub>P</sub> (R-1). Además, HO4<sub>GlcA</sub> establece una interacción con O6B<sub>GlcA</sub> (R-1), y O7'<sub>GalNAc</sub> (R-2 y R-3). Todas estas diferencias terminan en una distancia O4<sub>GlcA</sub>-C1<sub>GalNAc</sub> igual a 3.37, 3.65 y 3.66 Å en el R-1, R-2 y R-3, respectivamente, la cual no es marcadamente distinta, sin embargo las barreras de energía que se obtienen para R-2 y R-3 superan las 40 kcal/mol, en contraste con R-1 que alcanza un valor de 32 kcal/mol.



**Figura 59.** Representación de otro punto de vista de los 3 complejos de Michaelis seleccionados desde la MD, R-1 (gris), R-2 (morado) y R-3 (naranja). Las interacciones se indican a través de líneas discontinuas y se mantiene el código de colores que se ha establecido para los reactivos. El sustrato dador, UDP-GalNAc, del cual sólo se representan el grupo difosfato y GalNAc, se dibuja en barras; GlcA se dibuja con líneas al igual que los residuos enzimáticos. En el caso del Mg<sup>2+</sup> y el agua W<sub>604</sub>, se utilizan esferas. Se señalan las interacciones que forma parte del sustrato dador con la primera esfera de coordinación del Mg<sup>2+</sup> y algunos residuos importantes para los que se observan variaciones de un complejo a otro: Arg293, Asp246, Asp151 y Gly215. Adicionalmente, se dibuja a la Ser281 para mostrar que interactúa con W<sub>604</sub>.

Tal como se mencionó al inicio de esta sección, los complejos de Michaelis que resultan de la simulación de MD no muestran en ningún caso la interacción  $O2\beta-W_{604}$ , y en lugar de esto, la  $W_{604}$  aparece interaccionando con la Ser281 invariablemente. Modelar esta interacción pues (que en el complejo binario 1OMZ no aparece claramente,  $O2\beta$  y  $O_{W604}$  están a 5.2 Å), parece tener aquí una cierta complejidad metodológica aunque de estar presente sería una buena forma de estabilizar al grupo saliente. Otra de las interacciones que ha sido difícil de modelar en EXTL2 corresponde a la que establece el grupo NAc de GalNAc, aunque al menos en este caso se han observado dos modos de interacción en la MD. En uno de ellos, el  $HN2'$  forma una interacción anión-dipolo con Asp151 mientras  $O7'$  se orienta hacia el sustrato aceptor, formando una interacción con  $HO4_{GlcA}$  (R-2 y R-3); el segundo modo de interacción parece ser más adecuado para el proceso catalítico que se intenta modelar, ya que  $O7'$  forma un enlace de hidrógeno con el oxígeno carbonílico de la Gly215 y de este modo,  $HN2'$  se gira hacia GlcA quedando a una distancia de interacción de  $O3_{GlcA}$  ( $< 2.5$  Å en R-1). De la misma manera que ocurre con  $O2\beta-W_{604}$ , en las estructuras cristalográficas no hay indicio de la interacción  $HN2'_{GalNAc}-O3_{GlcA}$ , pero esta vez es debido a que el sustrato dador y aceptor han sido cristalizados en diferentes estructuras. Aun así, esta interacción llegó a aparecer en la MD y una de las estructuras representativas de este modo de interacción se usó para optimizar el complejo R-1.

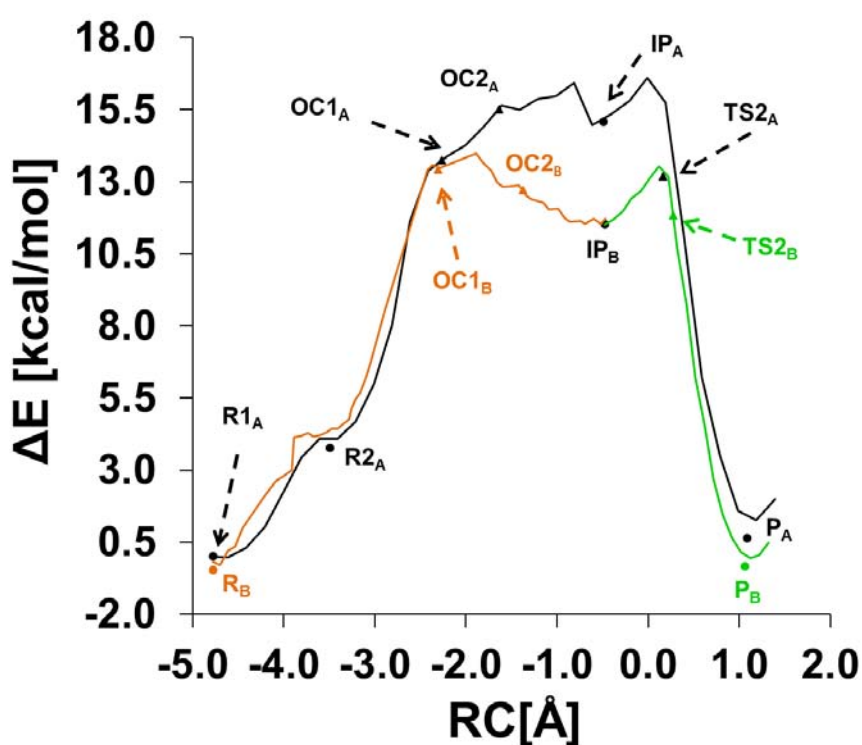
Por otra parte, si se observa la cara  $\beta$  del GalNAc es posible notar que la pareja Asp246-Arg293 varía sutilmente su modo de interacción. Uno de los casos extremos lo representa el complejo de Michaelis R-1, en el que los residuos interaccionan utilizando cada extremo de su respectiva cadena lateral y notablemente, ambas cadenas laterales quedan casi en un mismo plano. Por otro lado, en R-2 el Asp246 y la Arg293 interaccionan a través de uno de los oxígenos de Asp246, y ambas cadenas laterales caen más bien en planos perpendiculares (semejante a lo observado en 1ON8 y 1OMZ). Es importante notar que en 1ON8, la Arg293 interacciona cercanamente con  $COO^-_{GlcA}$  (3.1 Å) pero en nuestros resultados de MD, la Arg293 es capaz de posicionarse en un rango amplio de distancias con respecto a GlcA, que va desde 4 a 9 Å. Inicialmente, esto parece ser una consecuencia del ligero desplazamiento que sufre el sustrato aceptor, ubicándose un poco más lejos de la Arg293 y del sustrato dador. Sin embargo, los motivos que subyacen este comportamiento posiblemente están muy relacionados con la complejidad de la modelización de las interacciones entre los sustratos, así como entre la esfera de coordinación y el sustrato dador, las cuales resultan ser fundamentales no sólo para la reactividad, sino que también para que el aceptor se mantenga en el sitio activo durante la simulación.

El complejo QMMM en el que Arg293 y  $COO^-_{GlcA}$  están a una distancia que les permitiría interaccionar es el modelo R-1, con una distancia de 3.9 Å, bastante más pequeña que la encontrada en R-2 y R-3 ( $> 6$  Å en ambos complejos). Adicionalmente, la distancia de Arg293 con respecto a GalNAc (su carbono anomérico) en 1OMZ, resulta ser muy semejante a lo que hemos obtenido a través de la MD. De este modo, considerando que se trata del complejo a partir del cual se obtiene la barrera de energía menos alta y que  $COO^-_{GlcA}$ -Arg293 están dentro de una distancia de interacción, se seleccionó el producto del complejo R-1 para llevar a cabo la MD a nivel QM(SCC-DFTB)/MM(CHARMM) (ver apartado 5.3.2).

De esta manera, se obtiene un nuevo conjunto de complejos de Michaelis que son utilizados como punto de partida para la segunda exploración del mecanismo de ataque frontal. El nuevo conjunto de barreras de energía así obtenidas, se usan para calcular una estimación

de la barrera energética a nivel QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM), usando la ecuación 4.1. El valor calculado es de 18 kcal/mol, un número bastante más aceptable que concuerda con la barrera de energía libre fenomenológica derivada experimentalmente para otras GTs-ret.<sup>268</sup> Cabe mencionar que estas nuevas barreras difieren entre ellas en valores de hasta 10 kcal/mol, pero la variabilidad estructural (por ejemplo del sustrato aceptor) es menor que anteriormente por el modo en que se han generado (a partir de una posición del aceptor en el producto y con simulaciones más cortas y a nivel QMMM). Con el objetivo de profundizar el estudio mecanístico de EXTL2, se seleccionó uno de los complejos obtenidos, específicamente, el que presentaba la barrera de energía que más se aproximaba al valor medio estimado. Con este complejo, se prosiguió la investigación, se probaron distintos métodos QM (ver métodos) y, como se ha discutido, se acabó seleccionando el nivel de teoría QM(BP86/TZVP)/MM(CHARMM), con cálculos puntuales de energía M05-2X/TZVP. También se llevaron a cabo simulaciones de MD US para calcular la energía libre al nivel QM(SCC-DBTB)/MM(CHARMM), y además se realizaron análisis electrostáticos y de cargas para indagar en el rol de los diferentes residuos del sitio activo, especialmente el de la pareja Asp246-Arg293.

La primera exploración del mecanismo de ataque frontal realizada a este nivel, genera la curva de energía potencial que se muestra en la siguiente Figura 60 (perfil A, color negro). Se seleccionaron diferentes estructuras a lo largo del camino que se usaron para buscar estados de transición (TS) y mínimos (el intermedio IP).



**Figura 60.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM para el mecanismo de ataque frontal usando: la coordenada de reacción definida como  $RC_A = [d(O_P-C1_{GalNAc}) - d(O4_{GlcA}-C1_{GalNAc}) - d(HO4_{GlcA}-O_P)]$  con  $R1_A$  como estructura de partida, color negro; una segunda coordenada de reacción definida como  $RC_B = [d(O4_{GlcA}-C1_{GalNAc}) - d(O_P-C1_{GalNAc}) - d(HO4_{GlcA}-O6B_{GlcA})]$ , usando  $IP_B$  como estructura inicial, color naranja; y por último,  $RC_{AB} = [d(O_P-C1_{GalNAc}) - d(O4_{GlcA}-C1_{GalNAc}) - d(HO4_{GlcA}-O_P)]$ , que también utiliza como punto de partida el  $IP_B$ , color verde. Los círculos y triángulos se utilizan para indicar la



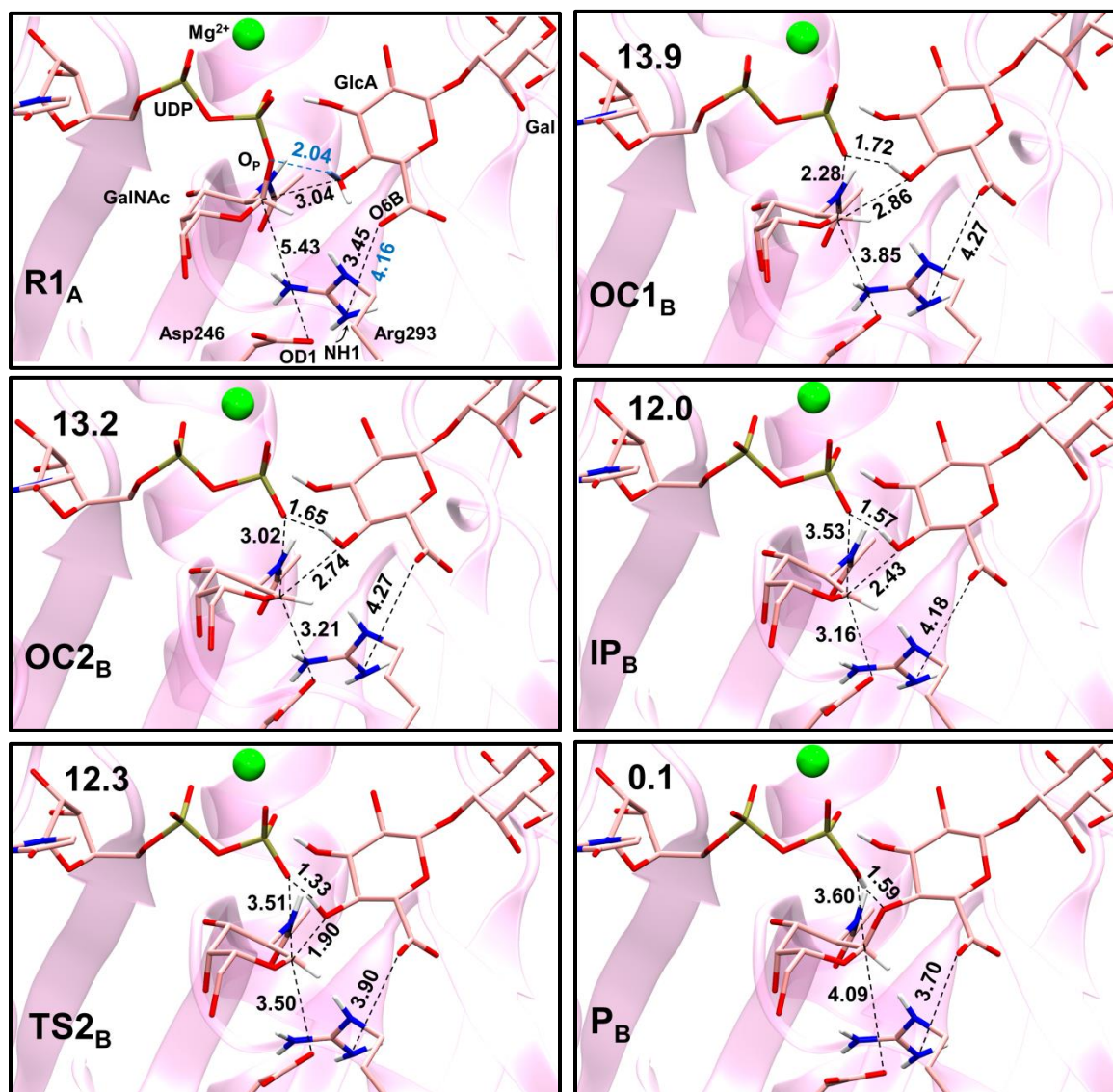
ubicación de mínimos (R, IP y P) y estados de transición (OC y TS), respectivamente, a lo largo del camino de reacción. Las especies OC no se denotan como TS debido a que no se encontró una frecuencia imaginaria.

Antes de entrar en la descripción detallada del mecanismo, es importante comentar que diferentes puntos estacionarios son localizados en la zona previa a la formación del IP<sub>B</sub> (especies OC). Las especies que se denotan como OC han sido optimizadas con algoritmos de búsqueda de TS, sin embargo no han podido caracterizarse con una frecuencia imaginaria, a diferencia del TS<sub>2</sub>. A pesar de esto, el análisis de la frecuencia real más baja que se encuentra en estas especies OC ( $\sim 50 \text{ cm}^{-1}$ ), nos ha permitido comprobar que el correspondiente vector describe la ruptura UDP-GalNAc, contrariamente a lo observado con los vectores asociados a las frecuencias reales más pequeñas que se obtienen para el verdadero mínimo intermedio (IP). Esto probablemente se debe a que la superficie de energía potencial es muy plana en esta región y existe un alto número de grados de libertad para el sistema, lo cual hace bastante difícil la caracterización de estas especies.

Retomando las exploraciones energéticas, el punto estacionario asociado al perfil negro con el valor más alto de energía corresponde a la estructura OC<sub>2A</sub>, con 15.5 kcal/mol y una distancia OD<sub>1Asp246</sub>-C<sub>1GalNAc</sub> = 4.60 Å. Luego, a partir de dos estructuras distintas del camino, dos IPs fueron optimizados, el IP<sub>A</sub> (15.1 kcal/mol) y el IP<sub>B</sub> (12.0 kcal/mol), en los que la distancia OD<sub>1Asp246</sub>-C<sub>1GalNAc</sub> es 4.16 Å y 3.16 Å, respectivamente. El primero quedaba claramente ajustado en el camino de reacción explorado a través de RC<sub>A</sub>, sin embargo el IP<sub>B</sub> se escapaba de la curva de energía (Figura 60). Por este motivo, se decidió iniciar nuevas exploraciones usando dos coordenadas de reacción diferentes, RC<sub>B</sub> para ir a hacia el reactivo y RC<sub>AB</sub> para ir hacia el producto (ver subapartado 5.3.3.1), con el fin de obtener el camino que conecta el IP<sub>B</sub> con reactivos y productos, y para ver si el reactivo obtenido era el mismo que antes (R<sub>1A</sub>) o había otro con el Asp246 más cerca del carbono anomérico. Para ambas exploraciones se usó como punto de partida el IP<sub>B</sub> y el reactivo obtenido de esta manera, R<sub>B</sub>, es prácticamente igual a R<sub>1A</sub> (R<sub>1A</sub> se conserva como estructura de referencia). A partir de estas nuevas exploraciones, se buscan estados de transición y se logra comprobar que OC<sub>1A</sub> y TS<sub>2A</sub> están conectados al IP<sub>B</sub>, aunque OC<sub>2A</sub> e IP<sub>A</sub> resultan ser parte de un camino de más alta energía. Así, se concluye que el perfil B ((Figura 60, color naranja-verde) es más favorecido que el perfil A y es por eso que el mecanismo que se describe a continuación corresponde a aquellos puntos estacionarios que conectan el camino de más baja energía (Figura 61).

El primer cambio observado en la exploración se relaciona con la reorientación del HO<sub>4GlcA</sub> hacia el O<sub>P</sub> aproximándose hasta 2.04 Å, un complejo (R<sub>2A</sub>) que tiene una energía 4 kcal/mol más alta que aquél en el que HO<sub>4GalNAc</sub> interacciona con O<sub>6BGalNAc</sub>. La reorientación de HO<sub>4GlcA</sub> provoca una pequeña rotación del COO<sup>-</sup><sub>GlcA</sub> en la cual O<sub>6BGalNAc</sub>, que acaba de perder su interacción con HO<sub>4GlcA</sub>, se gira hacia arriba distanciándose ligeramente de NE y NH1 de Arg293, aunque más marcadamente de NH1 (Figura siguiente). Una vez que la interacción HO<sub>4GlcA</sub>-O<sub>P</sub> se ha formado, la distancia C<sub>1GalNAc</sub>-O<sub>P</sub> comienza a elongarse hasta alcanzar la primera especie UDP----OC (OC<sub>1B</sub>), en la que esta distancia alcanza un valor de 2.28 Å mientras que HO<sub>4GlcA</sub>-O<sub>P</sub> y O<sub>4GlcA</sub>-C<sub>1GalNAc</sub> tienen valores de 1.72 y 2.86 Å, respectivamente. La interacción HO<sub>4GlcA</sub>-O<sub>P</sub> tiene un rol importante en el mecanismo que catalizan las GTs-ret, facilita el rompimiento del enlace UDP-GalNAc estabilizando la carga negativa que se desarrolla sobre UDP (catálisis asistida por sustrato). La especie OC<sub>1B</sub> tiene una energía de 13.9 kcal/mol y muestra dos características importantes: se ha desarrollado una carga de 0.34 u.a en el centro anomérico, que junto al hecho de que la

distancia (C1-O5)<sub>GalNAc</sub> se ha reducido hasta 1.28 Å, evidencian el carácter de doble enlace entre estos dos átomos. Adicionalmente, el azúcar se distorsiona desde una silla <sup>4</sup>C<sub>1</sub> en el reactivo, hasta una conformación intermedia entre <sup>4</sup>E y <sup>4</sup>H<sub>3</sub>; como consecuencia de esto, el Asp246 rota ligeramente, posicionándose por la suma de los dos factores a 3.85 Å del carbono anomérico. Después, la especie OC2<sub>B</sub> (13.2 kcal/mol) tiene un carácter bastante más disociado que OC1<sub>B</sub>, cuyas distancias C1<sub>GalNAc</sub>-O<sub>P</sub> y HO4<sub>GlcA</sub>-O<sub>P</sub> que han alcanzado valores de 3.02 y 1.63 Å, respectivamente; por la cara α de GalNAc, el O4<sub>GlcA</sub> se ha aproximado hasta una distancia de 2.74 Å, 0.29 Å más cerca que lo que estaba en el reactivo. Por la cara β de GalNAc, el Asp246 se ubica a 3.21 Å del carbono anomérico, un acercamiento de 2.22 Å con respecto al reactivo (5.43 Å). Como se acaba de mencionar, parte de la causa de este acercamiento se debe a la distorsión de GalNAc, que se encuentra en una conformación intermedia entre <sup>4</sup>E y <sup>4</sup>H<sub>3</sub>, según los parámetros de Cremer-Pople. Este tipo de conformación lleva al C1<sub>GalNAc</sub> hacia abajo, acercándolo al Asp246, por lo que se presume que la distorsión y el desarrollo de carga positiva en el azúcar, promueven este discreto cambio conformacional del grupo carboxilato en el Asp246.



**Figura 61.** Representación molecular de los puntos estacionarios para el mecanismo de ataque frontal. En el caso del reactivo, la representación de R1<sub>A</sub> (igual a R1<sub>B</sub>) incluye también la distancia HO4<sub>GlcA</sub>-O<sub>P</sub> y NH1<sub>Arg293</sub>-O6B<sub>GlcA</sub> correspondiente a R2<sub>A</sub> (2.04 y 4.16 Å, respectivamente, color celeste), que son los únicos cambios importantes al pasar de R1<sub>A</sub> desde R2<sub>A</sub>. Las distancias se dan Å y las energías relativas en kcal/mol.

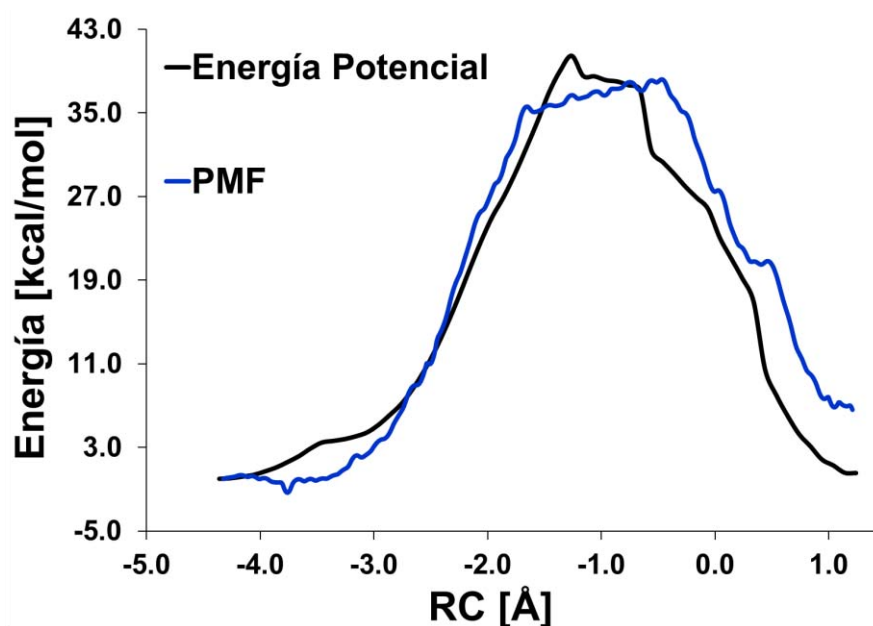
En el IP<sub>B</sub>, la distancia OD1<sub>Asp246</sub>-C1<sub>GalNAc</sub> es de 3.16 Å y UDP se encuentra 3.53 Å del carbono anomérico. Este distanciamiento adicional de UDP ocurre simultáneamente con el acercamiento de Asp246, y permite que O4<sub>GlcA</sub> se posicione a 2.43 Å de C1<sub>GalNAc</sub> por la cara  $\alpha$  de GalNAc. La distancia HO4<sub>GlcA</sub>-O<sub>P</sub> es de 1.57 Å, interacción que contribuye a que UDP y GlcA sigan interaccionando cuando GlcA se acerca al carbono anomérico. La conformación de GalNAc en el IP<sub>B</sub> es una <sup>4</sup>E y la energía es de 12.0 kcal/mol, 2 kcal/mol más estable que OC1<sub>B</sub>. Es importante indicar que además de que el hidroxilo nucleofílico se ha acercado a UDP y a GalNAc a través de HO4<sub>GlcA</sub> y O4<sub>GlcA</sub>, respectivamente, el mayor grado de asociación entre GlcA-GalNAc y GlcA-UDP también se observa en otras de las interacciones, estas son; O3<sub>GlcA</sub>-HN2'<sub>GalNAc</sub>, que disminuye hasta 1.77 Å (2.50 Å en el reactivo); HO3<sub>GlcA</sub>-O1 $\beta$  ha disminuido 0.12 Å con respecto al reactivo, quedando en un valor

de 1.60 Å en  $IP_B$ ; y  $O2\beta-W_{604}$  ha alcanzado un valor de 1.60 Å en  $IP_B$  (1.81 Å en el reactivo). Creemos que el  $IP_B$  es un complejo cuyas características electrónicas y estructurales lo configuran como el precursor de nucleófilo activado ( $HO4_{GlcA}-O_P = 1.57$  Å y  $O4_{GlcA}-C1_{GalNAc} = 2.43$  Å). En este intermedio, el nucleófilo de la reacción (GlcA) se mantiene interaccionando cercanamente con ambos fragmentos de sustrato dador, UDP y el azúcar, especialmente con UDP que actuará como base catalítica. De esta manera, la reacción de transferencia procede a través del  $TS2_B$  con una energía de 12.3 kcal/mol y una frecuencia de  $85i\text{ cm}^{-1}$ , en el que la distancia  $O4_{GlcA}-C1_{GalNAc}$  y  $HO4_{GlcA}-O_P$  tienen valores de 1.90 Å y 1.33 Å, respectivamente, mientras que UDP se mantiene prácticamente a la misma distancia que en el  $IP_B$ . Adicionalmente, el Asp246 se ha distanciado de  $C1_{GalNAc}$  (3.50 Å) en el  $TS2_B$ , en parte debido al sutil cambio de la conformación de GalNAc (que lleva  $C1_{GalNAc}$  hacia arriba), y a la ligera rotación del grupo carboxilato de Asp246, que le permite recuperar su orientación con respecto a la Arg293, mejorando la interacción entre estos. Así, el rompimiento del sustrato dador UDP-GalNAc sería el paso limitante de la reacción de transferencia, aunque las barreras que representan  $OC1_B$  y  $TS2_B$  son bastante similares. Finalmente, el producto ( $P_B$ ) tiene una energía de 0.1 kcal/mol y GalNAc recupera la conformación de silla  ${}^4C_1$ .

De este modo, el proceso de modelización realizado hasta ahora nos permite caracterizar un mecanismo de ataque frontal e identificar tres interacciones que podrían estar ayudando a facilitar la disociación del sustrato dador UDP-GalNAc. Una de ellas involucra a la molécula de agua de coordinación  $W_{604}$ , esta es la interacción  $O2\beta-W_{604}$ ; también se encuentra otra interacción en la que participa uno de los oxígenos del grupo fosfato que también es parte de la primera esfera de coordinación de  $Mg^{2+}$ , esto es,  $O1\beta-HO3_{GlcA}$ ; por último, se tiene la interacción  $HN2'_{GalNAc}-O3_{GlcA}$ . Como se comentó antes, esta interacción fue una de las más difíciles de modelar debido a que el grupo NAc aparecía interaccionando con el Asp151, lo cual se debe a que UDP-GalNAc y GlcA $\beta$ 1-3Gal han sido cristalizados en dos estructuras diferentes, 1OMZ y 1ON8, respectivamente. Adicionalmente, estudios de Calorimetría Isoterma de Titulación (ITC, por su sigla en inglés) han indicado que el grupo NAc es clave en la unión del sustrato dador, dado que ni UDP-Glc ni tampoco UDP-Gal se unen a EXTL2. Los autores atribuyen el incremento de entropía luego de unirse UDP-GalNAc/UDP-GlcNAc a la expulsión de aguas del sitio activo por parte del grupo NAc. Nuestro estudio sugiere que el grupo NAc podría estar también involucrado en la correcta orientación del sustrato aceptor a través de la interacción  $HN2'_{GalNAc}-O3_{GlcA}$ .

### 5.3.1.1 Cálculo de Energía Libre para el Mecanismo de Ataque Frontal

Con el objetivo de obtener una estimación de los efectos térmicos y entrópicos a lo largo de la coordenada de reacción, se han llevado cabo simulaciones de MD US al nivel QM(SCC-DFTB)/MM(CHARMM). A pesar de que este método sobreestima la barrera energética de esta reacción (el valor calculado es de 40.5 kcal/mol), se asume que nos permitirá obtener una estimación cualitativa de la diferencia entre la barrera de energía potencial y la de energía libre.



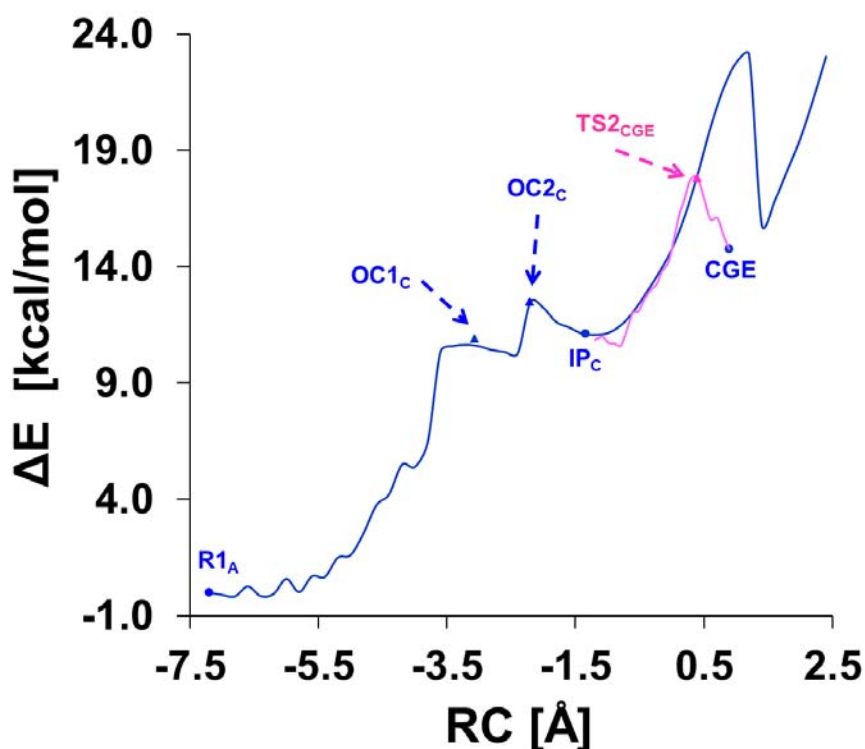
**Figura 62.** Perfil de energía potencial (línea negra) y perfil de energía libre (línea azul), calculados al nivel QM(SCC-DFTB)/MM(CHARMM).

Al comparar las curvas de energía potencial y energía libre para el mecanismo de ataque frontal, se observa que la segunda presenta una zona de meseta entre los valores de RC - 1.7 y -0.5 Å, que corresponden a la zona de coexistencia de las especies UDP----OC. Como se ha mencionado previamente, esta región de la superficie de energía es bastante complicada y muy sensible a las reorganizaciones de los grupos con habilidad nucleofílica como Asp246, GlcA y UDP con respecto al carbono anomérico. Aun así, la estimación de los efectos térmicos y entrópicos en el valor de la barrera resultan ser relativamente pequeños y por consiguiente, no alteran la viabilidad del mecanismo propuesto. De esta manera, nuestros resultados nos permiten indicar que el mecanismo de ataque frontal es posible en EXTL2.

### 5.3.2 Mecanismo de Ataque del Asp246: Formación del CGE

Aunque inicialmente no se había contemplado simular este proceso debido a que el residuo nucleofílico estaba a 5.43 Å del carbono anomérico, la aparición del  $IP_B$  en el cual la distancia  $OD1_{Asp246}-C1_{GlcNAc}$  tiene un valor de 3.16 Å, motivó el estudio de la formación del CGE con el Asp246 actuando como nucleófilo.

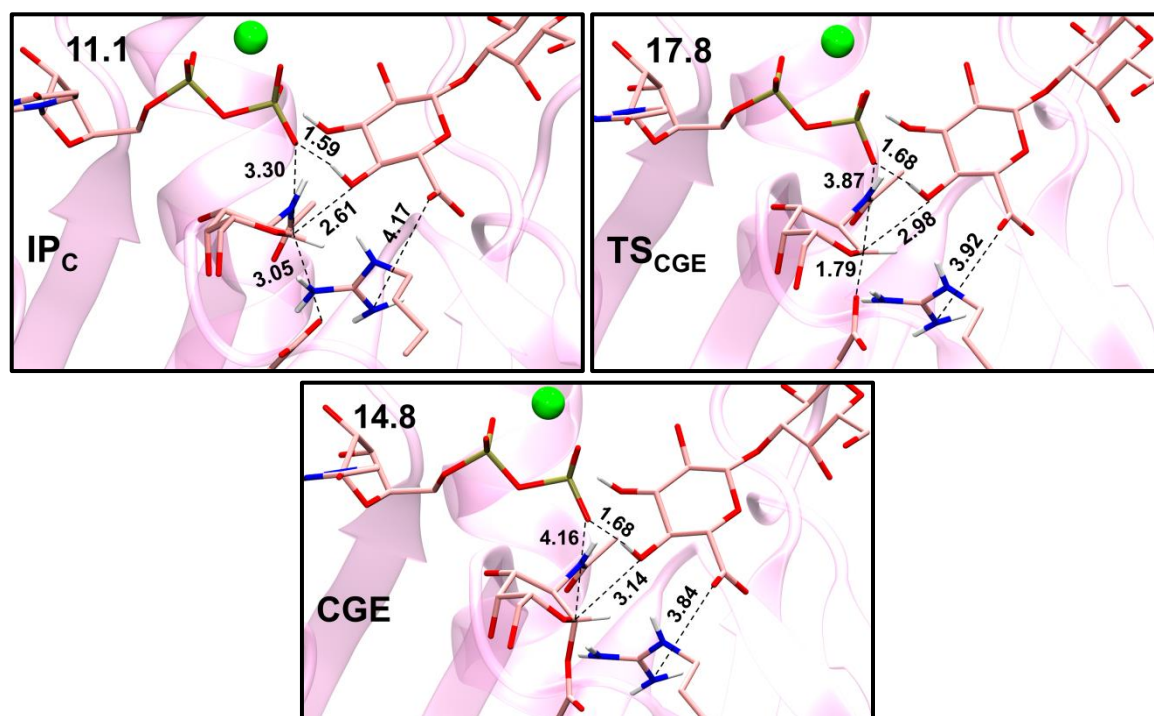
Para realizar la exploración de este mecanismo, se utilizó el complejo de Michaelis ( $R1_A$ ) y la coordenada de reacción definida como  $RC_{DA} = [d(O_P-C1_{GlcNAc}) - d(OD1_{Asp246}-C1_{GlcNAc}) - d(HO4_{GlcA}-O_P)]$ . A partir de esta exploración se seleccionan varias estructuras para iniciar la búsqueda de estados de transición (OC y TS) y mínimos (CGE). Curiosamente, la curva de energía para el ataque del Asp246 muestra una zona relativamente plana que sugiere la presencia de un IP. De hecho, la optimización completa de uno de estos puntos nos permitió caracterizar el  $IP_C$  previamente a la formación del CGE, lo cual indica que no se trata de un desplazamiento directo del UDP por parte del Asp246 (mecanismo  $S_N2$ ), sino de un mecanismo en dos etapas tipo- $S_N1$ . Este tipo de proceso también ha sido descrito para la enzima  $\alpha$ -1,3-galactosiltransferasa nativa (GTB) y su mutante en la posición del residuo nucleofílico E303 ( $GTB_{E303C}$ ).<sup>132</sup> La optimización del CGE utilizando las estructuras finales del camino de reacción, condujo a un intermedio covalente que quedaba fuera de la curva de energía potencial, tanto a nivel QM(BP86/TZVP)/MM(CHARMM) como a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM (Figura 63). A continuación, se muestra la exploración del camino de reacción a este último nivel.



**Figura 63.** Curva de energía potencial a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM para el mecanismo de ataque del Asp246 usando la coordenada de reacción definida como  $RC_{DA} = [d(OP-C1_{GalNAc}) - d(OD1_{Asp246}-C1_{GalNAc}) - d(HO4_{GlcA-OP})]$ , con  $R1_A$  como estructura de partida, color azul. También se define una segunda coordenada de reacción,  $RC_E = [d(OD1_{Asp246}-C1_{GalNAc}) - d(OP-C1_{GalNAc})]$ , usando el CGE como punto inicial, color rosado. Los círculos y triángulos se utilizan para indicar la ubicación de mínimos (R, IP y CGE) y estados de transición (OC y TS), respectivamente, a lo largo del camino de reacción.

Como se puede observar en la Figura 63, luego de que se forma el  $IP_C$  el comportamiento de la curva de energía potencial azul sugiere términos espurios de energía, los cuales posiblemente están asociados a la alta tensión de las estructuras exploradas bajo la  $RC_{DA}$  en esta última zona. En vista de esto, se decidió explorar esta zona final bajo otra coordenada de reacción ( $RC_E$ ), utilizando el CGE optimizado como estructura de partida (ver subapartado 5.3.3.2). Esta es la curva de energía potencial rosada, que permite conectar el  $IP_C$  con el CGE a través del  $TS_{CGE}$ .

Al igual que ocurre con el mecanismo de ataque frontal, el ataque nucleofílico del Asp246 se inicia con la disociación del dador UDP-GalNAc. Dado que las características e interacciones más relevantes del proceso de disociación UDP-GalNAc han sido ya descritas y no se ven afectadas por la coordenada de reacción que se utilice ( $RC_{DA}$  o  $RC_B$ ), a continuación se describirá en detalle la formación del CGE a partir del  $IP_C$ .



**Figura 64.** Representación molecular de los puntos estacionarios para el ataque del Asp246, obtenidos al nivel QM(BP86/TZVP)/MM(CHARMM). Sólo se muestran las estructuras a partir del  $IP_C$ , indicando en cada caso las distancias en Å y las energías relativas en kcal/mol.

La  $RC_{DA}$  que se aplicó para simular el ataque nucleofílico del Asp246 conduce a un  $IP_C$  en el que existe una carga positiva adicional de 0.37 u.a. con respecto a  $R1_A$ , y la conformación

de GalNAc corresponde a una intermedia entre  ${}^4E$  y  ${}^4H_3$ . La formación de esta especie es facilitada por la catálisis asistida por sustratos ( $HO4_{GlcA}-O_P$ ), y otras interacciones claves que hemos identificado en EXTL2, como  $O2\beta-W_{604}$ ,  $HN2^i_{GalNAc}-O3_{GlcA}$  y  $O1\beta-HO3_{GlcA}$ , que al ser interacciones inter-sustratos también forman parte de esta estrategia. La formación del intermedio covalente CGE ocurre a través del  $TS_{CGE}$ , en el que la conformación de GalNAc se ve un poco más distorsionada, aproximándose casi exactamente a la  ${}^4E$ . Este  $TS_{CGE}$  tiene una carga de 0.10 u.a. (0.09 u.a. menor que lo calculado para el  $TS2_B$ ), y representa una energía de 17.8 kcal/mol. Además, posee una frecuencia de  $70.8i\text{ cm}^{-1}$  asociada al vector propio que describe la formación del enlace  $C1_{GalNAc}-OD1_{Asp246}$ . La barrera que se debe superar para formar el CGE es 5.5 kcal/mol más alta que la que se obtiene para el ataque del GlcA ( $TS2_B$ ), que es de 12.3 kcal/mol, por lo que la formación del intermedio CGE es un camino desfavorecido frente al mecanismo de ataque frontal a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM.

Ahora que ya se ha descrito el mecanismo de reacción con el Asp246 como nucleófilo, es posible examinar y discutir ciertos detalles estructurales. Como ya se mencionó, la síntesis del CGE ocurre a través de un mecanismo en dos etapas que involucra la formación del  $IP_C$ . Las estructuras del camino que conduce al  $IP_C$  ( $RC_{DA}$ ) son bastantes parecidas a las que conducen al  $IP_B$  ( $RC_B$ ), aunque con algunas diferencias que se relacionan con la posición de los grupos con habilidad nucleofílica Asp246 ( $OD1$ ), UDP ( $O_P$ ) y GlcA ( $O4$ ), con respecto al carbono anomérico. Además, las especies  $OC_B$  y  $OC_C$  optimizadas a partir de los diferentes caminos son energéticamente muy similares.

**Tabla 13.** Energías QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM relativas al reactivo  $R1_A$ ; para la formación de las especies oxocarbenio exploradas bajo la  $RC_A$ , especies con subíndice A; la  $RC_B$ , en el caso  $OC1_B$  y  $OC2_B$ ; y la  $RC_{AB}$ , para el  $IP_B$  y el  $TS2_B$ . En el caso de las especies  $OC1_C$ ,  $OC2_C$  e  $IP_C$ , se ha utilizado la  $RC_{DA}$ , mientras que para obtener el  $TS_{CGE}$ , se usó la  $RC_E$ . Además se resalta en *negrita* la barrera de energía global calculada para cada uno de los procesos.

| UDP----OC                | $\Delta E$ [kcal/mol] | $\nu$ [ $\text{cm}^{-1}$ ] | $C1_{GalNAc}-O_P$ | $C1_{GalNAc}-OD1_{Asp246}$ | $C1_{GalNAc}-O4_{GlcA}$ | $HO4_{GlcA}-O_P$ |
|--------------------------|-----------------------|----------------------------|-------------------|----------------------------|-------------------------|------------------|
| <b>OC1<sub>A</sub></b>   | 13.8                  | 56.4                       | 2.29              | 4.86                       | 2.84                    | 1.71             |
| <b>OC2<sub>A</sub></b>   | <b>15.5</b>           | 56.8                       | 2.77              | 4.60                       | 2.76                    | 1.64             |
| <b>IP<sub>A</sub></b>    | 15.1                  | >65.0                      | 3.50              | 4.16                       | 2.46                    | 1.56             |
| <b>TS2<sub>A</sub></b>   | 13.2                  | 47.6                       | 3.56              | 3.40                       | 2.01                    | 1.38             |
| <b>OC1<sub>B</sub></b>   | <b>13.9</b>           | 53.3                       | 2.28              | 3.85                       | 2.86                    | 1.72             |
| <b>OC2<sub>B</sub></b>   | <b>13.2</b>           | 60.9                       | 3.02              | <b>3.21</b>                | 2.74                    | 1.65             |
| <b>IP<sub>B</sub></b>    | 12.0                  | >67.0                      | 3.53              | 3.16                       | 2.43                    | 1.57             |
| <b>TS2<sub>B</sub></b>   | 12.3                  | 85.4 <i>i</i>              | 3.51              | 3.50                       | 1.90                    | 1.33             |
| <b>OC1<sub>C</sub></b>   | 11.3                  | 53.1                       | 2.29              | 3.92                       | 2.87                    | 1.71             |
| <b>OC2<sub>C</sub></b>   | <b>12.5</b>           | 55.6                       | 2.94              | <b>3.49</b>                | 2.72                    | 1.66             |
| <b>IP<sub>C</sub></b>    | 11.1                  | >80.0                      | 3.30              | 3.05                       | 2.61                    | 1.59             |
| <b>TS2<sub>CGE</sub></b> | <b>17.8</b>           | 70.8 <i>i</i>              | 3.87              | 1.79                       | 2.98                    | 1.68             |

Al analizar la información de la Tabla 13, se puede reparar en detalles que aunque parezcan nimios pueden dar información relevante respecto a las estrategias catalíticas de EXTL2. Por ejemplo; en  $OC1_C$  y  $OC1_B$  los tres grupos nucleofílicos tienen sus respectivos átomos  $O_P$ ,  $OD1_{Asp246}$  y  $O4_{GlcA}$  localizados a 2.3, 3.9 y 2.9 Å del  $C1_{GalNAc}$ , respectivamente. Así mismo, la distancia  $HO4_{GlcA}-O_P$  es de 1.7 Å en ambas especies, sin embargo  $OC1_C$  es 2.6 kcal/mol más estable que  $OC1_B$ . Contrariamente a lo observado para  $OC1_B$  y  $OC1_C$ , cuya



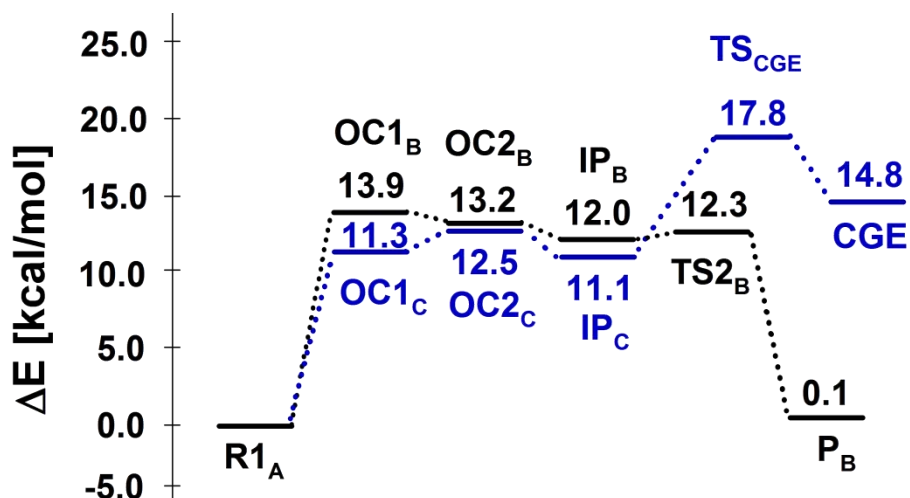
similitud estructural es alta pero energéticamente se diferencian por 3 kcal/mol, OC1<sub>A</sub> y OC1<sub>B</sub> poseen prácticamente la misma energía a pesar de que el Asp246 está 1 Å más lejos de C1<sub>GalNAc</sub> en OC1<sub>A</sub>. Esto deja ver que el Asp246 es perfectamente capaz de estabilizar el oxocarbenio desde 4 Å (OC1<sub>C</sub>), y además evidencia la complejidad de esta región de la superficie de energía potencial que parece ser muy sensible a las pequeñas reorganizaciones atómicas alrededor del C1<sub>GalNAc</sub>.

Retomando la descripción, las especies OC2<sub>C</sub> y OC2<sub>B</sub> continúan siendo muy semejantes, aunque aparece una pequeña diferencia en cuanto a la distancia OD1<sub>Asp246</sub>-C1<sub>GalNAc</sub>, que en OC2<sub>C</sub> es 0.28 Å más larga que en el OC2<sub>B</sub>, lo cual se traduce en una especie OC2<sub>C</sub> 0.7 kcal/mol más estable (valores resaltados en la Tabla 13). Esto llama la atención pues se esperaría que OC2<sub>B</sub> fuera más estable que OC2<sub>C</sub> al tener al Asp246 más cerca del centro anomérico con la carga positiva, sin embargo esto no ocurre así. Luego de estas pequeñas reorganizaciones, se forma el IP<sub>C</sub> que es 0.9 kcal/mol más estable que el IP<sub>B</sub>, y que tiene al O4<sub>GlcA</sub> 0.2 Å más lejos del C1<sub>GalNAc</sub> que el O4<sub>GlcA</sub> en el IP<sub>B</sub>. Esta ligera reorganización parece permitir que tanto el O<sub>P</sub> como el OD1<sub>Asp246</sub>, se posicionen 0.2 Å más cerca del C1<sub>GalNAc</sub> en el IP<sub>C</sub>. Resumiendo, queda claro que la rotura del enlace UDP-GalNAc lleva a una zona de alta energía relativamente plana en la que se encuentran varias estructuras posibles que se diferencian por la posición relativa de los tres átomos nucleofílicos (O<sub>P</sub>, O4<sub>GlcA</sub> y OD1<sub>Asp246</sub>) a C1<sub>GalNAc</sub>.

Por último, vale la pena mencionar la información que se puede obtener desde OC2<sub>A</sub>, un punto estacionario del camino de más alta energía para el mecanismo de ataque frontal (Figura 60, perfil A). Debido a que ya se trabaja sobre estructuras que poseen las interacciones óptimas para la catálisis, es altamente probable que esta especie UDP---OC sea menos estable porque está menos disociada (C1<sub>GalNAc</sub>-O<sub>P</sub> = 2.77 Å) que las encontradas en el perfil B o C, lo que lógicamente, debe suponer un mayor impedimento estérico para el acercamiento de GlcA. Este menor grado de disociación podría atribuirse a que el Asp246 está más lejos de C1<sub>GalNAc</sub> y por ende, la interacción electrostática puede aún mejorar. Adicionalmente, creemos que un acercamiento más pronunciado del Asp246 puede promover una especie UDP----OC más disociada no sólo por la estabilización electrostática que es capaz de entregar, sino por la capacidad de empujar o desplazar al UDP más lejos del C1<sub>GalNAc</sub>, creando un espacio más grande para la entrada y ataque de GlcA, y minimizando de esta manera la tensión estérica asociada a tal proceso.

Por último, pensamos que el conjunto de estructuras UDP----OC (hasta el IP), derivan de la descripción del mismo proceso químico, la rotura del dador UDP-GalNAc. Esto nos hace suponer que deben ser miembros del mismo conjunto de estructuras que el sistema puede adoptar cuando va desde el conjunto de reactivos hasta el conjunto de intermedios IP. En esta situación, pequeñas reorganizaciones atómicas alrededor del C1<sub>GalNAc</sub> pueden interconvertir estas especies,<sup>116</sup> y todas ellas están contribuyendo a la barrera de energía aparente de la disociación UDP-GalNAc.

A continuación se muestra un esquema energético resumen para los mecanismos explorados.



**Figura 65.** Esquema de energía a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM para el mecanismo de ataque frontal (color negro) y el mecanismo de ataque del Asp246 (color azul). La disociación del sustrato dador UDP-GalNAc, conduce a la formación del intermedio par iónico (IP<sub>B</sub> o IP<sub>C</sub>), un proceso que posiblemente resulte de la combinación de los caminos B y C. Una vez se alcanza el IP, el UDP puede actuar como base catalítica promoviendo el ataque de GlcA (TS<sub>2B</sub>), o el Asp246 puede seguir acercándose al carbono anomérico para formar un segundo intermedio, el CGE, a través del TS<sub>CGE</sub>.

Como puede verse claramente en la Figura 65, sin importar si es el GlcA o el Asp246 quien se aproxima más al C1<sub>GalNAc</sub> mientras se disocia UDP-GalNAc, se formará un IP estructural y energéticamente muy similar. Una vez que se ha alcanzado esta especie en la que la distancia HO<sub>4GlcA</sub>-O<sub>P</sub> < 1.6 Å, se puede producir el ataque de GlcA promovido por la base catalítica UDP (TS<sub>2B</sub>), y también podría atacar el Asp246 (TS<sub>CGE</sub>). La diferencia energética de estos dos ataques nucleofílicos (5.5 kcal/mol), se traduce en una  $k_{cat}$  varios órdenes de magnitud más lenta para la formación del intermedio covalente CGE. De este modo, parece ser que una vez formado el intermedio IP, el ataque del aceptor ocurrirá sin más reorganizaciones que las necesarias para formar el producto final, sin pasar por un intermedio covalente como el CGE.

### 5.3.2.1 Interacciones Favorables para la Catálisis

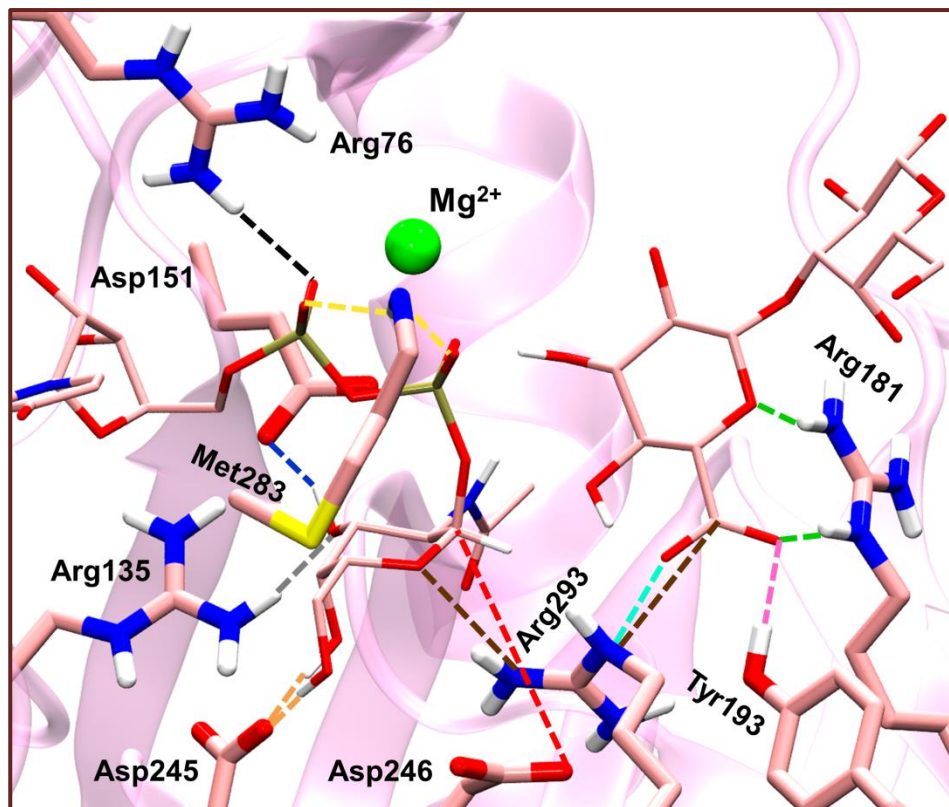
En los apartados previos se han caracterizado mecanismos que involucran especies altamente disociadas, en las que la distancia C1<sub>GalNAc</sub>-O<sub>P</sub> alcanza valores máximos en las estructuras de IP o TS (Tabla 13). También se ha mostrado que a pesar de que el Asp246 llega a ubicarse suficientemente cerca del carbono anomérico durante la disociación UDP-GalNAc, su ataque requeriría de la formación de un TS<sub>CGE</sub> altamente disociado (C1<sub>GalNAc</sub>-O<sub>P</sub> = 3.87 Å), que representa una etapa varios órdenes de magnitud más lenta que el ataque

del aceptor GlcA. De esta manera, podemos concluir que EXTL2 cataliza un mecanismo de ataque frontal evitando la tarea adicional de estabilizar especies intermedias tan altamente disociadas como el  $TS_{CGE}$  y por consiguiente, la formación del CGE.

Algunas de las interacciones que se han modelizado e identificado como claves en este mecanismo incluyen aquellas que se establecen entre los sustratos, o entre dador y esfera de coordinación del  $Mg^{2+}$  (en lo sucesivo, sólo esfera). Por ejemplo, las interacciones que forma el grupo fosfato  $\beta$  con la agua de coordinación  $W_{604}$  ( $O2\beta-W_{604}$ ), y la que establece con el sustrato aceptor ( $O1\beta-HO3_{GlcA}$ ), ambas han resultado ser importantes debido a que ayudan a estabilizar la creciente carga negativa que se desarrolla sobre el UDP a consecuencia de la disociación UDP-GalNAc. También se ha modelizado otra interacción que involucra al sustrato aceptor pero esta vez interaccionando con GalNAc ( $O3_{GlcA}-HN2'_{GalNAc}$ ), aunque esta interacción parece tener más relación con la correcta orientación de GlcA que con la estabilización de la especie disociada. Por último, pero la más importante, se tiene la interacción  $HO4_{GlcA}-O_P$  que se forma tempranamente en el camino de reacción. Esta interacción involucra al hidroxilo nucleofílico y ha sido descrita en todas las GTs-ret para las que se ha estudiado el mecanismo a nivel computacional hasta la fecha.<sup>111,112,116,123-125,131,220,268,285</sup> Su importancia no sólo radica en la estabilización de la creciente carga negativa de UDP sino que también es vital en el ataque nucleofílico, cuando UDP actúa como base catalítica activando al nucleófilo (GlcA). Independiente de la naturaleza química del sustrato aceptor o el monosacárido que se transferirá hasta éste, esta última interacción siempre se establece y en ese sentido, puede ser considerada una estrategia catalítica transversal a todas las GTs-ret, la catálisis asistida por sustratos. A pesar de ello, el resto de interacciones inter-sustratos o dador-esfera también pueden clasificarse dentro de la catálisis asistida por sustrato, aunque en este caso habrá variaciones de una GT-ret a otra, en dependencia del tipo de azúcar transferible y el tipo de molécula aceptor. En este contexto, vale la pena mencionar ciertas interacciones que caen dentro de la catálisis asistida por sustratos y que han sido identificadas en otras GTs-ret estudiadas en el grupo. En la enzima  $\alpha$ -1,4-galactosiltransferasa de *Nesseira meningitidis* (LgtC),<sup>125</sup> tanto el hidroxilo nucleofílico ( $HO4-O4$ ) como su vecino ( $HO3-O3$ ), interaccionan con el fosfato  $\beta$  a través del  $O_P$ ; en  $\alpha$ -1,3-galactosiltransferasa ( $\alpha$ 3GalT), la interacción entre el hidroxilo nucleofílico y  $O_P$  se forma pronto en el camino de reacción y además se establece una interacción dentro de UDP-Gal que involucra al fosfato  $\beta$  y un hidroxilo de la galactosa ( $HO2$ ); en GalNAc-T2, en la que el sustrato aceptor no es un carbohidrato sino que un aminoácido (Thr7), se establece la interacción entre hidroxilo nucleofílico y  $O_P$ , y también una segunda interacción entre el fosfato  $\beta$  y el segmento base de la Thr7; en PpgS se observa la interacción entre el hidroxilo nucleofílico y  $O_P$ , y además el grupo carboxilato del aceptor está estructuralmente bien posicionado para estabilizar la creciente carga positiva de la especie OC. Por último, en otras GTs-ret para las que se ha investigado el mecanismo, también ha sido indicada la presencia de la interacción entre hidroxilo nucleofílico y  $O_P$ , como es el caso de OtsA y Krep2p/Mnt1p.

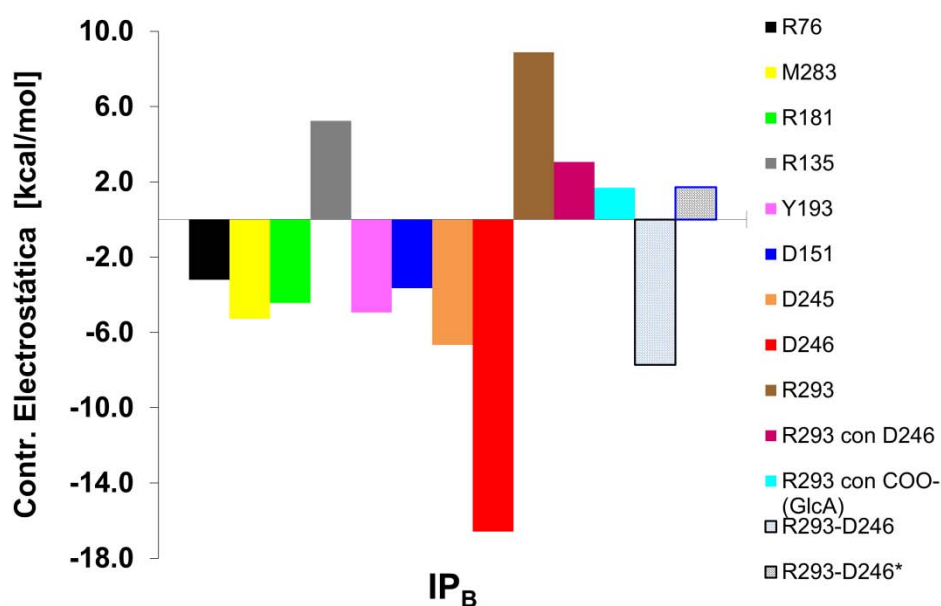
Además de las implicancias mecánicas que tienen las interacciones inter-sustratos, es razonable pensar que estas interacciones también son capaces de modular la orientación relativa entre el sustrato dador y el aceptor y por consiguiente, podrían estar relacionadas con la obtención de la regio y estereoespecificidad deseada. En este contexto, pensamos que el enlace hidrógeno  $HN2'_{GalNAc}-O3_{GlcA}$  tiene el rol de orientar el GlcA apropiadamente para la catálisis.

A continuación, se utilizan los resultados de los análisis electrostáticos para indagar en el rol de los residuos del sitio activo, identificando las interacciones que más ayudan a estabilizar las especies intermedias durante la reacción. En la siguiente Figura se representan todos aquellos residuos cuya contribución resultaba ser  $\geq |3|$  kcal/mol.



**Figura 66.** Representación molecular del reactivo señalando las interacciones (líneas discontinuas) en las que participan todos aquellos residuos para los que se calculó una contribución electrostática  $\geq |3|$  kcal/mol. Las líneas discontinuas se dibujan en diferentes colores: Arg76 (negro), Met283 (amarillo), Arg181 (verde), Tyr193 (rosado), Asp151 (azul), Asp245 (salmón), Asp246 (rojo), Arg135 (gris). En el caso de la Arg293, se utiliza el color marrón para indicar sus interacciones con los sustratos y el color celeste para indicar su interacción con el grupo carboxilato de GlcA ( $COO^-_{GlcA}$ ). Este código de color se relaciona con la Figura siguiente.

El análisis electrostático nos entrega valiosa información acerca de cómo la enzima contribuye a estabilizar las especies intermedias (en este caso el  $IP_B$ ), que se generan durante este mecanismo. Además, se han incluido dos términos adicionales que representan una estimación de la interacción electrostática entre Arg293 y Asp246 (color fucsia, Figura 67), y entre Arg293 y el  $COO^-_{GlcA}$  (color celeste, Figura 67).



**Figura 67.** Interacción electrostática calculada a nivel QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM para diferentes residuos del sitio activo en el intermedio  $IP_B$ . Los valores energéticos son relativos al reactivo y se han definido arbitrariamente negativos y positivos, según la interacción mejore o se deteriore con respecto al reactivo, respectivamente.

En la Figura 67, se observa que varios residuos estabilizan o interaccionan mejor con los sustratos en el  $IP_B$  que en el reactivo. Estos corresponden a la Arg76 (negro), la Met283 (amarillo), la Arg181 (verde), la Tyr193 (rosado), el Asp151 (azul), el Asp245 (salmón) y el Asp246 (rojo). La estabilización que estos residuos son capaces de proveer se relaciona por una parte con la carga negativa que se ha desarrollado en el grupo UDP: Arg76 y Met283. Por otro lado, hay un desarrollo de carga positiva en el centro anomérico que es acompañado por la distorsión conformacional de GalNAc. Los residuos que estabilizan esta especie son; el Asp151, a través de su interacción con  $HO3_{GalNAc}$ ; el Asp245, que interacciona con  $(HO4/HO6)_{GalNAc}$ ; y el Asp246, que ocupa parte de la cara  $\beta$  de GalNAc y cuya notable estabilización en el  $IP_B$  (-16.6 kcal/mol, ver Figura 67 color rojo), ha de relacionarse directamente con el aumento de carga en el centro anomérico y el mutuo acercamiento de los dos fragmentos. Este resultado indica el rol crucial del Asp246 en el mecanismo de ataque frontal, el cual es capaz de estabilizar las especies UDP----OC a través de una interacción de naturaleza electrostática (catálisis asistida por nucleófilo). Esto concuerda con los experimentos de mutagénesis en los cuales no se ha detectado actividad transferasa cuando el Asp246 es mutado por Ala.

Adicionalmente, Arg181 y Tyr193 son capaces de estabilizar moderadamente a lo largo de la reacción, lo cual muy probablemente tenga que ver con la interacción que forman con el  $COO^-_{GlcA}$ . Por otra parte, el deterioro de la interacción Arg135-sustratos al ir desde reactivo hasta  $IP_B$ , puede ser una consecuencia de la carga positiva adquirida por GalNAc.

Lo que ocurre con Arg293 es curioso y es por eso que hemos indagado un poco más profundamente en las interacciones que establece en el sitio activo. La Arg293 es un residuo altamente conservado dentro de la familia GT64<sup>285</sup> y se localiza en la cara  $\beta$  de GalNAc, en la vecindad del carbono anomérico. Los experimentos de mutagénesis han

mostrado que la enzima mantiene una actividad transferasa mínima (0.2 pmol/mg proteína/hora) luego de hacer la mutación Arg293Ala, lo que lleva a pensar que la Arg293 podría tener un rol importante en el mecanismo. El análisis de las estructuras cristalográficas (aquí se señalan los valores en 10M<sub>Z</sub>), llevó a los autores a sugerir que la red de interacciones que arginina está en posición de establecer con residuos como Asp246 ( $\text{NH2}_{\text{Arg293}}\text{-OD1}_{\text{Asp246}} = 3.16 \text{ \AA}$ ), Asn243 ( $\text{NE}_{\text{Arg293}}\text{-OD1}_{\text{Asn243}} = 3.35 \text{ \AA}$ ) y Tyr193 ( $\text{NH1}_{\text{Arg293}}\text{-OH}_{\text{Asn243}} = 3.24 \text{ \AA}$ ), podrían neutralizar parcialmente su carga positiva de modo de transformarla en una especie capaz de estabilizar la carga positiva que se desarrolla en GalNAc. El valor que obtenemos para la interacción electrostática de Arg293 (8.9 kcal/mol, Figura 67, color marrón) indica precisamente lo contrario, la interacción con ambos sustratos se ve perjudica con respecto al reactivo. Esto es comprensible dado el desarrollo de carga positiva sobre el carbono anomérico, y el ligero acercamiento entre  $(\text{C1-O5})_{\text{GalNAc}}$  y Arg293 luego de que GalNAc sufre el cambio conformacional. Por otra parte, dado que la Arg293 y el Asp246 interaccionan cercanamente, ha de existir una mutua polarización entre ellos. Por este motivo se repitió el análisis de cargas de modo que Arg293 pueda sentir el campo electrostático de Asp246 y viceversa. Los valores obtenidos siguen la misma tendencia, 11.6 para la Arg293 y -10.2 para el Asp246 (estos valores no se muestran en la Figura 67), y además sugieren que si Arg293 interacciona con la carga QM de Asp246 (y viceversa), el efecto de ambos residuos será menos favorable. Esto ha de ser una consecuencia de la reorientación de uno de los oxígenos del carboxilato de Asp246 (OD1) hacia el carbono anomérico. Este pequeño cambio no favorece la interacción Asp246-Arg293 ya que su interacción es mejor en el reactivo, cuando el carboxilato y el guanidinio caen en el mismo plano. En la Figura 67 se indica el valor del término electrostático entre ellos (color fucsia). También se ha calculado el término de interacción entre Arg293 y el  $\text{COO}^-_{\text{GlcA}}$ , el cual se ve desfavorecido en 2 kcal/mol y esto también tiene que ver con el pequeño giro del  $\text{COO}^-_{\text{GlcA}}$  a lo largo del camino de reacción (Figura 61), el cual provoca un ligero distanciamiento entre la Arg293 y el  $\text{COO}^-_{\text{GlcA}}$  (3.45 Å en R1<sub>A</sub> y 4.18 Å en el IP<sub>B</sub>, distancia  $\text{NH1}_{\text{Arg293}}\text{-O6B}_{\text{GlcA}}$ ).

De esta manera, los análisis electrostáticos realizados no permiten concluir un rol catalítico para la Arg293 y muestran que de hecho, el Asp246 puede estabilizar las especies UDP----OC una vez que se ha desarrollado la carga positiva sobre el centro anomérico  $(\text{C1-O5})_{\text{GalNAc}}$ , contrariamente a lo que Perderson *et. al.* habían sugerido en base a las estructuras cristalográficas. Creemos que el rol de la Arg293 puede tener más relación con la unión del sustrato aceptor, específicamente con en el reconocimiento del grupo  $\text{COO}^-_{\text{GlcA}}$ . Además, el grupo guanidinio de la Arg293 posiblemente permita que las cargas negativas de GlcA ( $\text{COO}^-$ ) y Asp246 se posicionen relativamente cerca en el sitio activo de EXTL2. Sin embargo, estas posibles funciones de la Arg293 no pueden ser concluidas en base al tipo de análisis que se acaba de explicar pues se llevan a cabo en el enzima nativa. Si Arg293 fuese mutada, probablemente se desencadenarían una serie de cambios estructurales que podrían afectar la reactividad de la enzima. Por ejemplo, si Arg293 no estuviese en el sitio activo podría verse afectada la unión o disposición relativa entre UDP-GalNAc y GlcA $\beta$ 1-3Gal, o la habilidad nucleofílica del Asp246, que al no interaccionar con la Arg293 sería un mejor nucleófilo que podría conducir al CGE.

Finalmente, es importante mencionar que la contribución conjunta Asp246-Arg293 en la cara  $\beta$  de GalNAc resulta en un valor de -7.7 kcal/mol (ver penúltima barra en la Figura 67), y de 1.7 kcal/mol (ver última barra en la Figura 67), cuando se omite o se tiene en cuenta,

respectivamente, la interacción entre ambos. Estos resultados sugieren que incluso en el peor caso, el término es desfavorable pero muy pequeño comparado con el término individual de la Arg293. De este modo, concluimos que el Asp246 amortigua el efecto desfavorable de tener un residuo con carga neta positiva en la cara  $\beta$  del azúcar, convirtiendo al par Asp246-Arg293 en una pareja eficiente catalíticamente.

## 5.4 Conclusiones

En este capítulo se ha abordado por primera vez el estudio del mecanismo de reacción de EXTL2 a un nivel atómico-molecular, combinando varias metodologías computacionales y las estructuras cristalográficas de EXTL2.

EXTL2 resulta ser un sistema bastante complejo para el cual ha sido difícil y laborioso modelar todas las interacciones que forman el complejo de Michaelis eficiente catalíticamente. Esta complejidad deriva en parte del hecho de que no se contaba con una estructura cristalográfica en la que estuvieran ambos sustratos, sino con dos estructuras a partir de las cuales no se podían anticipar un conjunto de interacciones inter-sustratos relevantes, con la excepción de  $\text{HO4}_{\text{GlcA}}\text{-O}_P$  que podía intuirse desde 1ON8. Nos referimos a las interacciones  $\text{HO3}_{\text{GlcA}}\text{-O1}\beta$  y  $\text{O3}_{\text{GlcA}}\text{-HN2}'_{\text{GalNAc}}$ , junto a las que se forman entre UDP y la esfera ( $\text{O2}\beta\text{-W}_{604}$ ), las cuales demostraron ser bastante influyentes en las barreras de energía obtenidas a partir de los primeros modelos ternarios obtenidos desde la MD MM. Todas estas interacciones se deben establecer óptimamente para la catálisis de modo de estabilizar la carga negativa creciente sobre UDP. Algunas de estas interacciones tampoco aparecieron en la MD, podría ser por deficiencias del campo de fuerzas.

Por otra parte, la presencia de un residuo nucleofílico en la cara  $\beta$  del azúcar ha sido durante años interpretada como evidencia en apoyo del mecanismo de doble desplazamiento a través del intermedio CGE, aunque esta idea se sustenta en una analogía histórica entre GTs-ret y GHs-ret, y no hay evidencia experimental concluyente al respecto. La enzima EXTL2 posee un Asp246 en la cara  $\beta$  de GalNAc a distancias  $> 5 \text{ \AA}$  en cualquiera de los reactivos. Nuestro estudio nos permite concluir que el Asp246 está relacionado con la estabilización de la especie oxocarbenio, cuya carga positiva está parcialmente estabilizada por deslocalización a lo largo del enlace  $\text{C1-O5}$  ( $sp_3 \rightarrow sp_2$ ), pero esta enzima parece requerir de una estabilización adicional por el efecto desfavorable de la Arg293 y porque la interacción  $\text{HO4}_{\text{GlcA}}\text{-O}_P$  no se ha formado en el complejo inicial. De este modo, el Asp246 resulta ser crucial en el mecanismo de ataque frontal ya que es capaz de estabilizar sustancialmente las especies OC a través de una interacción electrostática ( $\text{IP}_B$ ), sin necesidad de pasar por el intermedio covalente (CGE).

Tal como se comentó previamente, el análisis minucioso de la información estructural y energética obtenida desde las simulaciones del ataque frontal y el ataque del Asp246, nos permite notar que el Asp246 puede estabilizar la especie OC a partir de una distancia aproximada de  $4 \text{ \AA}$  (ejemplos:  $\text{OC1}_A$  y  $\text{OC1}_C$ ). Esta estabilización parece quedar ligeramente enmascarada en el mecanismo de ataque frontal (especies  $\text{OC}_B$ ), posiblemente por el impedimento estérico adicional que supone el hecho de que GlcA este acercándose al carbono anomérico más rápidamente que durante el ataque del Asp246. Curiosamente, en

el camino de más baja energía que logramos caracterizar para el mecanismo de ataque frontal (perfil B, naranja-verde), la distancias  $OD1_{Asp246^-} - C1_{GalNAc}$  son  $< 4.0 \text{ \AA}$  en todas las especies UDP----OC a diferencia de lo encontrado inicialmente en base al perfil A (perfil negro), en el que esta distancia siempre es  $> 4.0 \text{ \AA}$  hasta el  $IP_A$ . Basándonos en el hecho de que Asp246 puede estabilizar desde  $4 \text{ \AA}$ , creemos que el acercamiento adicional del Asp246 hasta  $C1_{GalNAc}$  puede relacionarse con algo más que la estabilización de la especie OC. Así, el mutuo acercamiento OC----Asp246 podría ocurrir para llenar la función que GlcA no puede llevar a cabo debido a la naturaleza intrínseca del mecanismo de ataque frontal, en el que el nucleófilo de la reacción (GlcA) no puede desplazar directamente al grupo saliente (UDP) porque esto tiene que ocurrir por el mismo espacio o cara de GalNAc (cara  $\alpha$ ), para conseguir mantener la configuración del carbono anomérico. Nuestra hipótesis, que ya fue propuesta por Lairson *et al.*<sup>45</sup> es que debido a que GlcA no puede llevar a cabo esta tarea por la alta tensión estérica que esto supondría, el Asp246 realiza esta función promoviendo así la salida de UDP a través del “empujón nucleofílico”. Y los resultados parecen confirmarlo. Así, la catálisis asistida por nucleófilo puede incluir además de la estabilización del oxocarbenio, un factor adicional de empuje nucleofílico para promover una especie UDP----OC más disociada.

Adicionalmente, nuestros análisis también indican que el Asp246 podría estar contrarrestando el efecto desfavorable de tener a la Arg293 en la cara  $\beta$  de GalNAc aproximadamente a  $3.0 \text{ \AA}$  del  $O5_{GalNAc}$  a lo largo del camino de reacción. Así, como Asp246 puede estabilizar a una distancia de  $4 \text{ \AA}$ , es esperable que la Arg293 pueda hacer lo contrario ubicándose a distancias pequeñas del centro anomérico ( $2.89 \text{ \AA}$  en el  $IP_B$ ). De hecho, así lo confirma el análisis electrostático, haciendo más lógica la idea de que el Asp246 puede contrarrestar su efecto desfavorable. Curiosamente, la mutación de Arg293 resulta en una actividad transferasa mínima, lo cual indica que la mutante no es capaz de catalizar la reacción ya sea por un cambio nefasto de la configuración del sitio activo o por problemas asociados a la unión de los sustratos, aunque lamentablemente no se ha informado una  $K_M$  para esta mutante. Con el trabajo realizado hasta ahora, se puede señalar que la importancia de Arg293 muy posiblemente se relacione con la unión, orientación y estabilización del sustrato aceptor a través del grupo  $COO^-_{GlcA}$ , ya que como se ve en la Figura de puntos estacionarios, el  $O6B_{GlcA}$  experimenta un distanciamiento de  $NH1_{Arg293}$  cuando pierde la interacción con  $HO4_{GlcA}$ , formándose un complejo de Michaelis  $R2_A$  casi  $4 \text{ kcal/mol}$  más alto en energía que  $R1_A$ .

Finalmente, como ya se resaltó en la introducción de esta tesis y en el inicio de cada capítulo, con la excepción de la pareja de residuos Arg135-Asp245 no existe una arquitectura conservada en la cara  $\beta$  del azúcar de las GTs-ret. El presente estudio aporta información importante respecto a la identidad de los residuos en esta cara  $\beta$ , que en EXTL2 es ocupada por la díada Asp246-Arg293. Estos residuos se relacionan íntimamente con la naturaleza química del sustrato aceptor y con la orientación que este adopta al unirse a la enzima, influyendo en la regio y estereoespecificidad de la transferencia.

Sin duda, más experimentos son necesarios para poder asignar mejor el rol de Arg293. Es por ello que en el próximo capítulo seguiremos investigando a la enzima EXTL2 pero utilizando su otro sustrato dador, UDP-GlcNAc. Además de adentrarnos en los detalles moleculares del mecanismo, construiremos mutantes para indagar profundamente en el rol de Arg293 y Asp246.





# 6 $\alpha$ 1,4-N-Acetilhexoaminiltransferasa: EXTL2-GlcNAc

## 6.1 Aspectos Generales

En este capítulo se investiga el mecanismo de reacción para el otro sustrato dador de EXTL2, el UDP-GlcNAc. Como se ha comentado previamente, el gen EXTL2 (del inglés EXT-like) presenta una altísima similitud con los genes EXT1 y EXT2 de la familia EXT, razón por la que es considerado otro de sus miembros. A través de análisis génicos, los locus que ocupan los genes de la familia EXT (EXT1-3) han sido relacionados con la aparición del desorden genético HME (Hereditary Multiple Exostoses), por lo que las enzimas que estos genes codifican suelen conocerse como supresoras de tumores.<sup>275</sup> Adicionalmente, la purificación y expresión de EXT2 desde suero bovino permitieron averiguar que esta enzima es activa para la transferencia de ácido D-glucurónico (GlcA) y N-acetil-D-glucosamina (GlcNAc), dos tipos de actividad transferasa que son necesarias en la biosíntesis del HS.<sup>276</sup> Tal como el gen EXT2, EXTL2 también podría codificar enzimas supresoras de tumores o estar involucrada en la biosíntesis del HS. Más tarde se comprobó que el gen EXTL2 efectivamente codifica una enzima que resulta poseer actividad transferasa para GalNAc y GlcNAc (unión  $\alpha$ 1-4).<sup>280</sup> Experimentos adicionales permitieron establecer que EXTL2 posee actividad transferasa de tipo GlcNAcT-I, lo cual llevó a los autores a postularla como una enzima clave en el inicio de la biosíntesis de HS.<sup>269</sup> Juntando esto con los resultados más recientes que sugieren que EXTL2 podría detener la polimerización de GAGs al transferir GlcNAc desde UDP a el tetrasacárido común de unión fosfatado en xilosa,<sup>277</sup> nos motiva a extender y a la vez, finalizar el estudio de esta enzima investigando los dos tipos de actividad transferasa que posee.

Aunque se trata de la misma enzima, para llevar a cabo este trabajo se tuvo que realizar un trabajo de modelización adicional debido al choque entre sustrato dador y sustrato aceptor que resultaba de superponer y alinear las correspondientes estructuras cristalográficas, códigos PDB 1ON6 y 1ON8, respectivamente. Esto se debe a que el fragmento GlcNAc aparece ligeramente rotado en la estructura cristalográfica con respecto a cómo estaba colocada GalNAc. En vista de estos primeros resultados se decidió llevar a cabo un estudio con cálculos de Docking. (Detalles en Metodología)

Debido a los roles que tiene EXTL2 en la biosíntesis de HS o H, existe un gran interés bioquímico por esta enzima. Adicionalmente, consideramos que EXTL2 representa un modelo paradigmático en la elucidación de las estrategias catalíticas que realizan las GTs que retienen la configuración del carbono anomérico del azúcar que transfieren. De esta manera, y con el objetivo de dar respuesta a los interrogantes mecanísticos que aún permanecen abiertos en torno a los residuos de la cara  $\beta$  del azúcar Asp246-Arg293, se ha llevado a cabo un estudio completo del mecanismo, caracterizando cada uno de los puntos estacionarios para tres diferentes complejos de Michaelis obtenidos de la combinación de

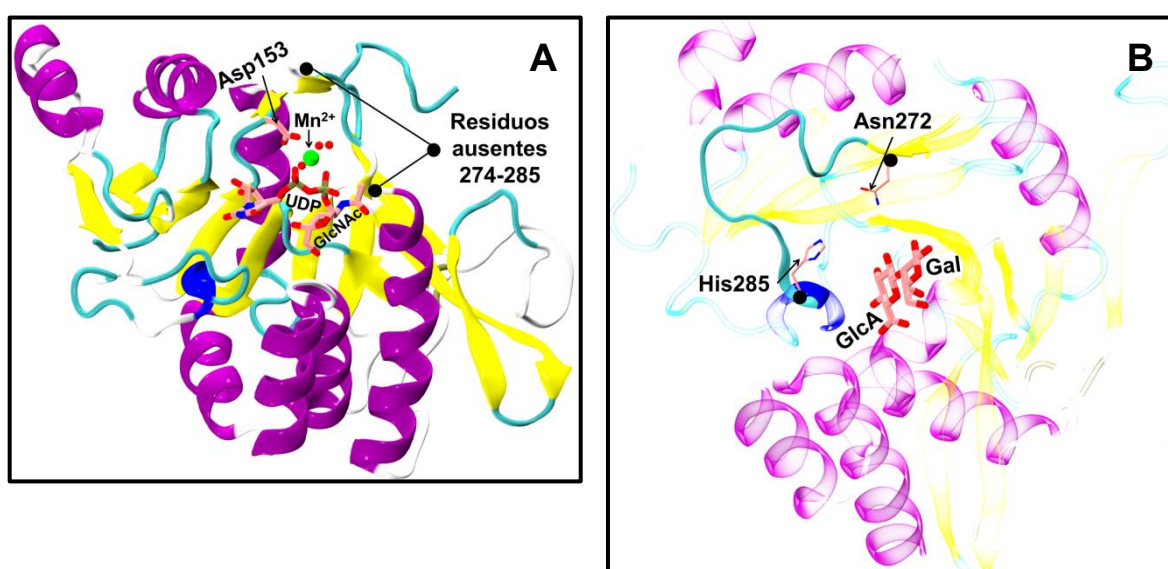
metodologías. Además, se han construido diferentes mutantes para uno de estos complejos con la idea de descifrar el rol mecánico de Arg293 y reevaluar los roles que fueron asignados al Asp246 en nuestro trabajo previo; esto es, evaluar el efecto de las diferentes mutaciones en la estrategia de empujón nucleofílico y en la capacidad de estabilizar el oxocarbenio (catálisis asistida por nucleófilo).

Los aspectos de estos trabajos pueden ser útiles en la ingeniería de enzimas aplicada a síntesis de glicanos, y esperamos que también sean un aporte relevante tanto en el campo del mecanismo catalítico de las GTs-ret, como en el interesante tópico de investigación que se desarrolla en torno a la biosíntesis del HS.

## 6.2 Modelos y Metodología

### 6.2.1 Construcción de los Modelos

El modelo ternario inicial, EXTL2·Mg<sup>2+</sup>·UDP-GlcNAc·GlcAβ1-3Gal, se construyó combinando dos diferentes estructuras cristalográficas. Una de ellas corresponde al complejo binario EXTL2·Mn<sup>2+</sup>·UDP-GlcNAc (código PDB 1ON6) y la otra corresponde al complejo ternario EXTL2·Mn<sup>2+</sup>·UDP-GlcAβ1-3Galβ1-Onaftalenmetanol (código PDB 1ON8), siendo el último un sustrato aceptor análogo del natural GlcAβ1-3Galβ1-Onaftalenmetanol. Las coordenadas del sustrato dador UDP-GlcNAc, el cofactor metálico y la proteína, con la excepción de un lazo sin coordenadas que abarca desde los residuos 274 al 285, se tomaron de la estructura o complejo 1ON6. Del complejo 1ON8 se tomaron las coordenadas del lazo recién mencionado y las del sustrato aceptor GlcAβ1-3Gal, cuyo sustituyente Onaftalenmetanol no tenía una densidad electrónica visible.<sup>70</sup>



**Figura 68.** Representación de la estructura secundaria de la EXTL2 para A: el complejo binario EXTL2·Mn<sup>2+</sup>·UDP-GlcNAc (código PDB 1ON6), en el que el sustrato dador UDP-GlcNAc y el Asp153 se dibujan en barras y el Mn<sup>2+</sup> y los átomos de oxígeno de las aguas de coordinación en esfera verde y rojas, respectivamente. Además, se indican los sitios entre los cuales no hay coordenadas (lazo 274-285, círculos negros). En B se representa la estructura secundaria del complejo ternario EXTL2·Mn<sup>2+</sup>·UDP-GlcAβ1-3Galβ1-Onaftalenmetanol (código PDB 1ON8), resaltando en color opaco las coordenadas que han sido usadas y en transparente aquellas que no se utilizaron. El sustrato aceptor y los residuos terminales del lazo perdido en 1ON6 se dibujan en barras, mientras que para el resto del lazo se muestra como una cinta celeste.

Como ya se mencionó en el capítulo anterior, la enzima muestra un plegamiento GT-A y ha sido cristalizada con un catión metálico, el Mn<sup>2+</sup> que es modelizado como Mg<sup>2+</sup>. El único aminoácido que coordina al metal es el Asp153, mientras que el resto de la esfera de coordinación está formada por el grupo difosfato y tres moléculas de agua; por efectos prácticos, se conserva la numeración previa de estas aguas (W<sub>604</sub>, W<sub>605</sub> y W<sub>606</sub>). El modelo ternario que se consigue de esta manera muestra choques directos entre los sustratos, específicamente entre el grupo N-acetilo (NAc) de GlcNAc y el C4 del ácido glurónico (GlcA); además, la distancia O4<sub>GlcA</sub>-C1<sub>GlcNAc</sub> tiene un valor de 1.98 Å, muy inferior a la suma de sus radios de Van der Waals que es 3.22 Å. Lo anterior da cuenta de que el modelo ternario que conseguimos construir al alinear 1ON6 y 1ON8 es altamente inestable y debido a esto, se decidió evaluar la posición relativa de los sustratos utilizando la metodología de docking. Para la realización de dichos cálculos contamos con la supervisión del Dr. Jean-Didier Maréchal (Universidad Autònoma de Barcelona).

## 6.2.2 Cálculos de Docking

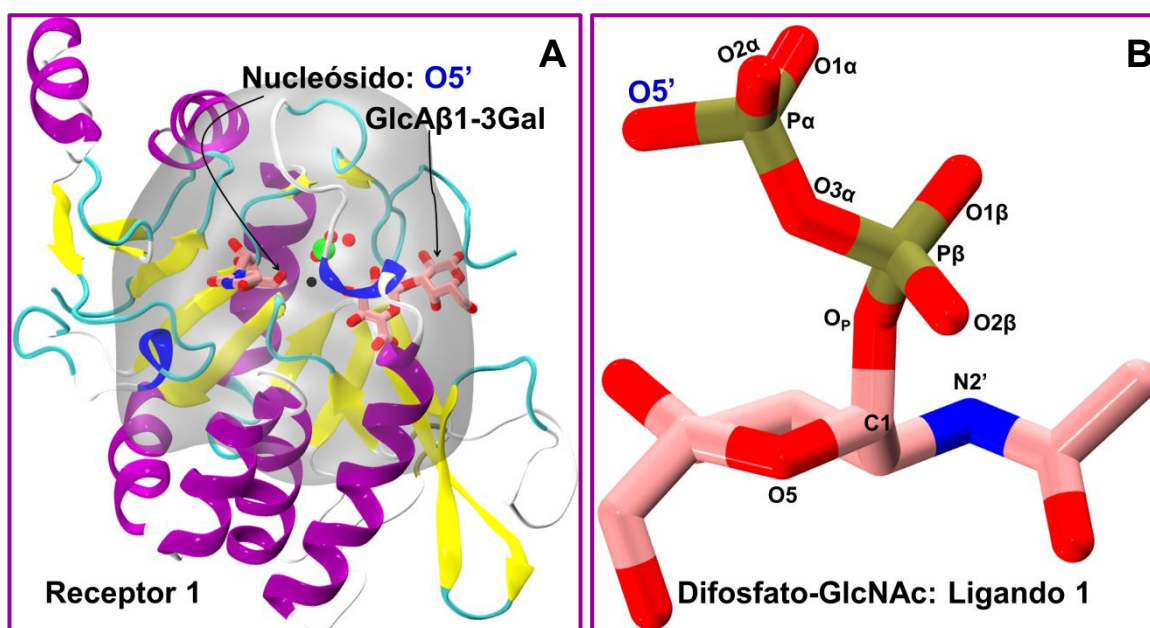
Para llevar a cabo estos cálculos se utilizó el programa GOLD (Genetic Optimisation for Ligand Docking),<sup>213,286</sup> que usa una estrategia evolutiva (algoritmo genético, GA) para explorar la variabilidad conformacional del ligando flexible mientras simultáneamente es capaz de considerar flexibilidad en la proteína, basándose para esto último, en librerías de rotámeros para los diferentes aminoácidos. Los buenos complejos enzima:ligando se definen en base a una función de puntuación que permite clasificar los modos de asociación obtenidos de acuerdo a la afinidad del ligando por el receptor. La afinidad es definida por interacciones como los enlaces de hidrógeno, interacciones hidrofóbicas y un término MM para calcular la energía interna del ligando.<sup>213</sup> A pesar de que nuestra primera inspección de los resultados estuvo centrada en las puntuaciones de los ligandos, los análisis posteriores priorizaron las posiciones de los ligandos y las diversas interacciones que estos pueden establecer desde aquellas posiciones. Este segundo criterio pudo ser aplicado gracias al conocimiento adquirido previamente del complejo de Michaelis que fue obtenido combinando estructuras cristalográficas junto a metodologías MM y QMMM en el Capítulo 5. De esta manera, la elección de soluciones de docking estuvo principalmente guiada por nuestro conocimiento previo de las interacciones entre sustratos y proteína en EXTL2-GalNAc (se obvia el sustrato aceptor y el UDP pues no cambia).

Aunque se realizaron dockings covalentes y no covalentes del sustrato dador y aceptor, sólo se describirán los resultados de los dockings covalentes del sustrato dador UDP-GlcNAc.

En éstos, se utilizó como receptor la proteína junto al sustrato aceptor, incluyendo el metal y su primera esfera de coordinación, y parte de UDP (ver a continuación).

### **6.3.2.1 Docking Covalente de difosfato-GlcNAc en el Complejo Ternario 1ON8**

La preparación de las estructuras de partida para el docking fue iniciada con el receptor, que corresponde al complejo EXTL2·Mn<sup>2+</sup>·UDP·GlcAβ1-3Gal (código PDB 1ON8). Todas las aguas, iones y el difosfato fueron eliminados de la estructura, con la excepción del metal y sus aguas de coordinación. A los aminoácidos ionizables de la proteína se le asignaron estados de protonación apropiados. Además, se indicó el tipo de coordinación del metal, que en este caso es de tipo octaédrica y finalmente se añadieron hidrógenos para terminar la preparación del receptor. En el caso del ligando sólo se añaden hidrógenos. El centro de la cavidad de asociación se colocó en el átomo O3α (origen, ver Figura 69), y el radio de la cavidad es de 12 Å. Ambos, el receptor y el ligando contienen el átomo O5' que forma parte del nucleósido uridina (molécula en la que uracilo y ribofuranosa se enlazan a través de un enlace glicosídico β1-N1), ya que el enlace se formará entre el nucleósido (O5') y el fragmento difosfato-GlcNAc (Pα). Automáticamente después de definir la esfera de asociación, todos los dadores de enlace de hidrógeno, así como átomos dadores y aceptores de hidrógeno cercanos a la cavidad, se identifican a través de los tipos de átomos SYBYL y se añaden pares de electrones no compartidos a los aceptores de hidrógeno. Si cualquiera de los dadores o aceptores estuvieran unidos por enlaces simples a un residuo(s) de la proteína, éstos se establecen como flexibles.<sup>213</sup> Por otra parte, se establecieron 4 residuos flexibles involucrados en la asociación del sustrato aceptor: Tyr193, Arg181, Trp284 y Arg293; aunque los dos últimos también forman interacciones con el sustrato dador a través del difosfato y el azúcar, respectivamente. Al hacerse el docking del fragmento difosfato-GlcNAc, consideramos adecuado añadir flexibilidad a estos residuos.

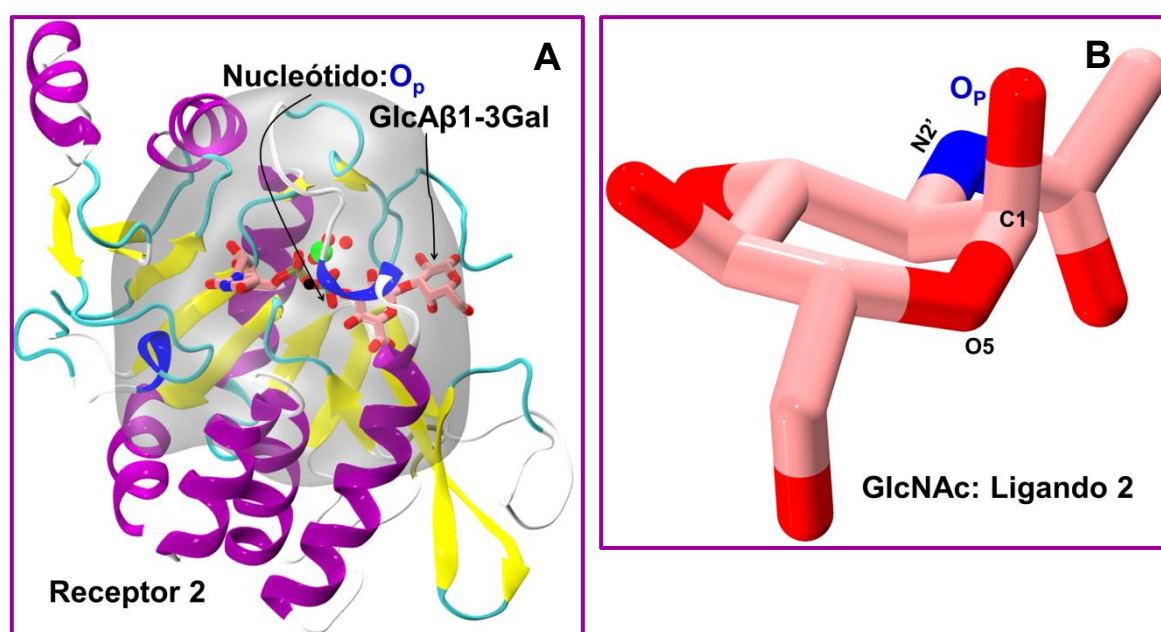


**Figura 69.** A. Estructura secundaria del complejo ternario 1ON8 que corresponde al receptor 1; la esfera verde corresponde al metal y las tres esferas rojas representan las moléculas de agua de coordinación; la zona de color negro indica el centro de la cavidad de asociación en la que se realizará el docking y la zona gris transparente tiene un radio de 12 Å con respecto a al centro de la cavidad (O3 $\alpha$ ), y se dibuja para representar aproximadamente la cavidad del docking; el sustrato aceptor GlcA $\beta$ 1-3Gal y el nucleósido se dibujan en barras de color rosa (carbonos), azul (nitrógeno) y rojo (oxígenos); además, se indica la posición de O5' (azul) y por claridad, los hidrógenos no se representan. B, El ligando 1 o difosfato-GlcNAc se representa en barras y se indican los nombres de algunos átomos, como por ejemplo, el O5'.

De un total de 20 modelos (soluciones del docking) evaluados, un número adecuado teniendo en cuenta que se trata de un algoritmo estocástico, sólo 4 parecen ser adecuados a pesar de que HO4<sub>GlcA</sub> no se orienta hacia O<sub>p</sub> sino que a O2 $\beta$ . Esta no es una buena interacción en el contexto de la reactividad pues el O<sub>p</sub> queda parcialmente ocluido por O2 $\beta$ . Es probable que el detalle en la definición de las interacciones que precisamos para la catálisis sea excesivo para la capacidad de predicción del programa de docking, por lo menos para el sistema estudiado. Además, los residuos involucrados en la asociación o unión del aceptor no se colocan en las orientaciones relativas más adecuadas y se sabe muy bien que todas estas interacciones escasamente óptimas tienen un efecto nefasto en las barreras de energía de la reacción química.

### 6.2.2.2 Docking Covalente de GlcNAc en un Complejo Ternario Catalítico

En este experimento se establece la cavidad de asociación de la misma manera que antes, definiendo el origen en las coordenadas del átomo  $O3\alpha$  y un radio 12 Å para la esfera (Figura 70). Esta vez, el enlace se formará entre  $C1_{GlcNAc}-O_P$  con el objetivo de contar con las interacciones entre el UDP y el GlcA, o el agua de coordinación que hemos identificado anteriormente como importantes para la catálisis, se utilizó como receptor la estructura generada en el estudio de EXTL2:GalNAc (Capítulo 5).



**Figura 70.** A. Se muestra la estructura secundaria de 1ON8 que ahora corresponde al receptor 2 pues no contiene sólo el nucleósido sino que todo el nucleótido. Igual que antes; la esfera verde y las rojas representan al metal y las tres moléculas de agua de coordinación, respectivamente; la esfera negra ( $O3\alpha$ ) señala el centro de la cavidad de asociación a partir de la cual se dibuja una zona gris transparente que representa aproximadamente el tamaño de la cavidad del docking; el sustrato aceptor  $GlcA\beta1-3Gal$  y el nucleótido se dibujan en barras; y por último, se indica la posición de  $O_P$  (azul) y por claridad los hidrógenos no se representan. B, El ligando 2 o GlcNAc se representa en barras y se indican los nombres de algunos átomos, como por ejemplo, el  $O_P$  en azul.

Además, a la luz de los resultados anteriores en los que Arg293 no adoptó posiciones adecuadas con respecto a GlcNAc, se decidió establecer restricciones en la distancia  $NH1_{Arg293}-O5_{GlcNAc}$ . Así, las soluciones del docking se reducen a aquellas en las que se cumpla la condición de que  $4.0 \text{ \AA} \leq NH1_{Arg293}-O5_{GlcNAc} \leq 4.5 \text{ \AA}$ . Alentadoramente, de un total de 20 modelos evaluados, uno de ellos mostró interacciones semejantes a las observadas en el complejo de Michaelis competente y fue seleccionado como modelo ternario de partida para las simulaciones de MD y los posteriores estudios de reactividad.

## 6.2.3 Simulaciones de Dinámica Molecular

### 6.2.3.1 Minimizaciones y Dinámica Molecular MM

Utilizando el modelo ternario explicado arriba, se realizó un procedimiento similar al ya descrito para EXTL2-GalNAc (apartado 5.2.2). La estructura se modificó con el servidor PDB2PQR versión 2.0.0,<sup>225,226</sup> sólo para establecer los tipos de átomos que usa el campo de fuerzas CHARMM.

Como ya se ha explicado anteriormente, una vez asignada las topologías y parámetros, el sistema es solvatado en una caja de aguas de dimensiones 80·75·87 Å<sup>3</sup> y neutralizado con 3 cationes Na<sup>+</sup> formando un sistema final de 48560 átomos. A continuación, éste fue minimizado y equilibrado a través de simulaciones sucesivas en las que se restringió el movimiento de grupos de átomos con potenciales armónicos gradualmente reducidos hasta cero. Mientras las aguas y cationes Na<sup>+</sup> se movían libremente, la proteína, el cofactor y los sustratos se desplazaban alrededor de sus posiciones iniciales hasta quitar las restricciones armónicas permitiendo la evolución libre del sistema. La etapa de equilibración tardó 5.4 ns y en el período final de 2.5 ns, en el que la proteína se equilibra libremente, el RMSD del segmento base de la proteína tiene un valor promedio de 1.47 Å. Luego, se llevaron a cabo 10 ns más de DM completando una simulación de 15.4 ns, lo que consideramos suficientemente largo para asegurar que nuestro modelo es estable.

Las simulaciones se llevaron a cabo con el programa NAMD2.9<sup>238,239</sup> y se establecieron condiciones periódicas de contorno. La presión y la temperatura se establecieron en 1 bar y 300 K, respectivamente. Para mantener la presión constante se aplicó un método que combina el método Nosé-Hoover<sup>200</sup> y el pistón de Langevin<sup>199</sup>, mientras que la temperatura se controló a través de fuerza adicionales incluidas en la dinámica de Langevin. El algoritmo SHAKE<sup>193</sup> fue usado para mantener fijas las distancias O-H de las aguas y el paso de integración de las ecuaciones de movimiento es de 2 fs. Una vez finalizada la simulación, se realizó un análisis de las estructuras seleccionando varios complejos de Michaelis que fueron usados como puntos iniciales para las primeras exploraciones del mecanismo de reacción.



### 6.2.3.2 Simulaciones de Dinámica Molecular QM/MM

Los complejos de Michaelis obtenidos en la etapa previa mostraban prácticamente las mismas interacciones que se encontraron en los primeros complejos de Michaelis de EXTL2-GalNAc (se omite el UDP y el sustrato aceptor). La mayoría de las estructuras muestreadas tienen al grupo NAc interaccionando con el Asp151 y no con GlcA, provocando que HO<sub>4</sub><sup>GlcA</sup> interaccione con O7' del grupo NAc. Además, la molécula de agua de coordinación ( $W_{604}$ ) aparece invariablemente interaccionando con Ser281 pero no con el grupo fosfato. La exploración del mecanismo de ataque frontal a nivel QM(BP86/SVP)/MM(CHARMM) resultó en barreras de energía potencial >37 kcal/mol, en algunos casos superando incluso las 60 kcal/mol cuando el grupo COO<sup>-</sup><sub>GlcA</sub> no establecía interacciones con Arg181, Arg293 y Tyr193.

A la luz de estos de estos resultados, y teniendo en cuenta que es una metodología que antes ha ayudado en la búsqueda del conjunto de interacciones óptimas para la catálisis, se decidió llevar a cabo simulaciones de MD QMMM de los complejos enzima-producto (EXTL2·Mg<sup>2+</sup>·UDP·GlcNAc $\alpha$ 1-4GlcA $\beta$ 1-3Gal) obtenidos previamente para dos de nuestros complejos (los que mostraron las barreras energéticas más bajas). Esto nos permite refinar las interacciones del producto, lo que luego se reflejará en el nuevo reactivo obtenido. Estas simulaciones adicionales de 60 ps fueron llevadas a cabo con el método QM(SCC-DFTB)/MM(CHARMM), el cual a pesar de ser conocido por sobreestimar las barreras de disociación UDP-azúcar, ha resultado útil anteriormente en la etapa de refinamiento de interacciones en el producto. En base a estas simulaciones se seleccionaron 6 complejos (sección 6.3) que se usaron como puntos iniciales para simular la reacción química inversa, consiguiendo un nuevo conjunto de complejos de Michaelis que serán los reactivos en las exploraciones del mecanismo de reacción posteriores. Los detalles de la metodología QMMM se expondrán en los siguientes apartados.

### 6.2.4 Simulaciones QM/MM

Para investigar el mecanismo de reacción de EXTL2 se seleccionaron todas las aguas que dentro de una esfera con radio de 30 Å desde el carbono anomérico C1<sub>GlcNAc</sub> y se incluyó toda la proteína, formando un modelo de 12476 átomos.



como región QM y MM. Por último, las flechas moradas señalan el movimiento de los electrones durante el mecanismo de ataque frontal.

La carga neta de la región QM es -2 u.a. y para completar la valencia de los átomos QM en los que se definió el corte con la región MM, se añadieron 6 *link atoms*.<sup>241,242</sup> Para reasignar la carga de los átomos en la interfaz QM/MM se usó el esquema *charge shift* (o desplazamiento de carga).<sup>179,243</sup> El cálculo de las interacciones electrostáticas se trató con el esquema *electronic embedding* (o integración electrostática)<sup>243</sup> y no se definió una distancia de corte para el cálculo de las interacciones no covalentes MM y QM/MM. En las optimizaciones se consideraron todas las aguas y residuos con al menos un átomo dentro de un radio de 15 Å desde el C1<sub>GlcNAc</sub> mientras que el resto del sistema se mantuvo fijo.

#### 6.2.4.1 Mecanismo de Ataque Frontal

Las primeras simulaciones del mecanismo de reacción utilizaron la coordenada de reacción (RC), definida como  $RC_A = [d(O_P-C1_{GlcNAc}) - d(O4_{GlcA}-C1_{GlcNAc}) - d(HO4_{GlcA}-O_P)]$ . La combinación de estas distancias permite conducir el sistema desde reactivos a productos, lo cual se lleva a cabo en pasos de 0.1 Å. En las primeras exploraciones, la región QM se describió a nivel BP86/SVP pero una vez comprobado que las barreras son razonables (< 28 kcal/mol), se llevaron a cabo los estudios al nivel QM BP86/TZVP y estos son los resultados que se discutirán más adelante. La región MM es descrita con el campo de fuerzas CHARMM. También se recalcularon las energías de todas las estructuras del perfil de reacción utilizando un nivel de teoría más sofisticado, QM(M05-2X/TZVP)/MM(CHARMM).

Basándonos en la exploración del mecanismo calculada a través de  $RC_A$  y a nivel QM(BP86<sup>152,154,244-246</sup>/TZVP<sup>248</sup>)/MM(CHARMM), se seleccionaron varias estructuras intermedias que se minimizaron en búsqueda del intermedio IP. En los casos en que las estructuras minimizadas conservaron sus características de intermedio (muchas colapsan a reactivo o producto), se continuó el estudio del mecanismo utilizando el IP como punto de partida para ir hacia reactivo y producto. La RC utilizada para ir hacia reactivo se define de la siguiente manera,  $RC_B = [d(O4_{GlcA}-C1_{GlcNAc}) - d(O_P-C1_{GlcNAc})]$ , mientras que la RC para guiar el sistema a productos es,  $RC_{AB} = [d(O_P-C1_{GlcNAc}) - d(O4_{GlcA}-C1_{GlcNAc}) - d(HO4_{GlcA}-O_P)]$ . A pesar de que  $RC_{AB}$  contiene la misma combinación de distancias que  $RC_A$ , se denotan con un subíndice distinto para indicar que ambas RC se inician desde configuraciones diferentes del sistema. En cada uno de los caminos obtenidos se buscaron estados de transición y se calcularon los modos normales de vibración.

#### 6.2.4.2 Mecanismo de Ataque del Asp246

De la misma manera que para el mecanismo de ataque frontal, las primeras exploraciones utilizaron el método QM(BP86/SVP<sup>247</sup>), pero después todos los cálculos se repiten al nivel QM(BP86/TZVP) y éstos son los resultados que se describirán más adelante en este

capítulo. Para llevar a cabo esta exploración se utilizaron dos coordenadas de reacción diferentes; en un principio se usó una RC triple  $RC_{DA} = [d(O_P-C1_{GlcNAc}) - d(OD1_{Asp246}-C1_{GlcNAc}) - d(HO4_{GlcA}-O_P)]$ , que luego se comparó con los resultados que da una RC doble  $RC_D = [d(O_P-C1_{GlcNAc}) - d(OD1_{Asp246}-C1_{GlcNAc})]$ , encontrándose que la  $RC_D$  es suficiente para el proceso que se intenta simular. Esto tiene que ver principalmente con el hecho de que  $HO4_{GlcA}$  está orientado hacia  $O_P$  (a diferencia de lo observado en EXTL2-GalNAc), además de que en este mecanismo UDP no actuará como base catalítica deprotonando a GlcA y por ende, ya no es necesario definirlo en la RC. En este caso, también se buscaron intermedios (IP) que se usaron como puntos de partida para ir hacia CGE, pero estas exploraciones adicionales dieron los mismos resultados que al utilizar  $RC_D$ . También se intentaron localizar estados de transición usando varias estructuras asociadas al proceso de disociación UDP-GlcNAc y al ataque de Asp246; las estructuras obtenidas se usaron para calcular los modos normales de vibración. Finalmente, en la sección de resultados se informan los datos obtenidos a partir de la  $RC_D$ .

Los cálculos QMMM fueron hechos con el programa modular ChemShell,<sup>181,250</sup> igual que en los capítulos anteriores.

### 6.2.5 Cálculo de Contribuciones Electroestáticas, Cargas Naturales e Interacciones de Orbitales Naturales

Estos cálculos se llevan a cabo a nivel QM(M05-2X<sup>166</sup>/TZVP)/MM/QM(BP86/TZVP)/MM en cada uno de los puntos estacionarios de los 3 perfiles de reacción finalmente considerados. Los cálculos electrostáticos y los de cargas y orbitales de enlace naturales, se llevaron a cabo de igual manera que lo descrito anteriormente.

### 6.2.6 Mutantes *in silico*

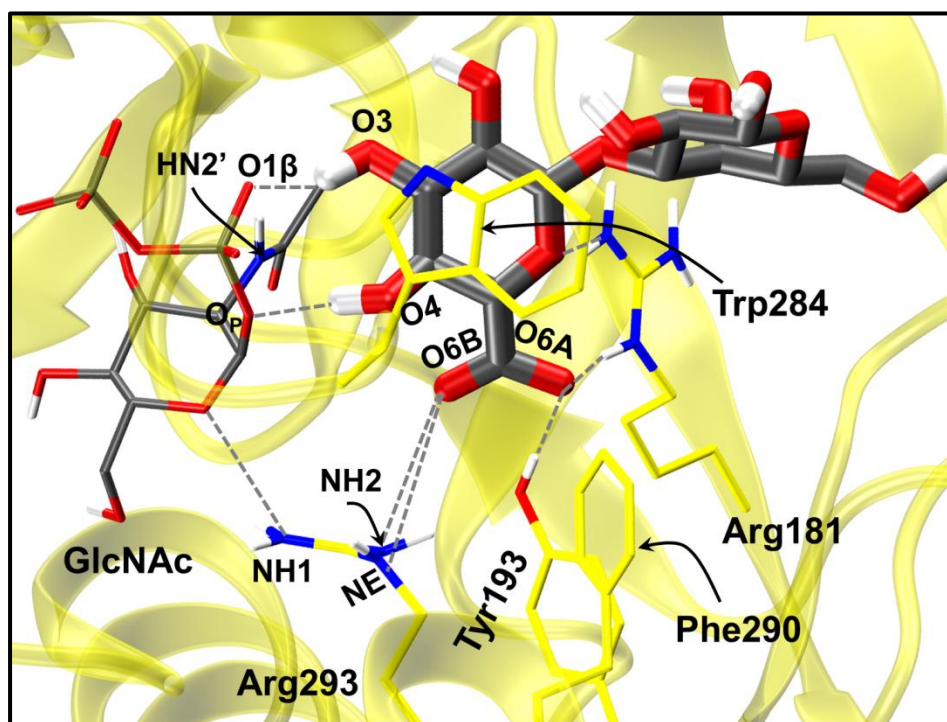
Las coordenadas iniciales en base a las cuales se construyeron las enzimas mutantes corresponden a las del complejo de Michaelis R-495 (ver Resultados). Para generar las 4 mutantes: Asp246Ala, Asp246Glu, Arg293Ala y Asp246Ala/Arg293Ala, se utilizó el *plugin Mutator* del programa VMD.<sup>240</sup>

### 6.2.7 Simulación de Umbrella Samplig

Simulaciones de MD empleando la técnica de *Umbrella Sampling* (US) y el nivel QM(SCC-DFTB<sup>172,173</sup>)/MM(CHARMM), se llevaron a cabo de modo análogo al descrito en el capítulo anterior. En este caso, se obtuvieron 61 estructuras (ventanas) a partir del camino de reacción explorado en pasos de 0.1 Å. Aquí, en cada ventana el tiempo de equilibración fue de 10 ps, seguido de una MD de producción de 50 ps, partir de la cual se realizó el muestreo para el cálculo del PMF. Finalmente, para calcular el PMF se utilizó el método *Umbrella Integration* (UI).

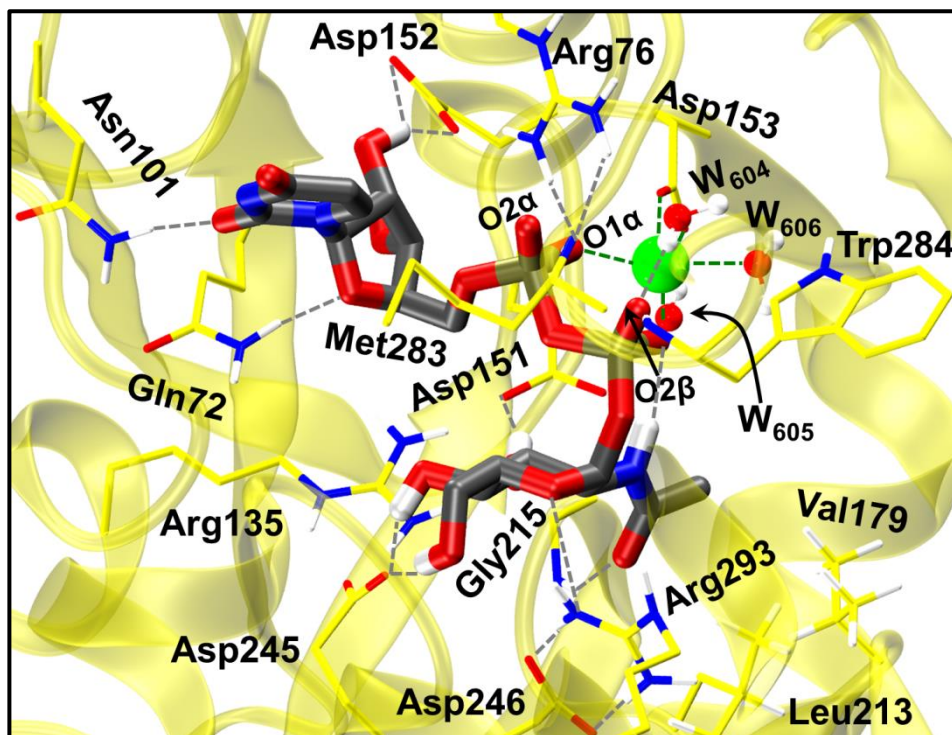
### 6.3 Resultados y Discusión

Esta sección se inicia con la descripción de los residuos involucrados en las interacciones de los sustratos (residuos de interacción). El modo de interactuar entre sustratos y proteína es básicamente el mismo que el encontrado para EXTL2-GalNAc, pero con algunas pequeñas diferencias que se comentarán a continuación.



**Figura 73.** Representación del sustrato aceptador GlcA $\beta$ 1-3Gal (barras), indicando con líneas discontinuas grises todas las interacciones en las que éste participa. Los residuos se dibujan en líneas (carbonos amarillos), al igual que la parte del sustrato dador que interacciona con GlcA.

El modo de interacción del sustrato aceptador es muy similar al ya descrito para EXTL2-GalNAc; la Phe290 ocupa un espacio muy cercano a la Tyr193 y el GlcA; la Tyr193 y la Arg181 interactúan con COO<sup>-</sup><sub>GlcA</sub> a través del O6A y a la vez, Arg181 también interactúa con O5 de GlcA. La Arg293 está a distancia de interacción del grupo COO<sup>-</sup><sub>GlcA</sub>, posicionándose a < 4 Å del O6B y además, interactúa con el O5<sub>GlcNAc</sub>. Las interacciones inter-sustratos resultan ser ligeramente distintas, observándose que HO4<sub>GlcA</sub> se orienta hacia O<sub>P</sub> y que HN2'<sub>GlcNAc</sub> interactúa con O1 $\beta$ .



**Figura 74.** Se representa el sustrato dador UDP-GlcNAc y sus residuos de unión a través de: barras (carbonos grises) y líneas (carbonos amarillos), respectivamente. Para representar el  $Mg^{2+}$  y las moléculas de agua de coordinación se utilizan esferas.

Como puede observarse en la Figura 74, el grupo UDP forma varias interacciones con residuos neutros y cargados. Los residuos que interaccionan con el uracilo son: Asn101, Gln72 y Asp152, exactamente lo mismo que se encontró para EXTL2-GalNAc. En el caso de UDP, se observan algunas diferencias relacionadas con la Met283 y la  $W_{604}$ ; en este complejo (EXTL2-GlcNAc), la Met283 interacciona con  $O2\alpha$  y no con  $O3\alpha$ , lo que puede tener relación con el hecho de que la Tyr280 aparece más alejada del UDP y ya no interacciona con  $O2\alpha$ , a diferencia de lo que se encontró en EXTL2-GalNAc. Ambos residuos, Tyr280 y Met283, pertenecen al lazo que no tiene coordenadas en los complejos binarios 1ON6 y 1OMZ, por lo que no parece extraño que aparezcan sutiles variaciones en sus posiciones. Sin embargo, a pesar de que la Tyr280 ya no interacciona con  $O2\alpha$ , la Met283 se desplaza ligeramente para llenar esa interacción, de modo que a nivel catalítico no se observan consecuencias. En el caso de  $W_{604}$ , ésta aparece interaccionando sólo con  $O2\beta$  y no con  $O2\alpha$ . Nuevamente, modelizar esta interacción requirió de estudios de reactividad pues en los complejos de Michaelis obtenidos de las MD MM, esta interacción nunca fue observada.

Otra de las diferencias encontradas con respecto al modelo EXTL2-GalNAc, tiene que ver con el grupo NAc de GlcNAc. En la Figura 74 se puede ver que  $HN2'$  apunta hacia  $O1\beta$  y no hacia el sustrato aceptor. Es difícil encontrar la explicación de este suave cambio en el patrón de interacciones, aunque se puede adelantar que  $HN2'_{GlcNAc}$  y  $O3_{GlcA}$  están lo suficientemente cerca como para recuperar su interacción durante la reacción de transferencia. En cualquier caso, la Gly215 sigue interaccionando con  $O7'_{GlcNAc}$  al igual que Val179 y Leu213 lo hacen con el grupo NAc.

En cuanto al fragmento Glc, el Asp151 interacciona con HO3<sub>GlcNAc</sub>; la Arg135 interacciona con O3<sub>GlcNAc</sub> y con Asp245; el Asp245 interacciona con HO4/HO6 de la Glucosa. En este caso, el modo de interacción es igual al encontrado en EXTL2-GalNAc. Cabe recordar (ver subapartado 1.3.3.2 y el capítulo de EXTL2-GalNAc, después de la Figura 57), que Asp245, Arg135 y las interacciones que éstos pueden formar corresponden a las pocas características estructurales conservadas en la cara  $\beta$  de las GTs-ret (GT-A).<sup>45</sup> En las discusiones que se harán más adelante en esta sección, se analizarán las implicancias de los residuos que se ubican esta región no conservada de la cara  $\beta$  de GlcNAc. Como ya se ha comentado previamente, en EXTL2 se han identificado dos residuos en esta cara  $\beta$ ; uno de ellos es la Arg293, que puede establecer interacciones con O5<sub>GlcNAc</sub> ( 3.0 Å en cada monómero de 1ON6), y con el grupo carboxilato de GlcA (NH2<sub>Arg293</sub>-O6B<sub>GlcA</sub> 3.3 Å en ambos monómeros de 1ON8). Por último, la Arg293 también interacciona con el Asp246, ambos con sus cadenas laterales dispuestas en planos perpendiculares tanto en 1ON6 y 1ON8. Por otra parte, el Asp246 no interacciona con ningún grupo hidroxilo del azúcar pero sí establece interacciones con Arg293. En 1ON6, el Asp246 se encuentra a una distancia de 5.10 Å del C1<sub>GlcNAc</sub>; 5.13 Å  $\pm$  0.45 Å en la MD de producción y 4.49 Å  $\pm$  0.06 Å en las estructuras de reactivos minimizadas a nivel QMMM. Dado que el Asp246 se encuentra un poco más cerca de C1<sub>GlcNAc</sub>, y teniendo en cuenta los resultados de EXTL2-GalNAc, también se explorará el ataque del Asp246.

Respecto a la Arg293, los resultados de la MD muestran que ésta se encuentra a una distancia promedio de 3.98 Å  $\pm$  0.26 Å del C1<sub>GlcNAc</sub> y 3.35 Å  $\pm$  0.22 Å del O5<sub>GlcNAc</sub>, respectivamente. En los reactivos, las distancias promedio con respecto al C1<sub>GlcNAc</sub> y el O5<sub>GlcNAc</sub> son 3.83 Å  $\pm$  0.04 Å y 3.24 Å  $\pm$  0.06 Å, respectivamente, valores ligeramente más pequeños que los encontrados en la MD pero también un poco más grandes que los observados en el complejo binario 1ON6 (cadena A), que son 3.20 Å y 3.00 Å para el C1<sub>GlcNAc</sub> y el O5<sub>GlcNAc</sub>, respectivamente. Aunque desde la perspectiva de las interacciones de GlcA la presencia de la Arg293 es comprensible, desde el punto de vista mecanístico sigue resultando intrigante tener un residuo con carga neta positiva ubicado a poco menos de 4 Å del centro anomérico. El Asp246 y la Arg293 son residuos esencialmente conservados dentro de la familia GT64,<sup>285</sup> por lo que uno de los objetivos de este trabajo es averiguar si existe un rol mecanístico para Arg293 y por otro lado, profundizar en las características de la catálisis asistida por nucleófilo que encontramos para el Asp246. Para lograr esto, se llevaron a cabo varios experimentos computacionales en los que se diseñaron los mutantes Asp246Ala, Asp246Glu, Arg293Ala y el doble mutante Asp246Ala/Arg293Ala, los cuales se estudiaron combinando simulaciones clásicas y QMMM.

### 6.3.1 Mecanismo de Ataque Frontal

Como se ha explicado en el apartado 6.2.3.2, seis estructuras del complejo E:S se seleccionaron para los estudios de reactividad, las cuales están en acuerdo con las barreras de energía potencial que suelen mostrar las GTs-ret. Como son exploraciones iniciales del mecanismo, no se han buscado estados de transición, sin embargo son útiles para reconocer las interacciones que más impacto tienen sobre la barrera energética de los procesos de ruptura UDP-GlcNAc y formación de la especie UDP----OC (OC, como antes, indica oxocarbenio), así como del ataque nucleofílico de GlcA.

**Tabla 14.** Resumen de las exploraciones de energía potencial (kcal/mol) a partir de un subconjunto de reactivos minimizados a nivel QM(BP86/TZVP)/MM(CHARM). Se indican las energías relativas correspondientes a: los puntos de energía más altos del camino de reacción ( $\Delta E^\ddagger$ ), estructura no optimizada; energías de reacción ( $\Delta E$ ), las estructuras de productos han sido optimizadas. El color con el que se escribe cada modelo o reactivo se correlaciona con las figuras que se representan más adelante.

| Reactivo     | QM(BP86/TZVP)       |            | QM(M05-2X/TZVP//BP86/TZVP) |            |
|--------------|---------------------|------------|----------------------------|------------|
|              | $\Delta E^\ddagger$ | $\Delta E$ | $\Delta E^\ddagger$        | $\Delta E$ |
| <b>R-175</b> | 6.64                | -13.79     | 13.23                      | -14.39     |
| <b>R-208</b> | 18.60               | -18.86     | 25.18                      | -18.12     |
| <b>R-488</b> | 18.20               | 9.15       | 24.90                      | 9.12       |
| <b>R-495</b> | 11.77               | -2.52      | 17.48                      | -3.78      |
| <b>R-350</b> | 18.52               | 2.72       | 23.78                      | 5.14       |
| <b>R-609</b> | 21.85               | 9.67       | 30.15                      | 11.02      |

Los valores de energía son diversos y esto se debe a sutiles variaciones en las interacciones inter-sustratos y enzima-sustratos. Tales diferencias nos han permitido identificar un conjunto de interacciones óptimas en el complejo de Michaelis que coinciden con las encontradas en el primer estudio de esta enzima. Entre tales interacciones destacan por ejemplo, las que se establecen entre los sustratos como:  $\text{HN2}'_{\text{GlcNAc}}-\text{O3}_{\text{GlcA}}$ ,  $\text{O1}\beta\text{-HO3}_{\text{GlcA}}$  y  $\text{O}_\text{P}\text{-HO4}_{\text{GlcA}}$ . En cuanto a las interacciones enzima-sustratos se tienen:  $\text{HH}_{\text{Tyr193}}-\text{O6A}_{\text{GlcA}}$ ,  $\text{NE}/\text{NH2}_{\text{Arg293}}-\text{O6B}_{\text{GlcA}}$ ,  $\text{OD1}/\text{OD2}_{\text{Asp246}}-\text{C1}_{\text{GlcNAc}}$  y  $\text{W}_{604}-\text{O2}\beta$ .



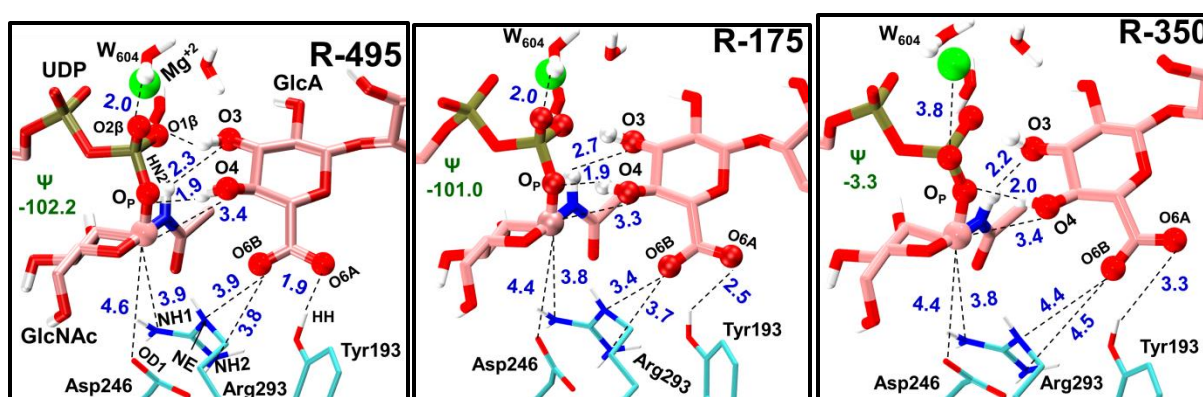






reacción se usaron para buscar IPs. La elección de estructuras de partida para buscar el IP se basa en un criterio geométrico en el que la distancia  $C1_{\text{GlcNAc}}-O_P \geq 2.9 \text{ \AA}$ . Esto se debe a que los perfiles obtenidos son bastante planos (Figura 76: R-175, R-495 y R-350), o muestran un aumento continuo de energía (Figura 76: R-488), lo cual no permite seleccionar una estructura que energéticamente indique la existencia de un mínimo intermedio. Los IPs obtenidos se utilizaron como puntos de partida para ir hacia reactivos y productos, a través de las coordenadas de reacción  $RC_B$  y  $RC_{AB}$ , respectivamente. Así, se consigue un nuevo conjunto de caminos refinados, aunque R-208 se descartó debido a que no se consiguió suavizar el cambio abrupto de energía que ocurre a consecuencia del cambio en el valor de  $\psi$ . Además, R-488 también se descarta ya que la barrera de energía para el ataque nucleofílico es de prácticamente 25.0 kcal/mol.

Cabe mencionar que si se calcula un promedio tipo Boltzmann de las barreras calculadas (ecuación 4.1), suponiendo una equiprobabilidad de los complejos, el valor obtenido es de 14.4 kcal/mol. En un estudio en el que calculaban las barreras para 40 complejos E:S de la proteasa de HIV-1 seleccionados de un total de 80ns de MD, Ribeiro *et al.*<sup>287</sup> comentaban que la reacción debía tener lugar desde las conformaciones transitorias asociadas a barreras bajas (en su estudio calcularon barreras entre 14.5 y 51.3 kcal/mol, dando un promedio de 16.5 kcal/mol). De este modo, parece adecuado centrarse en el estudio detallado de aquellos procesos reactivos que dominan el valor promedio de la barrera, aunque también mantenemos el reactivo R-350 para el análisis. En su estudio, dos de los factores identificados que modulaban las barreras eran una conformación de un puente de hidrógeno en el centro activo y las interacciones electrostáticas “instantáneas” con el entorno MM (más lejano). Aunque aquí no se ha evaluado específicamente para estas estructuras el efecto del entorno global en las barreras, en el estudio de GalNAc se observó que las variaciones descritas de las interacciones en el centro activo eran suficientes para explicar las diferencias observadas en las barreras. Y aquí, por lo comentado en este apartado, parece reproducirse de nuevo esta observación. A continuación se muestran los reactivos finales para las tres configuraciones seleccionadas siguiendo la coordenada  $RC_B$ .



**Figura 78.** Representación de barras y esferas para las estructuras de los 3 reactivos obtenidos mediante la  $RC_B$ . En cada una de estos se señalan varias distancias importantes, las cuales pueden ser examinadas con más detalle en las tablas. Los carbonos de los sustratos se dibujan en color rosa y los del resto de los residuos en color celeste. Además, en R-495 se señalan los identificadores de varios residuos pero en las siguientes imágenes no se señalaran para privilegiar la claridad de los valores de distancias.

Las siguientes tablas resumen toda la información estructural y electrónica de cada uno de los puntos estacionarios localizados en los caminos de reacción para R-495, R-175, R-350.

La localización de las especies denominada OC, al igual que el IP, se realiza en base a un criterio de geométrico. Por ejemplo: las estructuras del camino de reacción escogidas para buscar el TS1 tienen distancias  $C1_{\text{GlcNAc}}-O_{\text{P}}$  que están dentro del rango 2.0-2.6 Å, y las que se escogen para buscar el TS2 tienen distancias  $C1_{\text{GlcNAc}}-O4_{\text{GlcA}}$  que caen dentro del rango 2.0-2.4 Å. Como ya pasó en el estudio de EXTL2:GalNAc, la caracterización de estos puntos estacionarios en esta zona de alta energía resulta muy complicada, por lo que no siempre se consigue el análisis de segundas derivadas esperado. De todas formas, independientemente de los detalles concretos, parece evidente que la reacción procede por una zona de alta energía correspondiente a un conjunto de estructuras con carácter oxocarbenio (OC).

A continuación, se compara y discute la información resumida en las Tablas 15, 16 y 17, haciendo énfasis en los siguientes puntos:

1. El carácter disociativo del mecanismo, medido a través de la distancia  $C1_{\text{GlcNAc}}-O_{\text{P}}$  en la especie OC2-495, OC3-175 y OC3-350.
2. El valor de  $\psi$  en las especies UDP----OC, o sea, en las estructuras que se denotan como OC e IP.
3. Una tendencia cualitativa de la fuerza de la interacción electrostática Arg293-COO<sup>-</sup><sub>GlcA</sub>, basándonos en las distancias que separan a estos grupos.
4. Análogamente, a partir de la distancia  $OD1_{\text{Asp246}}-C1_{\text{GlcNAc}}$  se busca una tendencia cualitativa de la estabilización que puede brindar Asp246 a la especie UDP----OC. Por último, se discute acerca de su efecto en el distanciamiento de UDP y acercamiento de GlcA durante el camino de reacción.

**Tabla 15.** Información estructural y electrónica para el mecanismo de ataque frontal explorado a partir del IP-495. Se utiliza la  $RC_B$  para ir hacia Reactivos y la  $RC_{AB}$  para ir hacia Productos. Las geometrías se han optimizado al nivel de teoría BP86/TZVP para la región QM, las barreras de energía y las cargas atómicas se han calculado con el nivel M05-2X/TZVP. Las unidades de los parámetros geométricos, energías ( $\Delta E$ ) y cargas naturales ( $q$ ) corresponden a Å o grados ( $^\circ$ ), kcal/mol y u.a, respectivamente.

|   | R-495         | OC1-495       | IP-495        | OC2-495       | P-495         |
|---|---------------|---------------|---------------|---------------|---------------|
| <b>d(C1<sub>GlcNAc</sub>-O<sub>P</sub>)</b>       | <b>1.52</b>   | <b>2.56</b>   | <b>3.25</b>   | <b>3.38</b>   | <b>3.48</b>   |
| <b>d(C1<sub>GlcNAc</sub>-O4<sub>GlcA</sub>)</b>   | <b>3.45</b>   | <b>3.11</b>   | <b>2.49</b>   | <b>2.11</b>   | <b>1.52</b>   |
| <b>d(OD1<sub>D246</sub>-C1<sub>GlcNAc</sub>)</b>  | <b>4.62</b>   | <b>3.59</b>   | <b>3.38</b>   | <b>3.37</b>   | <b>3.98</b>   |
| d(C1 <sub>GlcNAc</sub> -O5 <sub>GlcNAc</sub> )    | 1.38          | 1.27          | 1.28          | 1.29          | 1.39          |
| <b><math>\Psi</math></b>                          | <b>-102.2</b> | <b>-122.8</b> | <b>-128.0</b> | <b>-125.5</b> | <b>-132.9</b> |
| d(HO4 <sub>GlcA</sub> -O <sub>P</sub> )           | 1.92          | 1.66          | 1.56          | 1.39          | 1.01          |
| d(HO4 <sub>GlcA</sub> -O4 <sub>GlcA</sub> )       | 0.98          | 1.01          | 1.03          | 1.09          | 1.69          |
| d(WAT604-O2B <sub>UDP</sub> )                     | 2.01          | 1.94          | 1.77          | 1.76          | 1.82          |
| d(HN2' <sub>GlcNAc</sub> -O1B <sub>GlcNAc</sub> ) | 2.32          | 3.15          | 3.45          | 3.63          | 3.86          |
| d(HN2' <sub>GlcNAc</sub> -O3 <sub>GlcA</sub> )    | 2.92          | 2.45          | 2.20          | 2.05          | 2.00          |
| d(O1B <sub>GlcNAc</sub> -HO3 <sub>GlcA</sub> )    | 1.71          | 1.55          | 1.53          | 1.57          | 1.75          |
| d(NH1 <sub>R293</sub> -O6 <sub>GlcNAc</sub> )     | 3.29          | 3.03          | 2.96          | 2.93          | 2.86          |
| <b>d(NH1<sub>R293</sub>-C1<sub>GlcNAc</sub>)</b>  | <b>3.89</b>   | <b>3.72</b>   | <b>3.68</b>   | <b>3.73</b>   | <b>4.05</b>   |
| <b>d(NH1<sub>R293</sub>-O5<sub>GlcNAc</sub>)</b>  | <b>3.23</b>   | <b>3.03</b>   | <b>2.99</b>   | <b>2.98</b>   | <b>3.09</b>   |
| d(HH22 <sub>R293</sub> -OD2 <sub>D246</sub> )     | 1.65          | 1.67          | 1.66          | 1.66          | 1.68          |
| d(HH12 <sub>R293</sub> -OD1 <sub>D246</sub> )     | 1.78          | 1.83          | 1.83          | 1.82          | 1.80          |
| d(NE <sub>R293</sub> -O6B <sub>GlcA</sub> )       | 3.90          | 3.83          | 3.70          | 3.55          | 3.29          |
| d(NH2 <sub>R293</sub> -O6B <sub>GlcA</sub> )      | 3.80          | 3.55          | 3.41          | 3.32          | 3.30          |
| d(HH21 <sub>R293</sub> -OH <sub>Y193</sub> )      | 2.03          | 1.99          | 1.98          | 1.98          | 1.97          |
| d(HH <sub>Y193</sub> -O6A <sub>GlcA</sub> )       | 1.88          | 1.84          | 1.81          | 1.78          | 1.76          |
| d(HH <sub>Y193</sub> -O6B <sub>GlcA</sub> )       | 2.41          | 2.35          | 2.34          | 2.37          | 2.44          |
| d(HE <sub>R181</sub> -O6A <sub>GlcA</sub> )       | 1.73          | 1.73          | 1.74          | 1.73          | 1.73          |
| d(HH21 <sub>R181</sub> -O5 <sub>GlcA</sub> )      | 1.66          | 1.66          | 1.68          | 1.69          | 1.71          |
| $\Delta q(C1+H1+O5)_{GlcNAc}$                     | 0.00          | 0.37          | 0.36          | 0.29          | -0.02         |
| $\varphi, \theta^\circ$                           | 225.1, 15.9   | 231.8, 36.6   | 235.2, 45.6   | 236.5, 41.8   | 229.7, 28.7   |
| $\Delta E$  | 0.00          | 15.84         | 16.58         | 14.03         | -3.46         |

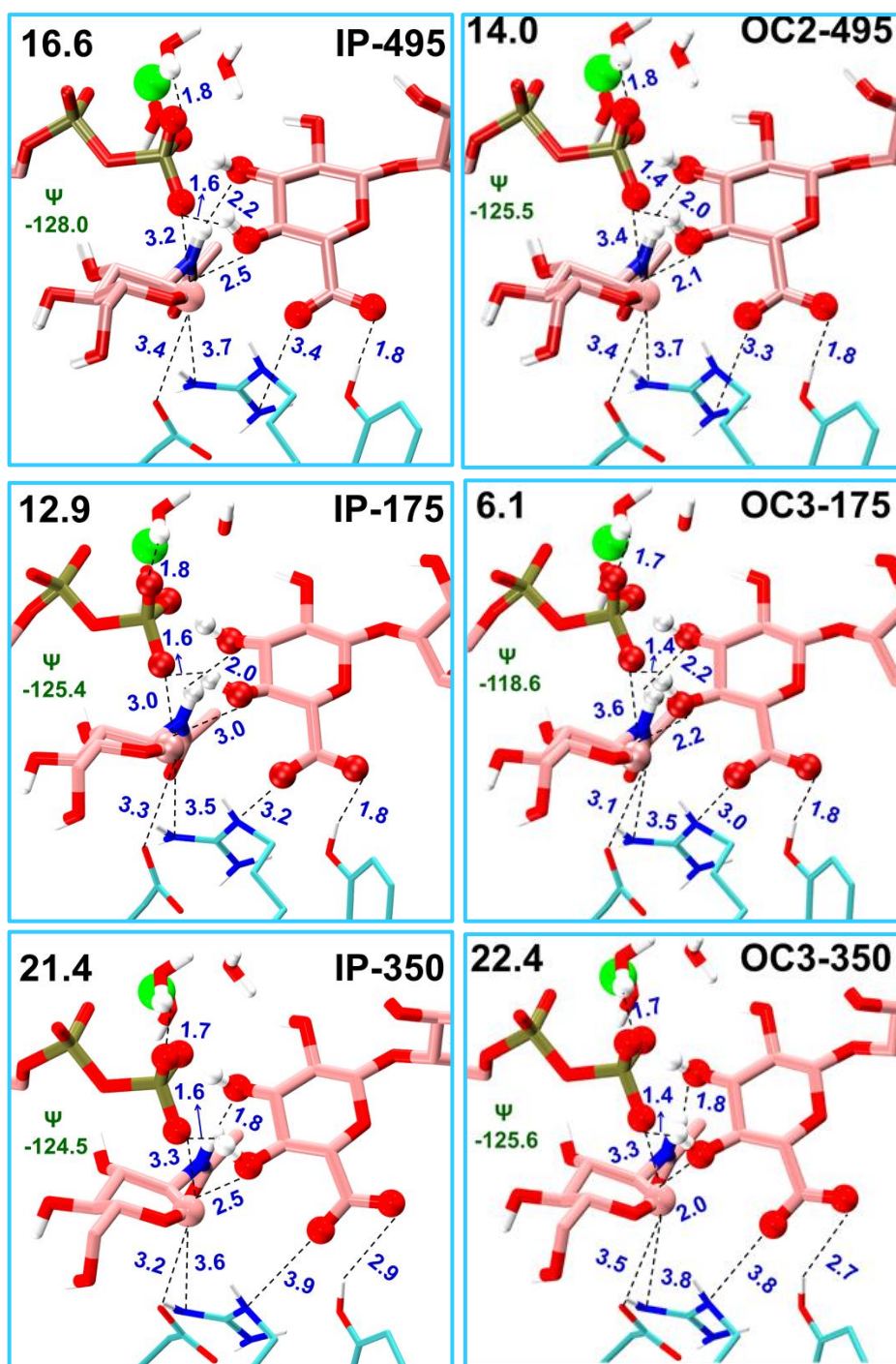
**Tabla 16.** Datos electrónicos y estructurales para el mecanismo de ataque frontal explorado a partir del IP-175 (se utiliza la  $RC_B$  para ir hacia Reactivos y la  $RC_{AB}$  para ir hacia Productos). Las geometrías se han optimizado al nivel de teoría QM(BP86/TZVP), las barreras de energía y las cargas atómicas se han calculado con el nivel QM(M05-2X/TZVP). Las distancias están en Å, el diedro en grados, las energías ( $\Delta E$ ) en kcal/mol y las cargas naturales (q) en u.a.

|  | R-175         | OC1-175       | OC2-175       | IP-175        | OC3-175       | P-175         |
|--|---------------|---------------|---------------|---------------|---------------|---------------|
| <b>d(C1<sub>GlcNAc</sub>-OP)</b>                 | <b>1.54</b>   | <b>2.14</b>   | <b>2.35</b>   | <b>2.99</b>   | <b>3.55</b>   | <b>3.56</b>   |
| <b>d(C1<sub>GlcNAc</sub>-O4<sub>GlcA</sub>)</b>  | <b>3.29</b>   | <b>3.25</b>   | <b>3.26</b>   | <b>3.04</b>   | <b>2.17</b>   | <b>1.50</b>   |
| <b>d(OD1<sub>D246</sub>-C1<sub>GlcNAc</sub>)</b> | <b>4.43</b>   | <b>3.80</b>   | <b>3.63</b>   | <b>3.30</b>   | <b>3.08</b>   | <b>3.63</b>   |
| d(C1 <sub>GlcNAc</sub> -O5 <sub>GlcNAc</sub> )   | 1.37          | 1.29          | 1.28          | 1.27          | 1.29          | 1.40          |
| <b><math>\Psi</math></b>                         | <b>-101.0</b> | <b>-109.1</b> | <b>-115.1</b> | <b>-125.4</b> | <b>-118.6</b> | <b>-119.6</b> |
| d(HO4 <sub>GlcA</sub> -OP)                       | 1.86          | 1.65          | 1.63          | 1.60          | 1.43          | 1.01          |
| d(HO4 <sub>GlcA</sub> -O4 <sub>GlcA</sub> )      | 0.98          | 1.00          | 1.01          | 1.02          | 1.08          | 1.69          |
| d(WAT604-O2B <sub>UDP</sub> )                    | 1.99          | 1.93          | 1.88          | 1.79          | 1.74          | 1.81          |
| d(HN2' <sub>GlcNAc</sub> -O1B <sub>UDP</sub> )   | 2.33          | 3.04          | 3.19          | 3.47          | 3.76          | 3.98          |
| d(HN2' <sub>GlcNAc</sub> -O3 <sub>GlcA</sub> )   | 2.71          | 2.14          | 2.11          | 2.07          | 2.19          | 2.14          |
| d(O1B <sub>UDP</sub> -HO3 <sub>GlcA</sub> )      | 1.72          | 1.57          | 1.57          | 1.54          | 1.60          | 1.78          |
| d(NH1 <sub>R293</sub> -O6 <sub>GlcNAc</sub> )    | 3.10          | 3.12          | 3.09          | 3.02          | 2.96          | 2.80          |
| <b>d(NH1<sub>R293</sub>-C1<sub>GlcNAc</sub>)</b> | <b>3.80</b>   | <b>3.58</b>   | <b>3.55</b>   | <b>3.48</b>   | <b>3.52</b>   | <b>3.87</b>   |
| <b>d(NH1<sub>R293</sub>-O5<sub>GlcNAc</sub>)</b> | <b>3.10</b>   | <b>3.00</b>   | <b>2.98</b>   | <b>2.97</b>   | <b>2.90</b>   | <b>2.99</b>   |
| d(HH12 <sub>R293</sub> -OD1 <sub>D246</sub> )    | 1.84          | 1.85          | 1.87          | 1.91          | 1.90          | 1.83          |
| d(HH22 <sub>R293</sub> -OD2 <sub>D246</sub> )    | 1.66          | 1.73          | 1.74          | 1.73          | 1.71          | 1.70          |
| d(NE <sub>R293</sub> -O6B <sub>GlcA</sub> )      | 3.39          | 3.14          | 3.14          | 3.16          | 2.99          | 2.94          |
| d(NH2 <sub>R293</sub> -O6B <sub>GlcA</sub> )     | 3.71          | 3.25          | 3.25          | 3.16          | 3.14          | 3.14          |
| d(HH21 <sub>R293</sub> -OH <sub>Y193</sub> )     | 1.89          | 1.96          | 1.95          | 1.96          | 1.93          | 1.91          |
| d(HH <sub>Y193</sub> -O6B <sub>GlcA</sub> )      | 1.77          | 2.22          | 2.23          | 2.23          | 2.30          | 1.91          |
| d(HH <sub>Y193</sub> -O6A <sub>GlcA</sub> )      | 2.51          | 1.84          | 1.82          | 1.80          | 1.77          | 1.74          |
| d(HE <sub>R181</sub> -O6A <sub>GlcA</sub> )      | 1.71          | 1.71          | 1.71          | 1.72          | 1.73          | 1.74          |
| d(HH21 <sub>R181</sub> -O5 <sub>GlcA</sub> )     | 1.79          | 1.73          | 1.73          | 1.73          | 1.76          | 1.75          |
| q(C1+H1+O5) <sub>GlcNAc</sub>                    | 0.00          | 0.31          | 0.35          | 0.36          | 0.31          | -0.03         |
| $\varphi, \theta^c$                              | 233.6, 18.5   | 222.8, 29.1   | 224.1, 32.6   | 226.6, 43.5   | 233.7, 44.8   | 227.5, 29.6   |
| $\Delta E$                                       | 0.00          | 12.48         | 12.99         | 12.88         | 6.14          | -15.52        |

**Tabla 17.** Información estructural y electrónica para el mecanismo de ataque frontal explorado a partir del IP-350 (la  $RC_B$  conduce hasta Reactivos y la  $RC_{AB}$  hasta Productos). Al igual que antes, las geometrías se han optimizado con el método QM(BP86/TZVP), y las barreras de energía y las cargas atómicas se han calculado con el método QM(M05-2X/TZVP). Las unidades de os

|   | R-350       | OC1-350       | OC2-350       | IP-350        | OC3-350       | P-350         |
|---|-------------|---------------|---------------|---------------|---------------|---------------|
| <b>d(C1<sub>GlcNAc</sub>-OP)</b>                  | <b>1.49</b> | <b>2.28</b>   | <b>2.41</b>   | <b>3.28</b>   | <b>3.34</b>   | <b>3.49</b>   |
| <b>d(C1<sub>GlcNAc</sub>-O4<sub>GlcA</sub>)</b>   | <b>3.35</b> | <b>3.03</b>   | <b>3.04</b>   | <b>2.52</b>   | <b>2.02</b>   | <b>1.51</b>   |
| <b>d(OD2<sub>D246</sub>-C1<sub>GlcNAc</sub>)</b>  | <b>4.39</b> | <b>3.60</b>   | <b>3.54</b>   | <b>3.18</b>   | <b>3.46</b>   | <b>4.01</b>   |
| d(C1 <sub>GlcNAc</sub> -O5 <sub>GlcNAc</sub> )    | 1.38        | 1.28          | 1.28          | 1.27          | 1.30          | 1.39          |
| <b><math>\psi</math></b>                          | <b>-3.3</b> | <b>-102.9</b> | <b>-105.6</b> | <b>-124.5</b> | <b>-125.6</b> | <b>-131.5</b> |
| d(HO4 <sub>GlcA</sub> -O1 <sub>UDP</sub> )        | 2.00        | 1.79          | 1.80          | 1.62          | 1.38          | 1.01          |
| d(HO4 <sub>GlcA</sub> -O4 <sub>GlcA</sub> )       | 0.98        | 1.00          | 1.00          | 1.02          | 1.10          | 1.66          |
| d(WAT604-O2B <sub>UDP</sub> )                     | 3.76        | 1.88          | 1.86          | 1.73          | 1.73          | 1.76          |
| d(HN2' <sub>GlcNAc</sub> -O1B <sub>GlcNAc</sub> ) | 3.50        | 3.27          | 3.30          | 3.61          | 3.73          | 3.97          |
| d(HN2' <sub>GlcNAc</sub> -O3 <sub>GlcA</sub> )    | 2.24        | 1.85          | 1.85          | 1.79          | 1.79          | 1.83          |
| d(O1B <sub>GlcNAc</sub> -HO3 <sub>GlcA</sub> )    | 1.78        | 1.55          | 1.55          | 1.53          | 1.57          | 1.77          |
| d(NH1 <sub>R293</sub> -O6 <sub>GlcNAc</sub> )     | 3.20        | 3.08          | 3.09          | 2.95          | 2.86          | 2.80          |
| <b>d(NH1<sub>R293</sub>-C1<sub>GlcNAc</sub>)</b>  | <b>3.84</b> | <b>3.66</b>   | <b>3.64</b>   | <b>3.62</b>   | <b>3.76</b>   | <b>4.02</b>   |
| <b>d(NH1<sub>R293</sub>-O5<sub>GlcNAc</sub>)</b>  | <b>3.31</b> | <b>3.01</b>   | <b>2.99</b>   | <b>2.95</b>   | <b>2.96</b>   | <b>3.04</b>   |
| d(HH12 <sub>R293</sub> -OD2 <sub>D246</sub> )     | 1.78        | 1.83          | 1.84          | 1.85          | 1.83          | 1.80          |
| d(HH22 <sub>R293</sub> -OD1 <sub>D246</sub> )     | 1.62        | 1.62          | 1.62          | 1.64          | 1.64          | 1.65          |
| d(NE <sub>R293</sub> -O6B <sub>GlcA</sub> )       | 4.42        | 4.28          | 4.32          | 3.93          | 3.80          | 3.67          |
| d(NH2 <sub>R293</sub> -O6B <sub>GlcA</sub> )      | 4.49        | 4.28          | 4.31          | 3.92          | 3.78          | 3.67          |
| d(HH21 <sub>R293</sub> -OH <sub>Y193</sub> )      | 2.00        | 1.95          | 1.95          | 1.94          | 1.94          | 1.96          |
| d(HH <sub>Y193</sub> -O6A <sub>GlcA</sub> )       | 3.28        | 3.18          | 3.20          | 2.86          | 2.70          | 2.56          |
| d(HH <sub>Y193</sub> -O6B <sub>GlcA</sub> )       | 1.73        | 1.71          | 1.72          | 1.74          | 1.80          | 1.86          |
| d(HE <sub>R181</sub> -O6A <sub>GlcA</sub> )       | 1.70        | 1.70          | 1.70          | 1.70          | 1.72          | 1.73          |
| d(HH21 <sub>R181</sub> -O5 <sub>GlcA</sub> )      | 1.75        | 1.73          | 1.73          | 1.71          | 1.71          | 1.71          |
| q(C1+H1+O5) <sub>GlcNAc</sub>                     | 0.00        | 0.36          | 0.38          | 0.37          | 0.25          | -0.01         |
| $\varphi, \theta$                                 | 189.2, 25.0 | 222.0, 30.8   | 223.1, 33.3   | 233.7, 46.5   | 235.1, 39.4   | 229.0, 26.6   |
| $\Delta E$  | 0.00        | 19.81         | 20.47         | 21.41         | 22.42         | 5.67          |





**Figura 79.** Representación de barras y esferas para las estructuras del IP, OC2-495, OC3-175 y OC3-350. En cada una de ellas se señalan varias distancias importantes en Å, las cuales pueden ser examinadas con más detalle en las Tablas 15, 16 y 17. Los carbonos de los sustratos se dibujan en color rosa y los del resto de los residuos en color celeste. No se han incluido los identificadores de cada residuo pues esto fue previamente indicado en la Figura 78.

El mecanismo más disociado se obtiene a partir de R-175, cuyo OC3 muestra la distancia  $O_P-C1_{GlcNAc}$  más larga (3.55 Å). Contrariamente, el IP-175 muestra la distancia  $O_P-C1_{GlcNAc}$  más corta entre los IPs (2.99 Å). Estos IPs tienen energías muy similares a las encontradas para las especies UDP----OC previas a éste (especies OC en las tablas), pero mayor energía que las especies OC posteriores al IP, con la excepción del OC3-350.

El valor promedio de  $\psi$  en las especies UDP----OC de R-495, R-350 y R-175 es  $-125.4^\circ \pm 1.8^\circ$ ,  $-114.7^\circ \pm 10.4^\circ$  y  $-117.1^\circ \pm 5.0^\circ$ , respectivamente. Los valores que más se desvían de su media se obtienen a partir de R-350, y también son las especies UDP----OC con las energías relativas más altas. Esto puede ser una consecuencia de la conformación inicial de UDP-GlcNAc, en la que  $\psi$  tiene un valor muy pequeño ( $-3.3^\circ$ ), lejano al que tendrá que alcanzar en las especies oxocarbenio. Es altamente probable que las conformaciones con valores absolutos pequeños ( $< |50|^\circ$ ) de  $\psi$  sean las más estables, ya que son las más muestreadas en las dinámicas moleculares (o el campo de fuerzas así lo da). A nivel QM(DFT)/MM, tales conformaciones continúan siendo las más estables, sin embargo es posible obtener valores de  $\psi$  que se aproximen a los  $-100^\circ$ , los cuales resultan aún más accesibles cuando se simula el mecanismo de reacción y se generan las especies UDP----OC.

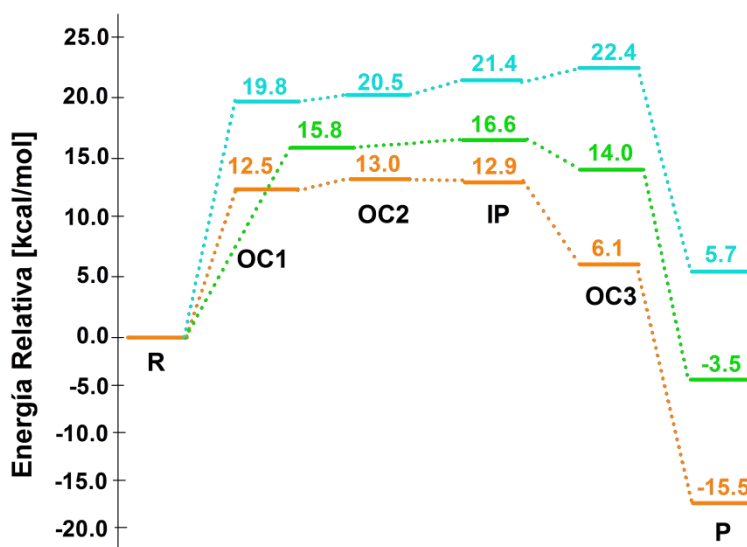
La interacción electrostática que es capaz de establecer Arg293 con  $COO^-_{GlcA}$  es un punto importante pues modula los movimientos del grupo carboxilato y por consiguiente, del hidroxilo nucleofílico. La revisión de las estructuras cristalográficas 10MZ y 1ON6 (UDP-GalNAc y UDP-GlcNAc, respectivamente), muestran que Arg293 (NH1) es capaz de posicionarse a 3.73 Å y 3.23 Å del  $C1_{GalNAc}$  y  $C1_{GlcNAc}$ , respectivamente. De hecho, en nuestros reactivos esta distancia tiene valores similares que caen dentro del rango 3.80-3.89 Å. Por otra parte, las distancias cristalográficas en 1ON8 para las parejas de átomos  $NE_{Arg293}-O6B_{GlcA}$  y  $NH2_{Arg293}-O6B_{GlcA}$  son de 3.07 Å y 3.30 Å, respectivamente (en 1ON8 no hay azúcar). Estos son valores más pequeños que los observados en los reactivos QMMM pero que se parecen bastante a los que se obtienen en los productos QMMM. Tales distancias menores a 5 Å, indican que los residuos están bien ubicados para interactuar electrostáticamente, más aun considerando que ambos grupos tienen una carga neta opuesta.

En cuanto a la distancia  $OD1_{Asp246}-C1_{GlcNAc}$  (OD2 en R-350), su análisis se aborda desde dos perspectivas: la primera tiene relación con el hecho de que el Asp246 tiene una carga neta negativa y por consiguiente, tiene la habilidad intrínseca de estabilizar el oxocarbenio (OC) que se forma durante el mecanismo. La segunda se asocia al llamado “*nucleophilic push*”<sup>45</sup> o “*empujón nucleofílico*”, en el cual se sugiere que el residuo nucleofílico que se ubica en la cara  $\beta$  de aquellas GTs-ret que lo poseen, puede facilitar la salida de UDP debilitando el enlace  $C1-O_P$ . Como ya se sabe, todas las GTs-ret con excepción de los miembros de la familia GT6, no poseen un nucleófilo bien posicionado en la cara  $\beta$  del azúcar para el ataque, y EXTL2 podría verse como un caso intermedio. En todos los reactivos, el Asp246 está a más de 4 Å del carbono anomérico interactuando fuertemente con Arg293, y según es posible apreciar en la Figura 78, sólo en el R-175 el Asp246 orienta uno de sus oxígenos hacia el carbono anomérico, sin embargo no se encuentra más cerca de éste que el Asp246 en R-350.

Al seguir la distancia  $OD1_{Asp246}-C1_{GlcNAc}$  a lo largo del mecanismo, se observa que el valor más pequeño se alcanza en el OC3-175 (3.08 Å), que es precisamente la estructura en la

que la distancia  $O_P-C1_{GlcNAc}$  alcanza su valor más largo (3.55 Å). Al pasar de IP-495 a OC2-495, el Asp246 se mantiene a una distancia relativa de GlcNAc bastante similar y  $O_P$  se ha distanciado sólo un poco más. Resulta curioso que a pesar de que en el IP-350 la distancia  $OD2_{Asp246}-C1_{GlcNAc}$  alcanza uno de los valores más pequeños (3.18 Å), es también la especie UDP----OC para la que se obtiene una barrera más alta. Los mayores valores de energía que se obtienen para R-495 y R-350 podrían entenderse por la tensión estérica adicional que experimentan estos sistemas al no poder catalizar un mecanismo igual de disociado que R-175. (Figura 79) Tanto en el OC2-495 como en el OC3-350, la distancia  $O_P-C1_{GlcNAc}$  no se alarga mucho más allá del valor visto en los respectivos IPs, y dado que el GlcA ya está bastante cerca del carbono anomérico en esta etapa del mecanismo, las tensiones estéricas podrían subir la energía de los intermedios y las barreras de esta reacción. El grado de disociación del mecanismo podría verse afectado por el empujón nucleofílico pues si el Asp246 se acerca lo suficiente rápido, podría forzar a UDP a alejarse rápidamente y así GlcA se podrá acercar sin un coste energético elevado (IP-175 → OC2-175). Estas distancias también están fuertemente determinadas por el movimiento de  $C1_{GlcNAc}$ , que al alejarse de UDP por el cambio de conformación del anillo se acerca a Asp246, y al mismo tiempo puede quedar más descubierto para el ataque de GlcA.

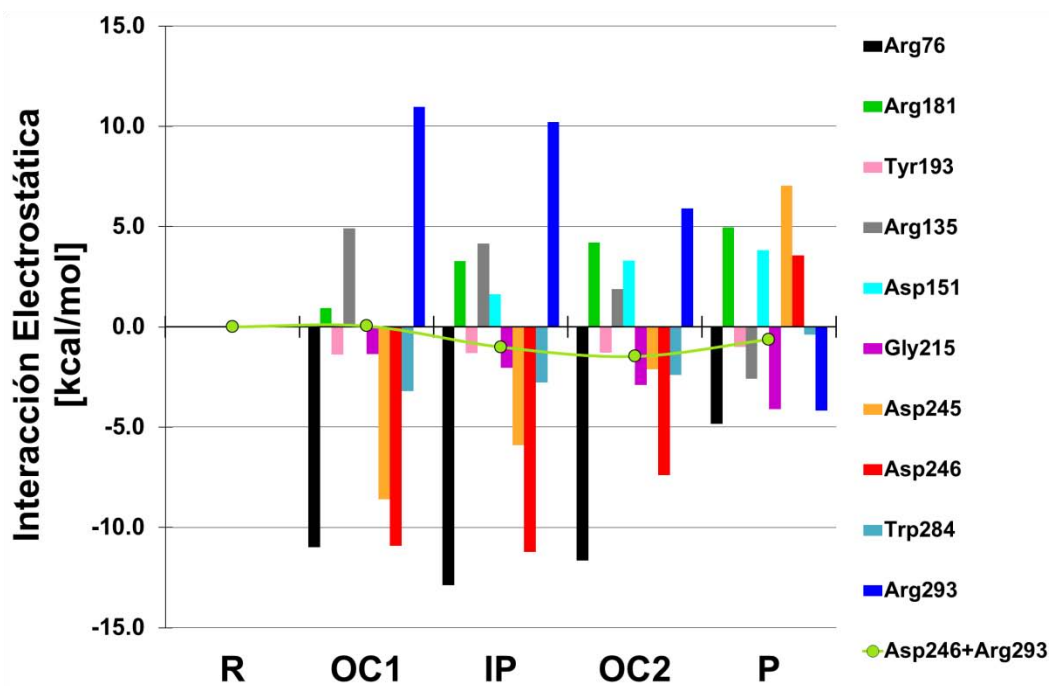
De este modo, los sutiles detalles apuntan a que Asp246 facilita un mecanismo altamente disociado no sólo estabilizando la carga de OC, sino que empujando más lejos el UDP para permitir una entrada relajada del sustrato aceptor. Entonces, ¿El Asp246 puede atacar para formar el intermedio covalente CGE? R-495 y R-175 indican que no es necesario pues una vez que se ha creado el espacio para GlcA, éste atacará por la cara  $\alpha$  del azúcar. Adicionalmente, el caso de R-350 muestra que si el Asp246 se ha aproximado más de lo necesario a  $C1_{GlcNAc}$ , éste tendrá que alejarse en orden a facilitar la entrada de GlcA.



**Figura 80.** Perfiles de energía a nivel QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM para cada uno de los caminos de reacción obtenidos a partir de: R-495 en color verde, R-175 en color naranja y R-350 en color celeste. También se indican las energías de algunas de las especies intermedias a lo largo del camino de reacción.

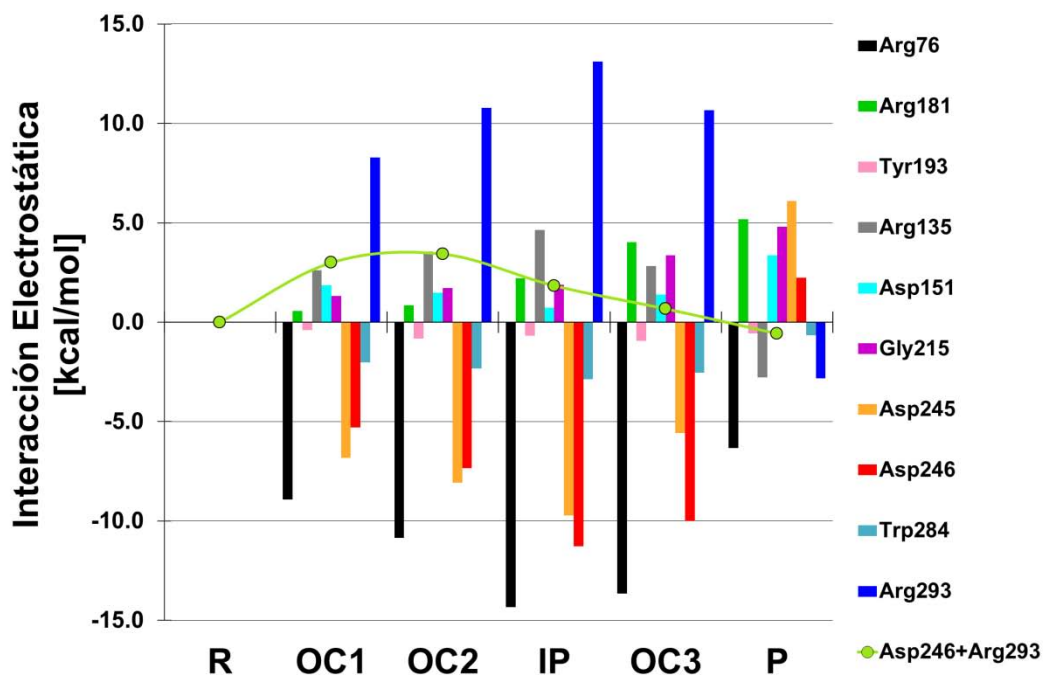


Los residuos que más estabilizan a las especies UDP----OC a lo largo del mecanismo de reacción a partir de R-495 son: el Asp245, el Asp246 y la Arg76. Los dos primeros mejoran sus interacciones con el azúcar y el último interacciona favorablemente con el UDP a lo largo del mecanismo de reacción (Figura 82).



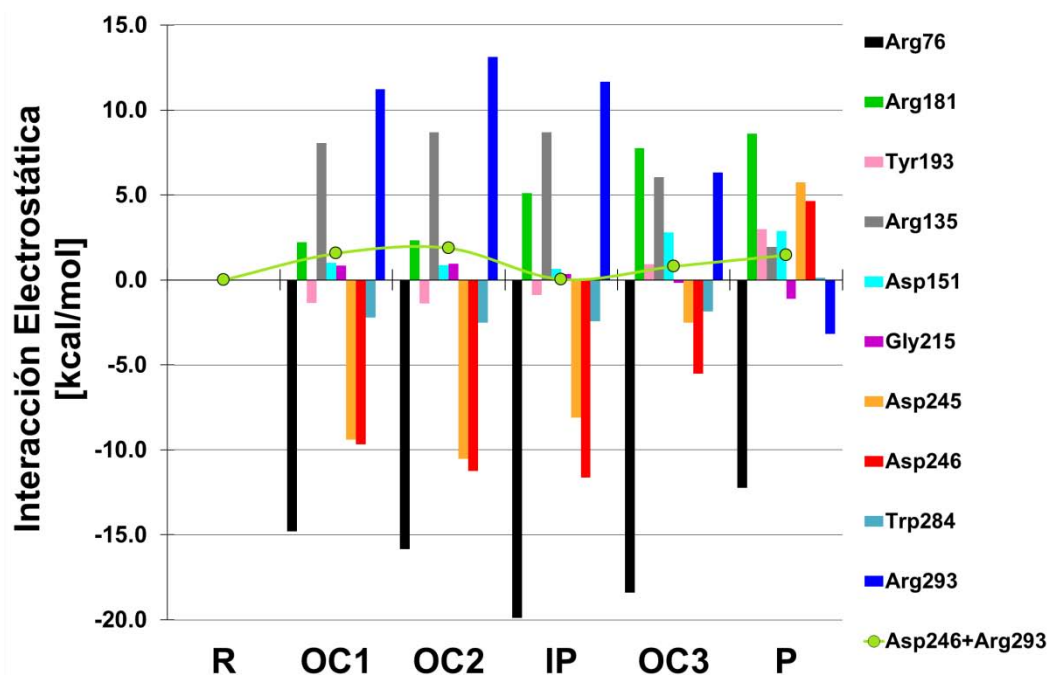
**Figura 82.** Contribución electrostática (o interacción electrostática) de diferentes residuos que interactúan directamente con los sustratos o con parte de la esfera de coordinación del  $Mg^{2+}$ . Los valores se calculan a nivel QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM para cada una de las especies caracterizadas en el mecanismo de reacción explorado a partir de R-495, utilizando como referencia energética el valor de la interacción electrostática en el reactivo.

Por otra parte, la Arg181 que interactúa fuertemente con GlcA deteriora sus interacciones durante la reacción, lo cual podría ser consecuencia del ligero distanciamiento con respecto a GlcA que sucede a lo largo del camino de reacción. Interaccionando con GlcNAc se tiene a la Arg135 y la Arg293, las que empeoran sus interacciones durante el mecanismo de reacción a causa del desarrollo de carga parcial positiva que experimenta el azúcar. Esto es más marcado con la Arg293, ya que se ubica a menos de 4 Å del  $C1_{GlcNAc}$  y a 3 Å del  $O5_{GlcNAc}$ . Es importante mencionar que esto no quiere decir que las interacciones entre estos residuos y los sustratos sean intrínsecamente malas, ya que en el reactivo contribuyen favorablemente a la energía de unión, sin embargo esto se ve afectado desfavorablemente al producirse la reacción. Por último, se estima la contribución electrostática conjunta de Asp246+Arg293, la cual resulta relativamente pequeña a lo largo del mecanismo de reacción aunque mostrando una ligera tendencia a favorecer las interacciones en OC2.



**Figura 83.** Interacciones electrostáticas de diferentes residuos calculadas a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM para cada una de las especies caracterizadas a lo largo del camino de reacción explorado a partir de R-175.

En el caso de R-175, los residuos cuyas interacciones se ven más favorecidas a lo largo de la reacción corresponden a Arg76, Asp245 y Asp246, aunque a diferencia de lo observado con R-495, el Asp245 consigue estabilizar casi con el doble de eficacia de lo que hace en R-495 (Figura 82, IP y OC2). Por otra parte, la Arg135, Arg181 y la Arg293 son los residuos cuyas interacciones se ven más deterioradas a lo largo de la reacción de transferencia. De hecho, la Arg135 deteriora sus interacciones notablemente pero es posible que Asp245 compense tal desestabilización (menos en el producto). Finalmente, la contribución conjunta Asp246+Arg293 resulta ser bastante reducida con una leve tendencia a la desestabilización que se hace prácticamente nula en OC3. En este sentido, el Asp246 menguaría significativamente la desestabilización debida a Arg293.



**Figura 84.** Contribuciones electrostáticas de diferentes residuos calculadas a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM para cada una de las especies caracterizadas en el mecanismo de reacción explorado a partir de R-350.

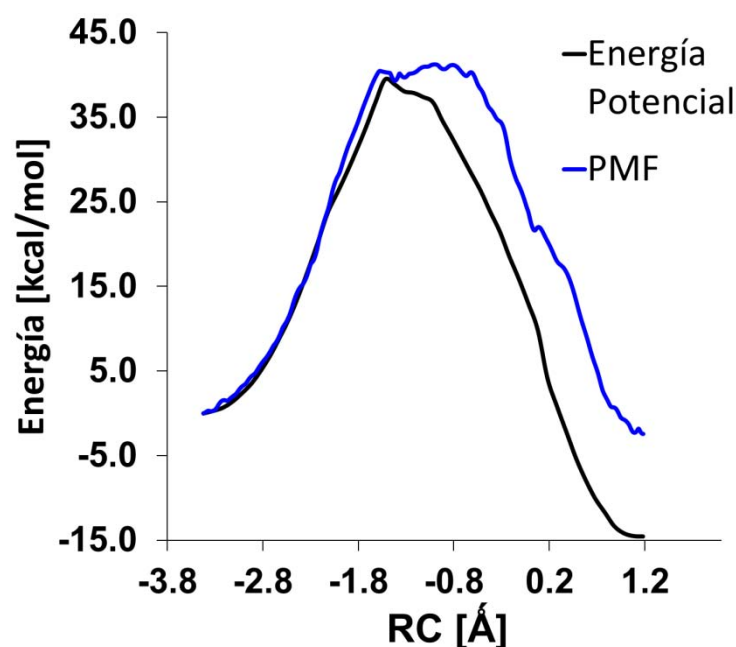
Al igual que ocurre con R-495 y R-175, los residuos cuyas interacciones se ven más favorecidas a lo largo de la reacción corresponden a Arg76, Asp245 y Asp246, y los que más deterioran sus interacciones son Arg135, Arg181 y Arg293. En este caso se observa que las contribuciones son más extremas pues la Arg293 llega casi a las 15 kcal/mol en OC2, pero la Arg76 alcanza una contribución de casi 20 kcal/mol en esa misma especie. De este modo, parece existir una compensación entre contribuciones favorables y desfavorables. Nuevamente, la contribución conjunta Asp246+Arg293 resulta ser bastante reducida con una leve tendencia a las desestabilización que es prácticamente nula en el IP.

Estos análisis permiten identificar la importancia de la estabilización que la proteína puede brindar a los grupos reaccionantes, por ejemplo; el desarrollo de carga negativa en el UDP durante la reacción, provoca que su interacción con Arg76 mejore sustancialmente; igualmente, el desarrollo de carga positiva en el GlcNAc durante la reacción, hace que las interacciones con Asp245 y Asp246 se hagan más favorables. La contribución conjunta Asp246+Arg293 sugiere que el Asp246 es capaz de compensar la desfavorable característica de tener una arginina en la cara  $\beta$  del azúcar. Es curioso que en nuestro análisis de la Arg293, no se observe el efecto favorable que debería causar la interacción Arg293-COO<sup>-</sup><sub>GlcA</sub> durante la formación de las especies UDP----OC, aunque cuando se forma el producto sí aparece una contribución favorable que corresponde precisamente a cuando

estos grupos están más cerca. De esta manera, se podría inferir que el efecto de la Arg293 está relacionado con la estabilidad relativa reactivo-producto.

### 6.3.1.2 Simulación de Umbrella Sampling

Como ya se ha visto previamente, el método SCC-DFTB tiende a sobreestimar considerablemente las barreras de energía potencial de este tipo de reacciones, 39.48 kcal/mol para R-175 y 41.20 kcal/mol para R-495. Sin embargo, al tratarse de un método semiempírico es posible calcular perfiles de energía a través de la técnica de US, los cuales estarían limitados en tiempo y recursos computacionales en caso de mantener el método DFT para la descripción de la partición QM.



**Figura 85.** Perfil de energía libre (azul) y perfil de energía potencial (negro) a nivel QM(SCC-DFTB)/MM(CHARMM). Se utilizó la coordenada de reacción  $RC_A = [d(O_P-C1_{GlcNAc}) - d(O4_{GlcA}-C1_{GlcNAc}) - d(O_P-HO4_{GlcA})]$ .

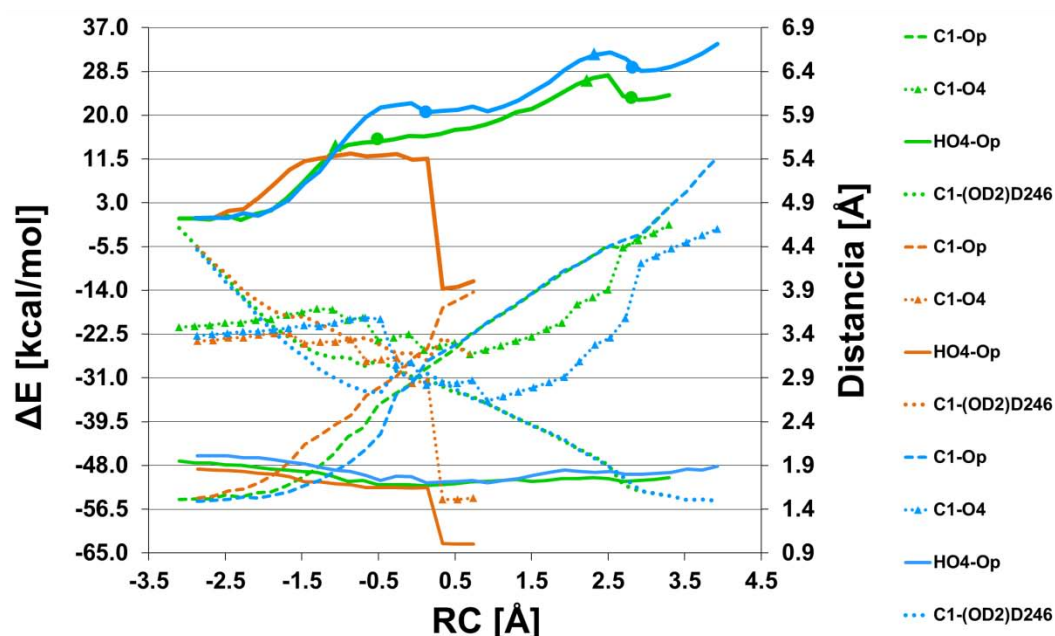
La comparación de ambos perfiles deja ver que el perfil de energía libre es más ancho y un poco más alto en energía que el perfil de energía potencial (cabe mencionar que el perfil de energía potencial a este nivel de cálculo siempre es estrecho en los sistemas estudiados en el grupo de GT-ret, con estructuras de máximo bastante asociativas). La zona de meseta que muestra el perfil de energía libre es semejante a lo que se suele encontrar utilizando por ejemplo, un nivel de teoría QM(BP86/TZVP) o QM(M05-2X/TZVP//BP86/TZVP). Además, esta región plana en el perfil de energía libre abarca desde una RC -1.6 Å hasta -0.6 Å, lo cual corresponde a un rango en el que se han muestreado principalmente especies UDP---OC. La barrera de energía libre es de 41.25 kcal/mol, 1.8 kcal/mol más alta que la barrera de energía potencial, lo cual indica que los efectos térmicos y entrópicos son



pequeños cuando se compara la barrera energética del proceso. Por último, se observa una diferencia más grande en la zona de productos del perfil de energía potencial y energía libre, el primero resultando más exoérgico.

### 6.3.2 Mecanismo de formación del CGE: Ataque Nucleofílico de Asp246

A pesar de que nuestros análisis apuntan a que el rol del Asp246 se relaciona con la estabilización de las especies UDP---OC y compensación del efecto desfavorable de Arg293, la posibilidad de que esté actuando como nucleófilo no se ha descartado y a continuación, se resume la información obtenida de estas simulaciones para R-495, R-175 y R-350.



**Figura 86.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM para el ataque nucleofílico de Asp246 (líneas gruesas superiores) simulado a través de la coordenada de reacción  $RC_D = [d(OP-C1_{GlcNAc}) - d(OD_{Asp246}-C1_{GlcNAc})]$ . Se mantiene el código de colores previamente usado para cada reactivo; verde para R-495 y todas las distancias involucradas con éste; naranja para R-175 y todas sus distancias; celeste para R-350 y todas las distancias que se relacionan con éste.

Como se puede apreciar en la Figura 86, la simulación del ataque de Asp246 se consigue llevar a cabo para R-495 y R-350 pero no para R-175. En el caso de las primeras dos, la reacción se inicia con el Asp246 acercándose a  $C1_{GlcNAc}$  mientras UDP se aleja de éste; simultáneamente, el GlcA se distancia ligeramente del carbono anomérico hasta que UDP se ubica a 2.7 Å. Después en R-350, GlcA, Asp246 y UDP quedan aproximadamente a 2.9 Å del carbono anomérico; en R-495, UDP y Asp246 también consiguen posicionarse a 2.9 Å, sin embargo GlcA se mantiene más alejado del carbono anomérico, apareciendo siempre a una distancia mayor que 3.0 Å. Este podría ser el motivo de que las energías son más bajas que las observadas para R-350, ya que teniendo al GlcA más alejado, Asp246

puede aproximarse en condiciones estéricas más relajadas. Finalmente, el CGE se forma al superar una barrera de 26.80 y 31.91 kcal/mol para R-495 y R-350, respectivamente.

En el caso de R-175, UDP se distancia más rápidamente del carbono anomérico que lo observado en R-495 y R-350; esto crea el espacio suficiente para que GlcA comience a aproximarse a C1<sub>GlcA</sub> sin aparecer una subida sustancial de energía durante el proceso. Esto indica la importancia de la naturaleza disociada del mecanismo ya que permite la aproximación rápida del nucleófilo a costa de un bajo costo energético. Finalmente el sistema colapsa, formándose el enlace glicosídico sin formarse el CGE. Notar que este hecho da cuenta del empujón nucleofílico que fue mencionado antes, ya que UDP se aleja bastante más rápidamente del carbono anomérico cuando se está acercando Asp246 en lugar de GlcA. De esta manera, además de estabilizar la carga positiva de GlcNAc, Asp246 puede facilitar la salida de UDP empujándolo lejos del carbono anomérico y por consiguiente, promoviendo una disociación rápida y pronunciada del sustrato dador, UDP-GlcNAc. En otras palabras, el residuo nucleofílico en la cara  $\beta$  del azúcar facilita el mecanismo altamente disociado que deben catalizar las GTs-ret.

A continuación, se resume la información de algunos de los puntos estacionarios que se obtienen a lo largo de la reacción de ataque del Asp246 para R-495 y R-350. La obtención de estos puntos se realizó de manera análoga a lo explicado para la obtención de los puntos estacionarios en los caminos de reacción para el mecanismo de ataque frontal.

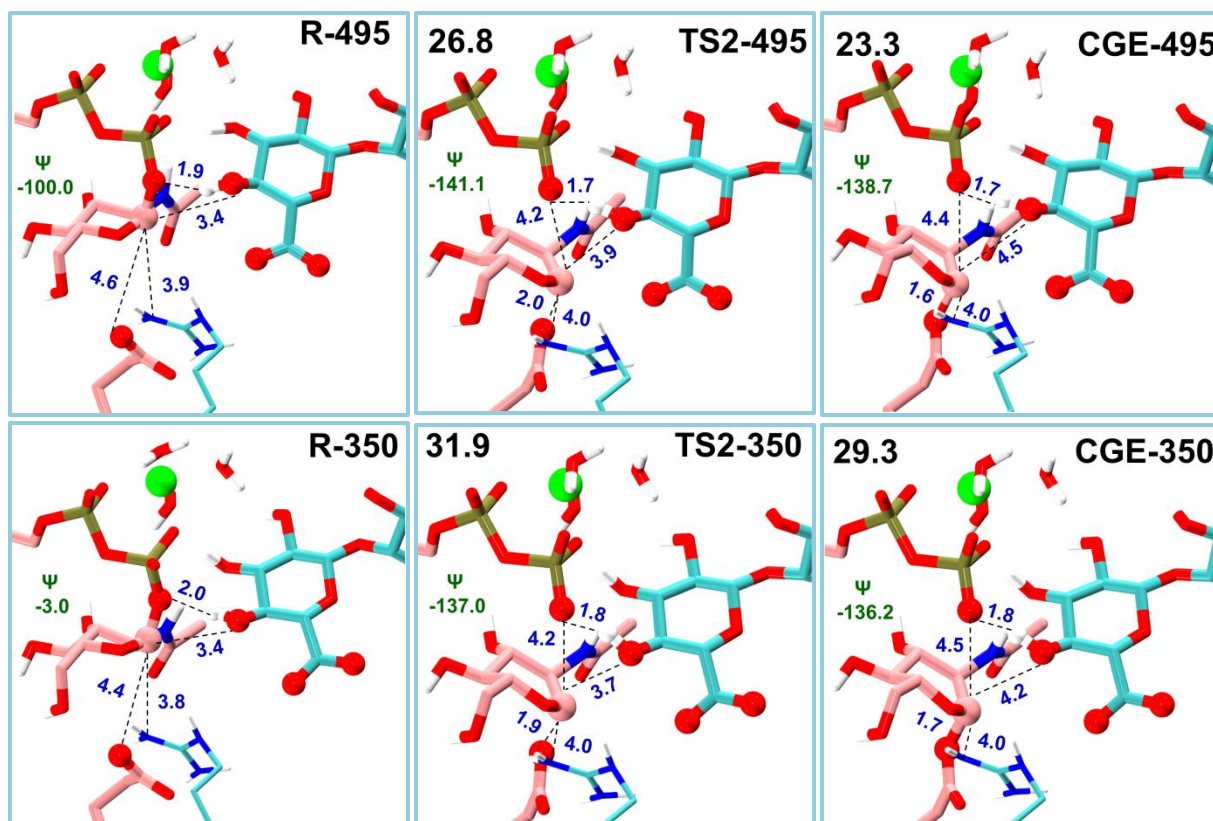
**Tabla 18.** Información estructural y electrónica para el mecanismo de ataque de Asp246 explorado a partir del reactivo, R-495. Las geometrías se han optimizado utilizando el método BP86/TZVP para la región QM. Las barreras de energía y las cargas atómicas se han calculado con el método M05-2X/TZVP para la región QM. Las unidades de los parámetros geométricos, energías ( $\Delta E$ ), cargas naturales ( $q$ ) y frecuencias normales de vibración corresponden a Å o grados ( $^\circ$ ) según corresponda, kcal/mol, u.a. y  $\text{cm}^{-1}$ , respectivamente.

|   | R-495         | OC-495         | IP-495         | TS2-495        | CGE-495        |
|---|---------------|----------------|----------------|----------------|----------------|
| <b>d(C1<sub>GlcNAc</sub>-O1<sub>UDP</sub>)</b>            | <b>1.52</b>   | <b>2.26</b>    | <b>2.73</b>    | <b>4.23</b>    | <b>4.45</b>    |
| <b>d(C1<sub>GlcNAc</sub>-O4<sub>GlcA</sub>)</b>           | <b>3.45</b>   | <b>3.32</b>    | <b>3.06</b>    | <b>3.93</b>    | <b>4.51</b>    |
| <b>d(OD1<sub>Asp246</sub>-C1<sub>GlcNAc</sub>)</b>        | <b>4.62</b>   | <b>3.53</b>    | <b>3.24</b>    | <b>2.01</b>    | <b>1.64</b>    |
| d(C1 <sub>GlcNAc</sub> -O5 <sub>GlcNAc</sub> )            | 1.38          | 1.28           | 1.27           | 1.30           | 1.35           |
| <b><math>\psi</math></b>                                  | <b>-99.95</b> | <b>-115.61</b> | <b>-128.18</b> | <b>-141.12</b> | <b>-138.74</b> |
| d(HO4 <sub>GlcA</sub> -O1 <sub>UDP</sub> )                | 1.92          | 1.68           | 1.65           | 1.71           | 1.73           |
| d(HO4 <sub>GlcA</sub> -O4 <sub>GlcA</sub> )               | 0.98          | 1.00           | 1.01           | 1.01           | 1.01           |
| d(WAT604-O2B <sub>UDP</sub> )                             | 2.01          | 2.03           | 1.88           | 1.80           | 1.84           |
| d(HN2' <sub>GlcNAc</sub> -O1B <sub>GlcNAc</sub> )         | 2.32          | 2.95           | 3.23           | 3.69           | 3.56           |
| d(HN2' <sub>GlcNAc</sub> -O3 <sub>GlcA</sub> )            | 2.92          | 2.63           | 2.44           | 2.43           | 2.60           |
| d(O1B <sub>GlcNAc</sub> -HO3 <sub>GlcA</sub> )            | 1.71          | 1.56           | 1.53           | 1.52           | 1.50           |
| d(NH1 <sub>R293</sub> -O6 <sub>GlcNAc</sub> )             | 3.29          | 3.11           | 2.99           | 2.78           | 2.89           |
| <b>d(NH1<sub>R293</sub>-C1<sub>GlcNAc</sub>)</b>          | <b>3.89</b>   | <b>3.71</b>    | <b>3.67</b>    | <b>4.00</b>    | <b>4.00</b>    |
| <b>d(NH1<sub>R293</sub>-O5<sub>GlcNAc</sub>)</b>          | <b>3.23</b>   | <b>2.98</b>    | <b>2.97</b>    | <b>3.06</b>    | <b>2.96</b>    |
| d(HH22 <sub>R293</sub> -OD2 <sub>D246</sub> )             | 1.65          | 1.67           | 1.67           | 1.86           | 1.94           |
| d(HH12 <sub>R293</sub> -OD1 <sub>D246</sub> )             | 1.78          | 1.86           | 1.93           | 3.57           | 3.65           |
| d(NE <sub>R293</sub> -O6B <sub>GlcA</sub> )               | 3.90          | 3.87           | 3.81           | 3.76           | 3.68           |
| d(NH2 <sub>R293</sub> -O6B <sub>GlcA</sub> )              | 3.80          | 3.60           | 3.51           | 3.21           | 2.95           |
| d(HH21 <sub>R293</sub> -OH <sub>Y193</sub> )              | 2.03          | 1.97           | 1.95           | 2.01           | 2.22           |
| d(HH <sub>Y193</sub> -O6A <sub>GlcA</sub> )               | 1.88          | 1.81           | 1.80           | 1.78           | 1.77           |
| d(HH <sub>Y193</sub> -O6B <sub>GlcA</sub> )               | 2.41          | 2.41           | 2.39           | 2.33           | 2.34           |
| d(HE <sub>R181</sub> -O6A <sub>GlcA</sub> )               | 1.73          | 1.74           | 1.73           | 1.73           | 1.73           |
| d(HH21 <sub>R181</sub> -O5 <sub>GlcA</sub> )              | 1.66          | 1.65           | 1.66           | 1.66           | 1.66           |
| $\Delta q(\text{C1}+\text{H1}+\text{O5})_{\text{GlcNAc}}$ | 0.00          | 0.35           | 0.38           | 0.24           | 0.00           |
| $\phi, \theta^\circ$                                      | 225.1, 15.9   | 233.6, 30.0    | 233.0, 38.9    | 242.1, 61.3    | 252.2, 69.3    |
| $\Delta E$  | 0.00          | 14.12          | 15.33          | 26.80          | 23.34          |
| $\nu (\text{cm}^{-1})$                                    |               | 43.74          | 107.94         | 7.00           |                |

**Tabla 19.** Detalles estructurales y electrónicos para el mecanismo de ataque de Asp246 explorado a partir de R-350. Las geometrías se calcularon al nivel QM(BP86/TZVP). Las cargas atómicas y las barreras de energía fueron obtenidas con el nivel QM(M05-2X/TZVP). Las unidades de los parámetros geométricos, energías ( $\Delta E$ ), cargas naturales ( $q$ ) y frecuencias normales de vibración corresponden a Å o grados ( $^\circ$ ) según corresponda, kcal/mol, u.a. y  $\text{cm}^{-1}$ , respectivamente.

|   | R-350        | OC-350         | IP-350         | TS2-350        | CGE-350        |
|---|--------------|----------------|----------------|----------------|----------------|
| <b>d(C1<sub>GlcNAc</sub>-O1<sub>UDP</sub>)</b>    | <b>1.49</b>  | <b>2.70</b>    | <b>3.08</b>    | <b>4.24</b>    | <b>4.48</b>    |
| <b>d(C1<sub>GlcNAc</sub>-O4<sub>GlcA</sub>)</b>   | <b>3.35</b>  | <b>3.05</b>    | <b>2.82</b>    | <b>3.68</b>    | <b>4.17</b>    |
| <b>d(OD2<sub>D246</sub>-C1<sub>GlcNAc</sub>)</b>  | <b>4.39</b>  | <b>2.99</b>    | <b>2.96</b>    | <b>1.93</b>    | <b>1.66</b>    |
| <b>d(C1<sub>GlcNAc</sub>-O5<sub>GlcNAc</sub>)</b> | <b>1.38</b>  | <b>1.27</b>    | <b>1.27</b>    | <b>1.30</b>    | <b>1.34</b>    |
| <b><math>\psi</math></b>                          | <b>-2.96</b> | <b>-109.32</b> | <b>-118.54</b> | <b>-137.24</b> | <b>-136.20</b> |
| d(HO4 <sub>GlcA</sub> -O1 <sub>UDP</sub> )        | 2.00         | 1.78           | 1.70           | 1.76           | 1.80           |
| d(HO4 <sub>GlcA</sub> -O4 <sub>GlcA</sub> )       | 0.98         | 1.00           | 1.01           | 1.00           | 1.00           |
| d(WAT604-O2B <sub>UDP</sub> )                     | 3.76         | 1.79           | 1.74           | 1.74           | 1.73           |
| d(HN2' <sub>GlcNAc</sub> -O1B <sub>UDP</sub> )    | 3.50         | 3.36           | 3.51           | 3.72           | 3.69           |
| d(HN2' <sub>GlcNAc</sub> -O3 <sub>GlcA</sub> )    | 2.24         | 1.86           | 1.82           | 1.85           | 1.88           |
| d(O1B <sub>GlcNAc</sub> -HO3 <sub>GlcA</sub> )    | 1.78         | 1.52           | 1.50           | 1.50           | 1.49           |
| d(NH1 <sub>R293</sub> -O6 <sub>GlcNAc</sub> )     | 3.20         | 3.03           | 2.99           | 2.78           | 2.91           |
| <b>d(NH1<sub>R293</sub>-C1<sub>GlcNAc</sub>)</b>  | <b>3.84</b>  | <b>3.56</b>    | <b>3.56</b>    | <b>4.01</b>    | <b>3.98</b>    |
| <b>d(NH1<sub>R293</sub>-O5<sub>GlcNAc</sub>)</b>  | <b>3.31</b>  | <b>2.95</b>    | <b>2.95</b>    | <b>3.01</b>    | <b>2.93</b>    |
| d(HH12 <sub>R293</sub> -OD2 <sub>D246</sub> )     | 1.78         | 2.03           | 1.97           | 3.71           | 3.66           |
| d(HH22 <sub>R293</sub> -OD1 <sub>D246</sub> )     | 1.62         | 1.65           | 1.63           | 1.80           | 1.85           |
| d(NE <sub>R293</sub> -O6B <sub>GlcA</sub> )       | 4.42         | 4.28           | 4.00           | 3.94           | 3.92           |
| d(NH2 <sub>R293</sub> -O6B <sub>GlcA</sub> )      | 4.49         | 4.34           | 3.99           | 3.77           | 3.62           |
| d(HH21 <sub>R293</sub> -OH <sub>Y193</sub> )      | 2.00         | 1.94           | 1.93           | 1.90           | 1.91           |
| d(HH <sub>Y193</sub> -O6A <sub>GlcA</sub> )       | 3.28         | 3.25           | 2.97           | 3.01           | 2.95           |
| d(HH <sub>Y193</sub> -O6B <sub>GlcA</sub> )       | 1.73         | 1.72           | 1.71           | 1.68           | 1.67           |
| d(HE <sub>R181</sub> -O6A <sub>GlcA</sub> )       | 1.70         | 1.71           | 1.71           | 1.73           | 1.72           |
| d(HH21 <sub>R181</sub> -O5 <sub>GlcA</sub> )      | 1.75         | 1.75           | 1.71           | 1.69           | 1.69           |
| q(C1+H1+O5) <sub>GlcNAc</sub>                     | 0.00         | 0.39           | 0.39           | 0.19           | 0.03           |
| $\varphi, \theta$                                 | 189.2, 25.0  | 225.9, 40.8    | 229.3, 46.2    | 245.4, 62.6    | 250.2, 67.1    |
| $\Delta E$  | 0.00         | 21.88          | 20.58          | 31.91          | 29.29          |
| $\nu$ ( $\text{cm}^{-1}$ ) <sup>1</sup>           |              | 76.87          | 89.73          | 37.12 <i>i</i> |                |

Al examinar las Tablas 18 y 19, se verifica que las energías son bastante similares a las obtenidas para el ataque frontal hasta la formación del IP, aunque GlcA está más lejos aquí de modo que Asp246 puede aproximarse más al carbono anomérico.



**Figura 87.** Representaciones moleculares de algunos de los puntos estacionarios de la superficie de energía potencial más relevantes para el mecanismo de ataque nucleofílico del Asp246. Primero para R-495 y después para R-350, se muestran el Reactivo, el TS2 y el CGE. Las distancias están en Å.

Tanto en el TS2-495 como en el TS2-350, UDP se ubica a más de 4 Å del carbono anomérico y GlcA a más de 3.5 Å de éste. Esto es principalmente causado por la distorsión adicional que debe sufrir el azúcar; en el mecanismo de ataque frontal los valores indican que las especies OC se encuentran entre conformaciones  ${}^4E$  y  ${}^4H_3$ , sin embargo en los TS2 del ataque nucleofílico de Asp246, las conformaciones se desplazan hacia un bote de tipo  ${}^{1,4}B$ . Esta es una conformación algo más distorsionada que las anteriores y por los valores de energía, parece ser también menos estabilizada por la enzima. En el CGE, la conformación del GlcNAc sigue siendo  ${}^{1,4}B$  y la energía no disminuye sustancialmente con respecto al TS2.

Comparando para una misma estructura de partida de reactivos las barreras de los dos mecanismos, se observa que el ataque del Asp246 es menos favorable que el ataque frontal (alrededor de 10 kcal/mol). De esta manera, todo apunta a que el rol del Asp246 no es actuar como nucleófilo para formar el intermedio CGE sino más bien, asistir al ataque

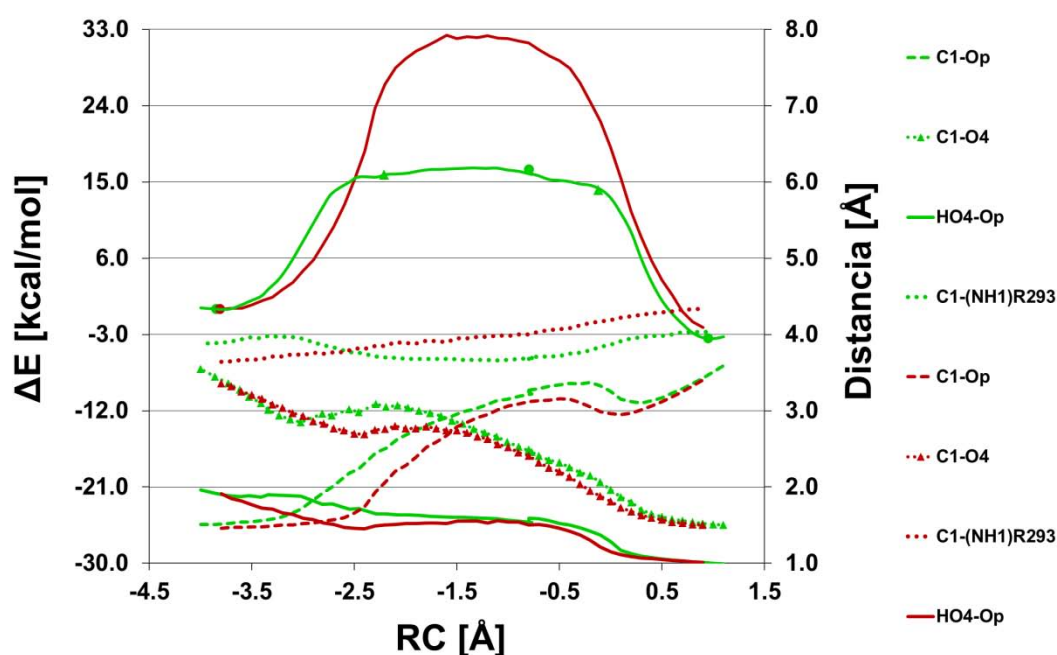
frontal del sustrato aceptor facilitando el mecanismo altamente disociado en dos maneras: estabilizando la especie oxocarbenio y empujando al UDP lejos del carbono anomérico. Creemos que la capacidad de llevar a cabo el empujón nucleofílico también puede ser realizada, aunque en menor medida, por residuos neutros, como Gln189 en LgtC, que ocupa una posición estructuralmente análoga a la de Asp246 en EXTL2. En el caso de  $\alpha$ 3GalT, esta posición la ocupa el Glu317 que además es capaz de brindar una sustancial estabilización a la especie OC gracias a su carga neta negativa. Es importante mencionar que en LgtC y  $\alpha$ 3GalT, Gln189 y Glu317 están cerca del carbono anomérico ( $<4.0 \text{ \AA}$ ) pero además forman un enlace de hidrógeno con el sustrato aceptor. En EXTL2, Asp246 no podría hacer esto porque se encontraría directamente con  $\text{COO}^-_{\text{GlcA}}$ , por lo que deducimos que la Arg293 podría estar ahí para evitar tal cosa. Arg293 interacciona electrostáticamente con  $\text{COO}^-_{\text{GlcA}}$  y hace lo mismo con Asp246 a través de su grupo guanidinio, de modo que provee un ambiente electrostático adecuado.

Para seguir investigando el rol de Asp246 y Arg293 en el mecanismo de reacción de EXTL2, se construyeron mutantes *in silico* del reactivo R-495. A continuación se describen los resultados de estos experimentos.

### 6.3.3 Mutaciones *in silico* y Mecanismo de Reacción

Como primer estudio del posible efecto de mutar Asp246 y Arg293 se escogió el reactivo R-495 y se mutó el residuo en cuestión, re-optimizando después la estructura y calculando el perfil de reacción para el mutante. Esta es una aproximación estática que omite posibles efectos estructurales mayores a consecuencia de la mutación, pero permite una comparación directa con la enzima nativa y así evaluar el rol del residuo en esta posición, suponiendo que los cambios estructurales asociados son menores. En cierta forma podría verse como un análisis parecido al de las contribuciones electrostáticas presentado previamente, pero un poco más elaborado (aquí también se incluyen efectos estéricos y se reoptimiza), y que permite evaluar también el cambio de naturaleza del residuo, no solo eliminarlo. En el siguiente apartado se presentaran los resultados de las MD que se hicieron de los mutantes para estimar las consecuencias estructurales de la sustitución.

#### 6.3.3.1 Mutante D246A: Mecanismo de Ataque Frontal

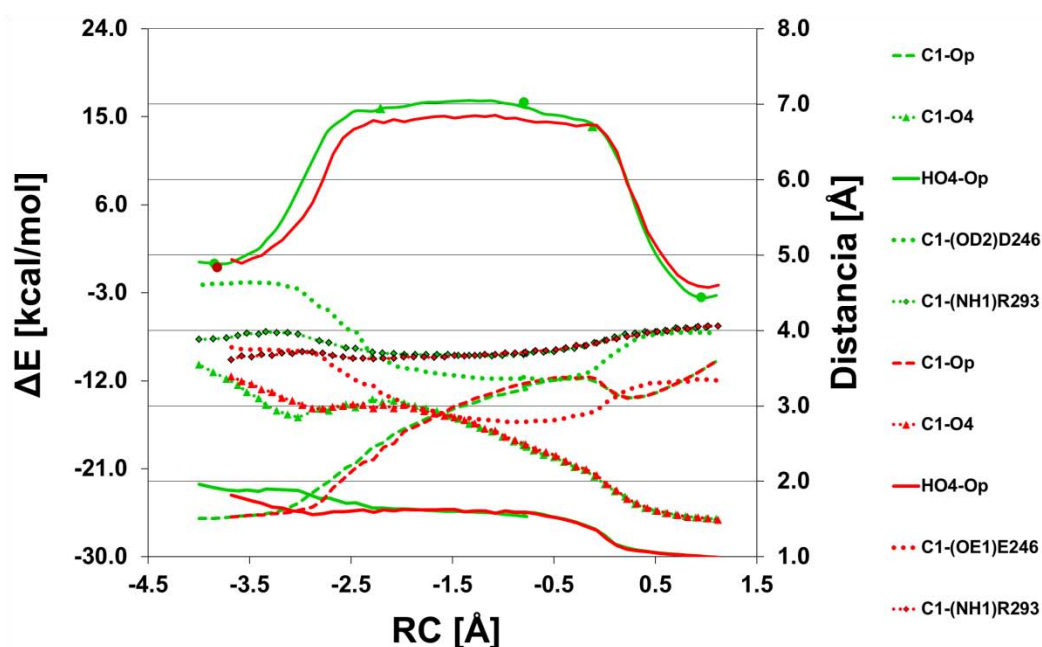


**Figura 88.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM para la enzima nativa (verde), y la mutante D246A (granate), líneas continuas superiores. También se muestran las distancias más relevantes obtenidas a nivel QM(BP86/TZVP)/MM(CHARMM) utilizando la correspondencia de colores verde para nativa y granate para mutante. Para los puntos estacionarios del mecanismo se usan triángulos (OC) y círculos (R, IP y P).

Lo primero que se observa es que la disociación UDP-GlcNAc ocurre más lentamente en el caso de la mutante, por ejemplo, en  $RC = -2.20 \text{ \AA}$  (26.52 kcal/mol) la distancia  $C1_{\text{GlcNAc}}-O_P$  ha alcanzado apenas los 2.06 Å mientras que en la enzima nativa, a ese valor de RC le

corresponde una distancia  $C1_{\text{GlcNAc}}-O_P = 2.51 \text{ \AA}$  (15.89 kcal/mol). La diferencia energética de estas dos estructuras es de 10.63 kcal/mol, una divergencia sustancial que está relacionada con diferentes factores. Asp246 es una alanina, por lo que la estrategia del empujón nucleofílico queda anulada así como la estabilización electrostática de la carga positiva emergente. Por otra parte, el GlcA se aproxima un poco más rápido al carbono anomérico al principio del mecanismo, sin embargo esto también podría subir la energía debido a que aún no se ha creado suficiente espacio para la entrada del aceptor. En otras palabras, la mutación de Asp246 ha disminuido la habilidad de la enzima para facilitar el mecanismo altamente disociado a través del empujón nucleofílico y la capacidad estabilizadora que este residuo es capaz de llevar a cabo. Este proceso resulta en una barrera de energía de casi 32.3 kcal/mol.

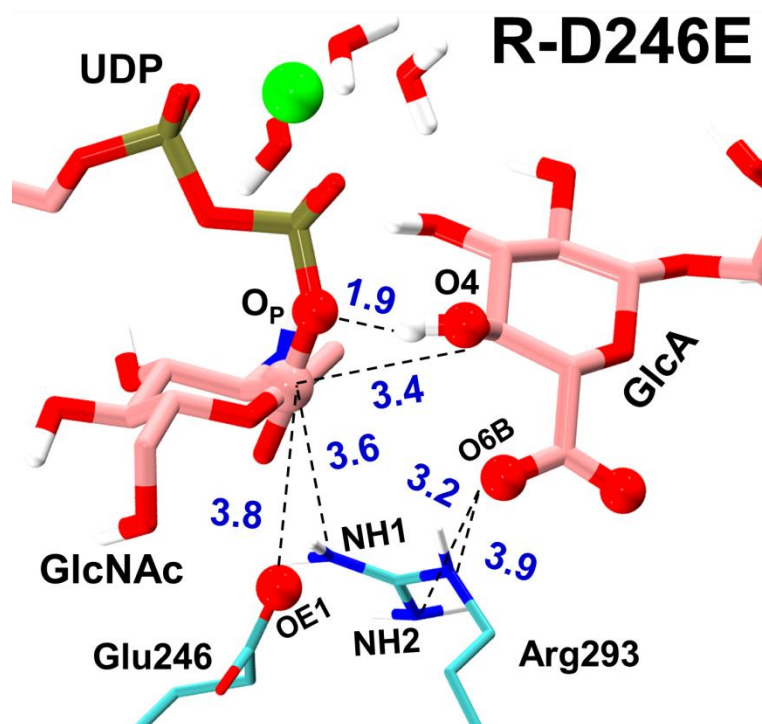
### 6.3.3.2 Mutante D246E: Mecanismo de Ataque Frontal



**Figura 89.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM (líneas continuas superiores), junto a las distancias más relevantes a nivel QM(BP86/TZVP)/MM(CHARMM) para la enzima nativa (verde), y la mutante D246E (rojo). También se incluyen las distancias  $C1-(OD2)D246$  (verde) y  $C1-(OE1)E246$  (rojo) para la enzima nativa y mutante D246E, respectivamente. Por último, los triángulos se refieren a especies OC y los círculos representan a R, IP y P.

Tal como se esperaba, Glu246 se coloca más cerca del carbono anomérico ya que tiene una cadena lateral más larga que Asp246. Glu246 se reacomoda de tal manera que uno de sus oxígenos (OE1) apunta hacia el carbono anomérico mientras que el otro, queda apuntando en dirección opuesta, perdiendo una de sus interacciones con Arg293 (Figura 90).

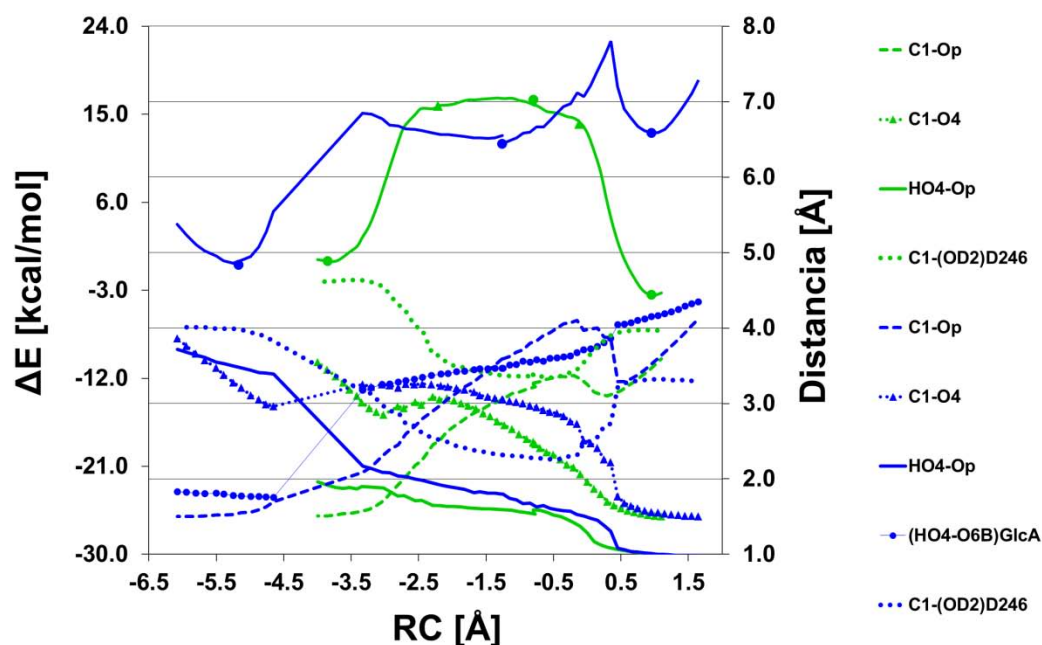




**Figura 90.** Reactivo optimizado a nivel QM(BP86/TZVP)/MM(CHARMM) para la enzima mutante D246E. Se señalan algunas distancias relevantes en Å.

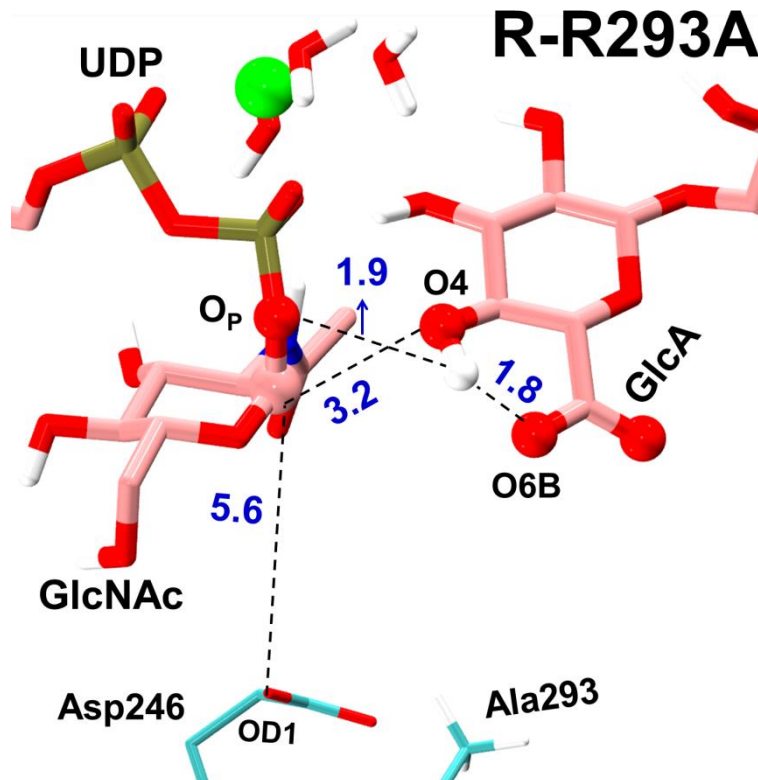
Aunque la Arg293 está un poco más cerca del carbono anomérico que en el reactivo utilizado (R-495), esto no parece afectar la energía pues se encuentra interaccionando fuertemente con el Glu246. Por otra parte, Glu246 logra posicionarse a menos de 3.0 Å del C1 a lo largo del mecanismo de reacción y aunque esto no afecta la posición relativa de UDP o GlcA (no se altera el grado de disociación del mecanismo), la interacción electrostática entre Glu246 y el azúcar debería mejorar a causa de la disminución de esta distancia. Esto explica porque la curva de energía de energía es ligeramente más baja comparada a la enzima nativa. Estos resultados sugieren nuevamente un rol estabilizador para el residuo nucleofílico de la cara  $\beta$  de GlcNAc, aunque hay que tener presente que hasta ahora no se han considerado los efectos que el Glu246 tendría en la unión del sustrato aceptor. Esto se considerará más adelante.

### 6.3.3.3 Mutante R293A: Mecanismo de Ataque Frontal y Ataque del Asp246



**Figura 91.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM, líneas continuas superiores. También se indican las distancias más relevantes a nivel QM(BP86/TZVP)/MM(CHARMM) para la enzima nativa (verde), y la mutante R293A (azul). Por último, también se representan los puntos estacionarios del mecanismo a través de triángulos (OC) y círculos (R, IP y P).

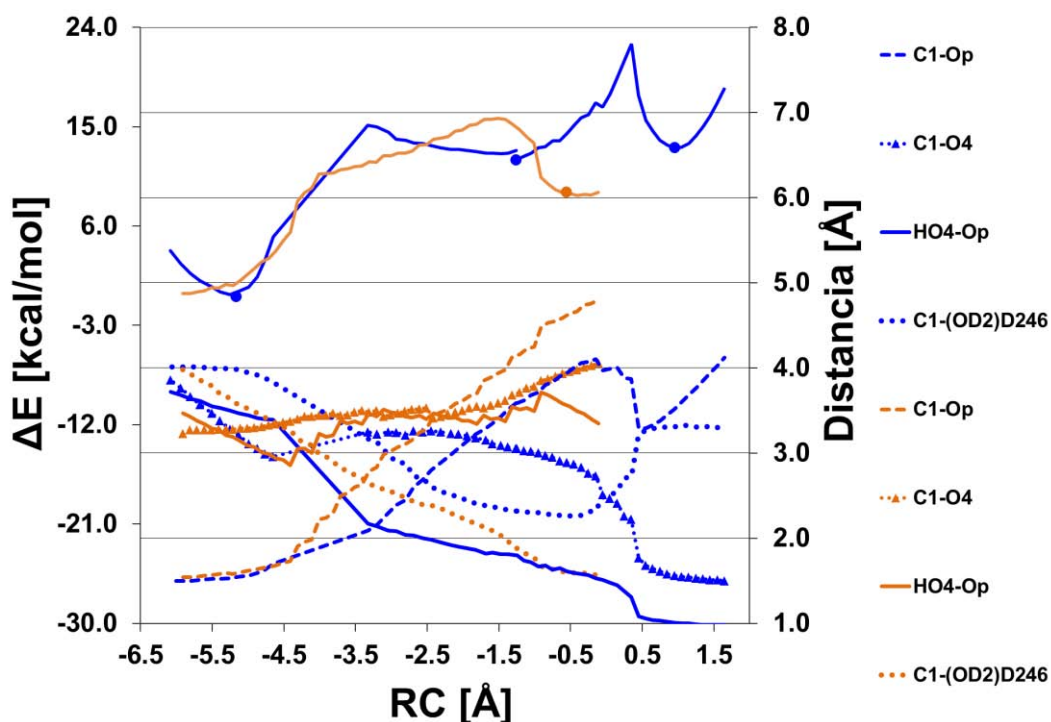
Al mutar la Arg293 por alanina se observa que la curva de energía potencial deja de ser plana. El IP tiene una energía un poco más baja debido a que el Asp246 logra acercarse bastante más al carbono anomérico ya que no interacciona con la Arg293, y posiblemente más importante, no existe el efecto desfavorable de la Arg293 durante la formación de las especies UDP---OC. Además, la disociación UDP-GlcNAc es mucho más rápida y pronunciada, esto probablemente se deba a que el empujón nucleofílico es más efectivo sin la presencia de Arg293. Como la catálisis del mecanismo disociado ha sido efectiva y ya se ha generado el espacio para la entrada de GlcA, se esperaría que el ataque ocurriera rápidamente. Sin embargo, en la etapa de ataque de GlcA la energía sube hasta 22.50 kcal/mol, lo cual parece ser una consecuencia de que la aproximación de GlcA hasta el carbono anomérico es más costosa energéticamente y más lenta. Es importante destacar que en esta enzima mutante no se establece la interacción  $HO4_{GlcA}-O_P$  en el reactivo sino que HO4 está interaccionando con  $O6B_{GlcA}$  (Figura 92). Esto podría ser consecuencia de la ausencia de la Arg293 pero en cualquier caso, al aplicar la coordenada de reacción ( $RC_A$ ), la interacción  $O_P-HO4_{GlcA}$  se forma justo antes de iniciarse la disociación UDP-GlcNAc.



**Figura 92.** Reactivo optimizado a nivel QM(BP86/TZVP)/MM(CHARMM) para la mutante R293A. Se señalan algunas distancias relevantes en Å.

Esto sugiere que la Arg293 es importante para la correcta orientación del sustrato aceptor, muy posiblemente gracias a su interacción electrostática con  $\text{COO}^-_{\text{GlcA}}$ . Esto es destacable pues aunque en los análisis electrostáticos la estabilización que esta interacción es capaz de brindar queda escondida por el efecto desfavorable de su interacción con las especies  $\text{UDP} \cdots \text{OC}$ , la simulación del ataque frontal de la enzima mutante R293A nos muestra la importancia de la Arg293 en la etapa de ataque nucleofílico, sugiriendo que el acercamiento de GlcA es más fácil si su grupo  $\text{COO}^-_{\text{GlcA}}$  interacciona con una carga neta positiva que, además, puede apantallar la carga del Asp246.

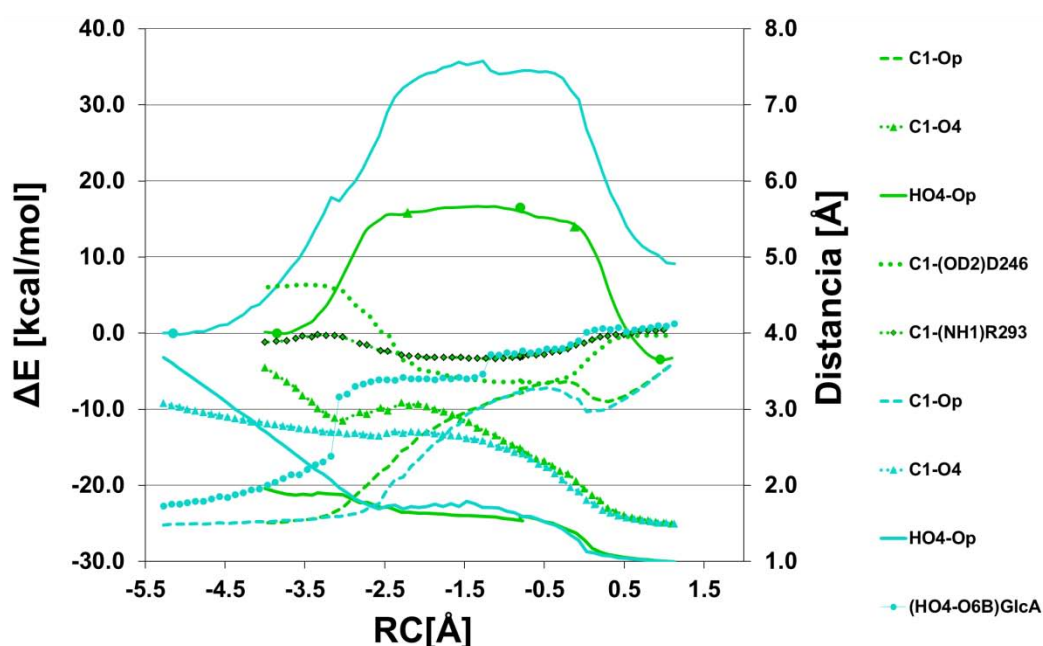
Como el Asp246 logra acercarse bastante al  $\text{C1}_{\text{GlcNAc}}$ , se decidió simular el ataque del Asp246 en el mutante R293A (Figura 93).



**Figura 93.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM para el mecanismo de ataque frontal (azul), y el para el ataque del Asp246 (marrón) en la mutante R293A, líneas continuas superiores. También se muestran las distancias más relevantes calculadas con el método QM(BP86/TZVP)/MM(CHARMM) utilizando la correspondencia de colores asignada a cada mecanismo. Para los puntos estacionarios del mecanismo se usan triángulos (OC) y círculos (R, IP y P).

En este caso, la formación del CGE ocurre con una barrera de 16.0 kcal/mol, un valor muy diferente del encontrado para la formación del CGE en la enzima nativa (26.80 kcal/mol). Además, es una barrera de energía 7.0 kcal/mol más baja que la formación del enlace glicosídico o ataque del GlcA en el mutante (Figura 93). Esto es porque la disociación UDP-GlcNAc ocurre muy rápidamente y GlcA se mantiene a más de 3.0 Å durante todo el proceso. En el caso del mecanismo de ataque frontal, la disociación del UDP-GlcNAc continúa siendo eficaz, sin embargo el ataque nucleofílico de GlcA no se cataliza adecuadamente porque Arg293 no está presente. Esto indica que la Arg293 no sólo es importante para anclar el sustrato aceptor sino que además, posee un rol catalítico en el paso final de la reacción. Es altamente probable que la Arg293 provea el ambiente electrostático adecuado para que GlcA se acerque al azúcar sin que la carga de  $\text{COO}^-_{\text{GlcA}}$  signifique un problema por la presencia del Asp246.

### 6.3.3.4 Doble Mutante D246A-R293A: Mecanismo de Ataque Frontal



**Figura 94.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM (líneas discontinuas superiores), y las distancias más relevantes a nivel QM(BP86/TZVP)/MM(CHARMM) para la enzima nativa (verde), y la doble mutante D246A-R293A(celeste). Por último, también se representan los puntos estacionarios del mecanismo a través de triángulos (OC) y círculos (R, IP y P).

La simulación del ataque frontal en el doble mutante da como resultado unas barreras de energía más altas, alcanzando valores de 35.0 kcal/mol. La doble mutante más que dar nueva información, sólo reafirma los comentarios anteriores demostrando que Asp246 y Arg293 son residuos catalíticos imprescindibles para que la enzima sea capaz de llevar a cabo la catálisis. Esto nos permite explicar los resultados experimentales de mutagénesis llevados a cabo por Pedersen *et. al.* que resultaron en mínima actividad transferasa (0.2 pmol/mg de proteína/h) para Arg293A y actividad transferasa indetectable, en el caso de Asp246A.<sup>70</sup>

De esta manera, podemos asignar los siguientes roles a esta pareja de residuos:

El Asp246 y la Arg293 son necesarios para llevar a cabo dos estrategias que utilizan las GT-ret para catalizar la reacción:

1. Permitir una alta disociación del sustrato dador, en este caso UDP-GlcNAc, a través de: el empujón nucleofílico. Para esto es necesario que un residuo nucleofílico se ubique en la cara  $\beta$  del azúcar, y puede tener una carga neta negativa, como Asp246, o ser neutro (Gln189 en LgtC); la estabilización de la carga del oxocarbenio será mejor si el residuo tiene una carga neta negativa. En

trabajos anteriores de nuestro grupo, ya se han hecho referencias a estos roles a través del concepto de catálisis asistida por nucleófilo.

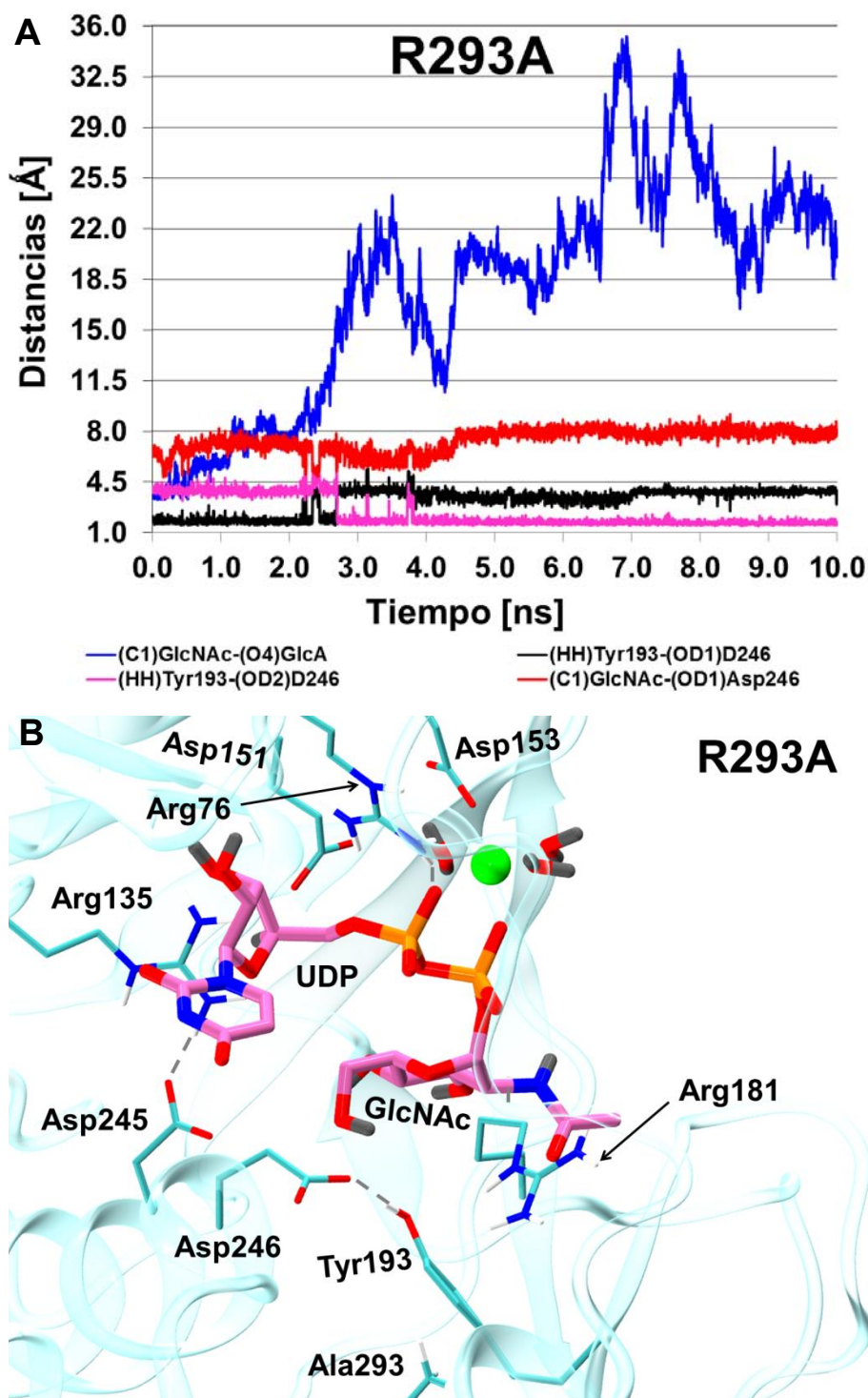
2. Permitir que el sustrato aceptor esté correctamente posicionado para el ataque a través de la interacción  $\text{HO4}_{\text{GlcA}}-\text{O}_P$ . La Arg293 promueve esta interacción al establecer interacciones electrostáticas con  $\text{COO}^-_{\text{GlcA}}$ , haciendo más fácil que la interacción intra-sustrato  $(\text{HO4-O6B})_{\text{GlcA}}$  se pierda (EXTL2-GalNAc) o no se forme (EXTL2-GlcNAc). Y Arg293 también parece facilitar el acercamiento del GlcA para el ataque por medio de la disminución de la repulsión Asp246 -  $\text{COO}^-_{\text{GlcA}}$ .

Por supuesto, el o los residuo(s) con carga neta positiva que interaccionan directamente con el grupo difosfato de UDP, así como las posibles interacciones sustrato-sustrato, son de vital importancia en la catálisis del mecanismo altamente disociado, estabilizando la disociación heterolítica del UDP-GlcNAc.

### 6.3.4 Simulaciones de Dinámica Molecular de las Enzimas Mutantes

Todos los experimentos anteriores son mutaciones *in silico* en las que no se ha considerado el efecto dinámico que cada mutación podría tener en la enzima. Debido a lo anterior, también se llevaron a cabo dinámicas moleculares de 10 ns, sólo para verificar la posición del sustrato dador en el sitio activo.

Los resultados de la mutante R293A muestran que la Arg293 es fundamental para la unión del sustrato aceptor pues si ésta no está, el aceptor se mantiene en el sitio activo menos de 1 ns (Figura 95.A y 95.B). La ausencia de la Arg293 también provoca cambios en la forma de interactuar de diferentes residuos, por ejemplo, se forma una interacción estable entre el Asp246 y la Tyr193 que se mantiene a lo largo de toda MD (Figura 95.A). De hecho, la Tyr193 ocupa el espacio que antes ocupaba el grupo guanidinio de la Arg293, sin embargo no es capaz de interactuar con GlcNAc como sí lo hacía la Arg293. Esto probablemente se relaciona con la alteración o pérdida de las interacciones más estables que forma el GlcNAc en la enzima nativa, pero que en esta mutante no aparecen. Se trata de las interacciones ión-dipolo entre  $OD1_{Asp151} \cdots HO3_{GlcNAc}$ , que se pierde a partir de los 2 ns; y  $HH22_{Arg135} \cdots O3_{GlcNAc}$ , que no se observa durante los 10 ns. Además, las interacciones  $OD2_{Asp245} \cdots HO4/HO6_{GlcNAc}$  también se pierden, aunque la interacción fuerte entre Arg135 y Asp245 se mantiene estable a lo largo de toda la MD; incluso la interacción  $HN2^i_{GlcNAc} \cdots OD1_{Asp151}$  se pierde antes de completar el primer ns, por lo que el azúcar se ha desplazado de su posición nativa. El UDP mantiene sus interacciones y en la Figura 95.B se muestra a la Arg76 como ejemplo de esto. Por último, así como la Tyr193 se desplazó de su posición nativa, la Arg181 (otro de los residuos que interactúa con el aceptor junto a Tyr193 y Arg293), sufre fluctuaciones debido a la entrada de aguas en el sitio activo y también aparece interactuando con el carbonilo del grupo NAc, aunque la interacción no es estable.



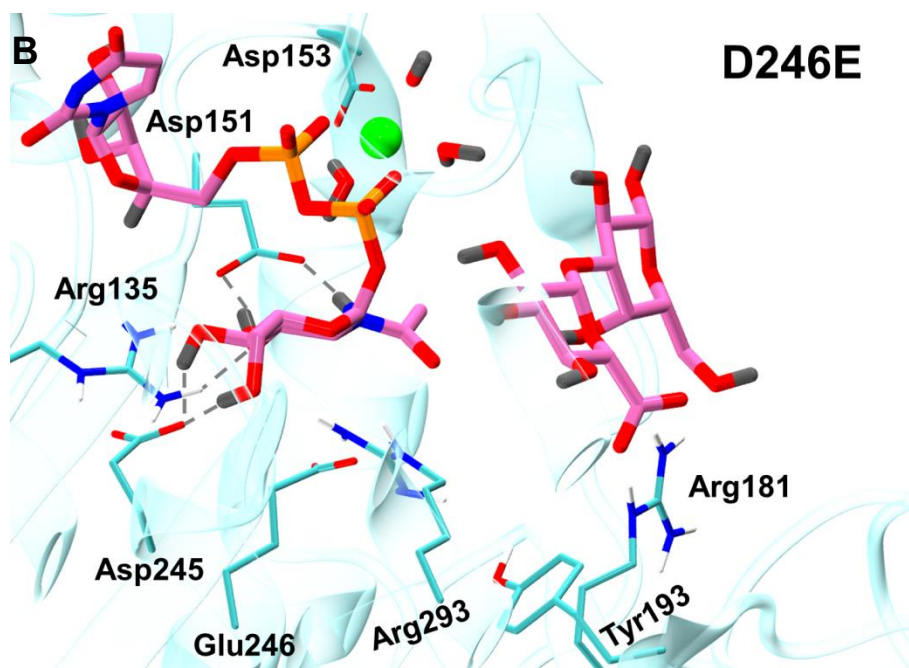
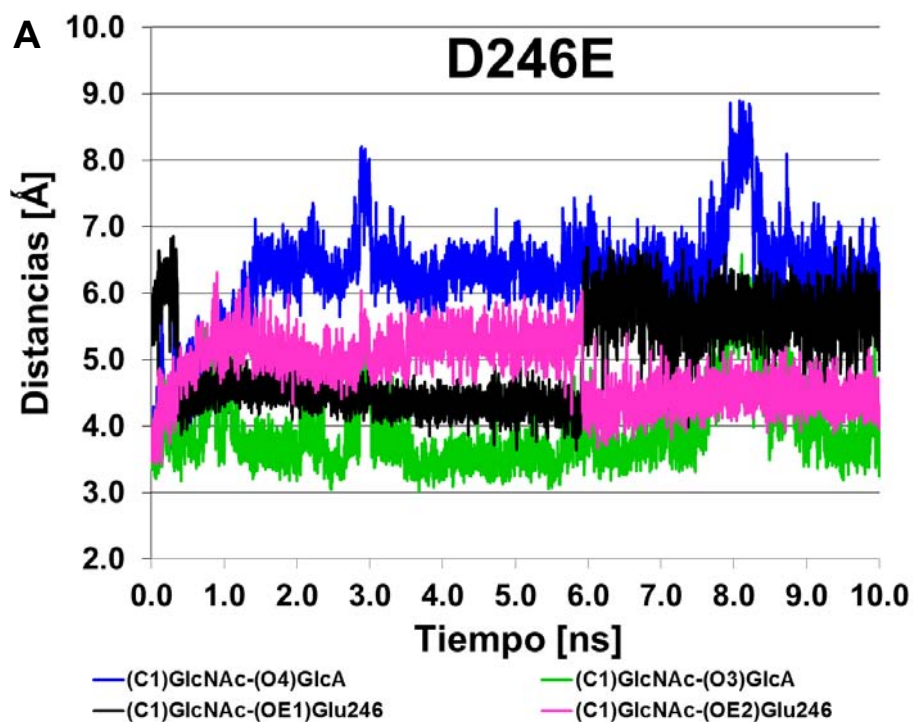
**Figura 95.** A: Variación de las distancias más relevantes para la reactividad que divergen completamente del comportamiento de la enzima nativa. B: Estructura representativa de la mutante R293A correspondiente al ns 9 de simulación de MD de 10 ns de producción. Además, se señala la interacción en la que participa Arg135-Asp245, la interacción Tyr193-Asp246 y la interacción de Arg76 con UDP (líneas discontinuas grises). El sustrato dador se muestra en color rosado.

Por último, cabe mencionar que en esta mutante la distancia  $OD1_{Asp246}-C1_{GlcNAc}$  se hace más larga ( $7.36 \pm 0.66$  Å), por lo que el Asp246 queda bastante alejado del carbono anomérico.



Esto se debe a que a que el Asp246 forma una interacción muy estable con Tyr193 ( $2.34 \pm 0.88 \text{ \AA}$  para  $OD2_{Asp246}$ ), la cual a su vez no interacciona con el sustrato dado a diferencia de la Arg293. Este es un dato adicional que indica porqué la Arg293 es importante para la catálisis ya que si ésta no está, es muy posible que el Asp246 no pueda posicionarse más cerca del carbono anomérico y por consiguiente, tampoco sería capaz de ejercer el empujón nucleofílico.

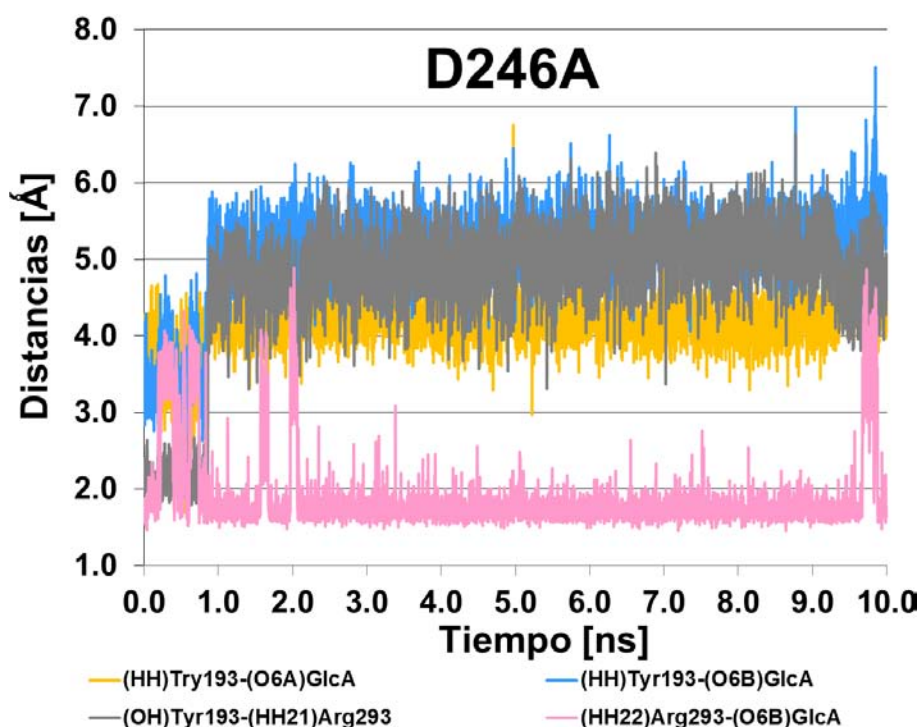
Continuando con la descripción de los resultados obtenidos para las mutantes, aunque el sustrato aceptor se une en las mutantes D246E y D246A, el modo de unión que establece el GlcA es diferente del observado en la enzima nativa.



**Figura 96.** A: Variación de algunas distancias importantes que involucran al Glu246 y al sustrato aceptor. B: Representación de la mutante D246E correspondiente al ns 9 de simulación de MD. También se señalan las interacciones que forma el GlcNAc en el sitio activo (líneas discontinuas grises) y los sustratos se dibujan en rosado.

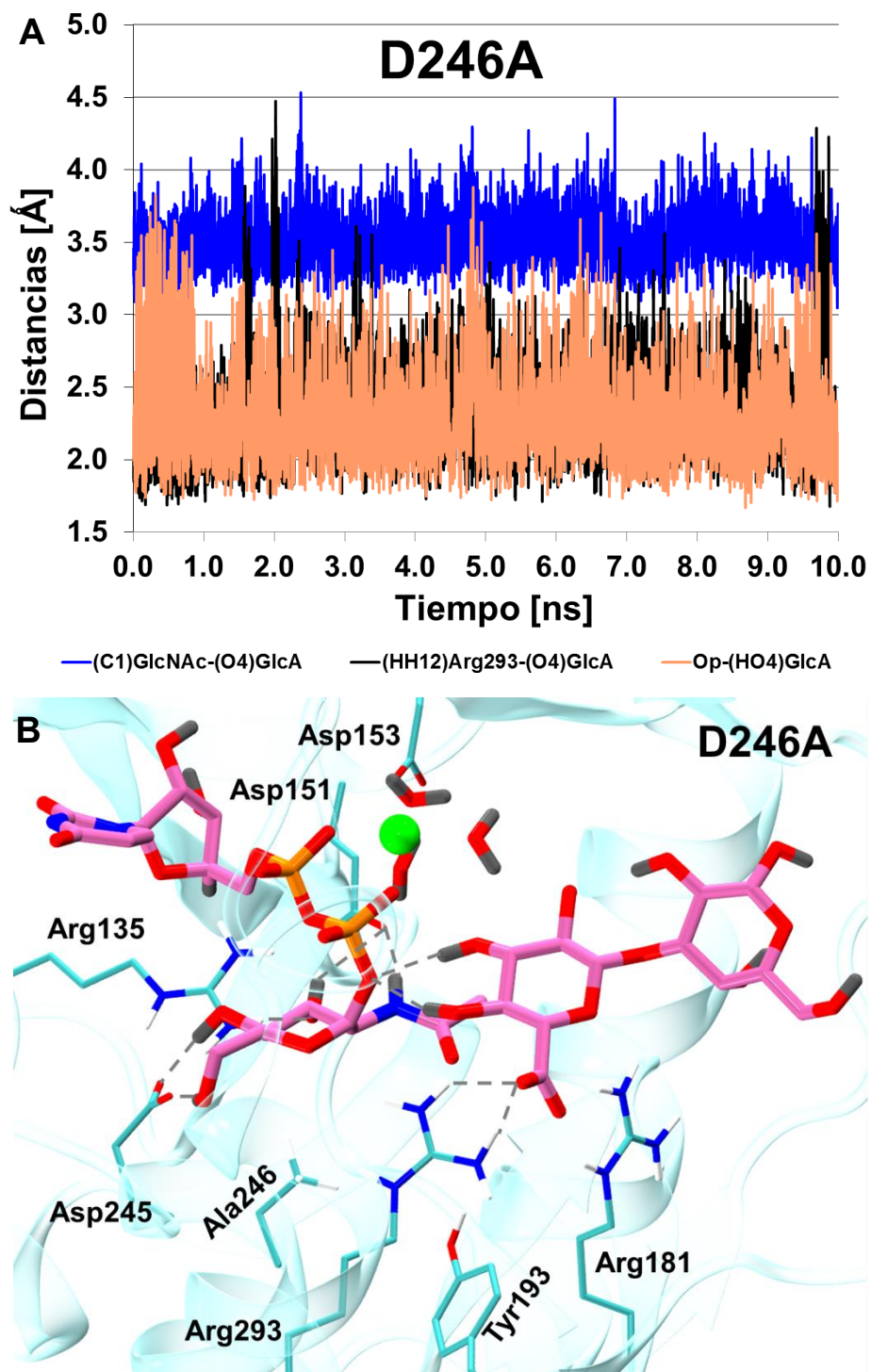
En la mutante D246E se produce un cambio en la orientación relativa entre sustrato dador y aceptor, de modo que ahora es  $O3_{GlcA}$  y no  $O4_{GlcA}$  el que queda en mejor posición para atacar al carbono anomérico. La distancia  $O4_{GlcA}-C1_{GlcNAc}$  tiene un valor promedio de  $6.24 \pm 0.49$  Å a lo largo de la MD del mutante (en la nativa es de  $4.83 \pm 0.82$  Å), superando los 6.0 Å aproximadamente a partir del primer ns, mientras el  $O3_{GlcA}$  se mantiene a una distancia promedio de  $3.92 \pm 0.43$  Å. Así pues, de ser la reacción viable a nivel de barrera energética, la mutante D246E podría dar un cambio en la regioespecificidad de la enzima. En este caso, no existen datos experimentales de este mutante para comprobar dicha posibilidad. La mutante D246E muestra que el Glu246 se ubica un poco más cerca del carbono anomérico ( $4.84 \pm 0.40$  Å para  $OE2_{Glu246}$ ) que lo observado para el Asp246 en la enzima nativa ( $5.13 \pm 0.45$  Å para  $OD2_{Asp246}$ ), lo cual es lógico pues la cadena lateral del Glu es 1.4 Å más larga que la del Asp246. En esta mutante las interacciones entre GlcNAc y la Arg135, el Asp151 y el Asp245 no se ven afectadas a diferencia de lo obtenido para R293A. Como se observa en la Figura 96, las interacciones que forma el sustrato aceptor en el sitio activo con residuos como la Tyr193, la Arg181 y la Arg293 también se han visto afectadas. De esta manera, el único residuo que permanece interaccionando con el GlcA es la Arg181, una interacción que también contribuye al cambio de orientación del aceptor. Al ser el Glu más largo, esto empuja a la Arg293 y a todo el aceptor de modo que el  $COO^-_{GlcA}$  se gira hacia la Arg181, orientando el  $O3$  hacia el carbono anomérico y no el  $O4$  (Figura 96).

En el caso de la mutante D246A, se observa que la Arg293 desplaza a la Tyr193 y no interacciona con ella (Figura 97), a diferencia de lo observado en D246E ( $OH_{Tyr193}-HH21_{Arg293} = 2.29 \pm 0.26$  Å) y en la WT ( $OH_{Tyr193}-HH21_{Arg293} = 2.24 \pm 0.23$  Å).



**Figura 97.** Variación de algunas distancias relacionadas con la Arg293, la Tyr193 y el GlcA a lo largo de la MD de la enzima mutante D246A.

Como puede observarse, Tyr193 ya no interacciona con GlcA y tampoco con la Arg293 (Figura 97). A falta del Asp246, la Arg293 parece moverse hacia el sustrato aceptor desplazando a la Tyr193 de su posición en la enzima nativa (Figura 98.B). Así, la Arg293 se coloca muy cerca del  $\text{COO}^-_{\text{GlcA}}$  y el  $\text{O4}_{\text{GlcA}}$  (Figura 97 y Figura 98.A, respectivamente). Por otra parte, el fragmento nucleofílico de GlcA ( $\text{O4}_{\text{GlcA}}$ ) se ubica más cerca del carbono anomérico, a un valor promedio de  $3.52 \pm 0.15 \text{ \AA}$  (Figura 98.A); esta distancia es más pequeña que la encontrada en la enzima nativa de  $4.83 \pm 0.82 \text{ \AA}$ , sin embargo la Arg293 forma una interacción bastante estable con  $\text{O4}_{\text{GlcA}}$  (Figura 98.A). A pesar de que tal interacción es efectiva para la unión del sustrato aceptor, no es adecuada para la formación del enlace glicosídico pues disminuye la nucleoficidad del  $\text{O4}_{\text{GlcA}}$ . Además, como se ha mostrado previamente, es de esperar que la ausencia del Asp246 haga que la formación de la especie oxocarbenio esté muy desfavorecida por la presencia de Arg293.



**Figura 98.** A: Variación de algunas interacciones en las que participa el nucleófilo GlcA en la mutante D246A a lo largo de la MD. B: Representación de una de la estructuras obtenidas de la MD que corresponde a los 9 ns. Los sustratos se dibujan en rosado y se incluyen las interacciones en las que participa GlcNAc y GlcA con varios residuos del sitio activo (líneas discontinuas grises).

Así, la mutación de Asp246 por Ala altera el modo de unión del sustrato aceptor provocando que la Arg293 interactúe cercanamente con GlcA. Esto causa que los sustratos estén más unidos que en la WT pero no en una orientación óptima para la catálisis (nótese en la

Figura 98.B, que nuevamente el O4<sub>GlcA</sub> ha bajado con respecto al plano del anillo, cosa que puede dificultar su ataque por la cara  $\alpha$ , y que el O3<sub>GlcA</sub> se encuentra en una posición que también podría ser reactiva). De todos modos, como se ha dicho, la presencia de la Arg293 pero no de un residuo como Asp/Glu que la contrarreste, hace difícil pensar que la disociación de UDP-GlcNAc pueda darse en este mutante, resultado que estaría en acuerdo con las observaciones experimentales. Finalmente, una de las conclusiones más interesantes de esta mutante es que la Arg293, que ocupa la cara  $\beta$  de GlcNAc, puede modular el modo de unión del sustrato aceptor.

## 6.4 Conclusiones

El trabajo realizado con esta enzima no ha permitido identificar la importancia que puede llegar a tener el o los residuos que se ubican en la cara  $\beta$  del azúcar. Gracias a las peculiares características que EXTL2 muestra en esta cara, una pareja Asp246-Arg293, hemos podido indagar en las características de esta zona y hemos asignado roles mecánicos específicos al Asp246 y a la Arg293. El Asp246 puede modular el mecanismo de reacción de dos maneras: debido a su carga negativa puede estabilizar sustancialmente las especies OC y además, puede empujar más lejos al UDP al acercarse al carbono anomérico, facilitando la disociación UDP-GlcNAc(GalNAc). Por otra parte, el Asp246 también puede compensar el efecto desfavorable de tener una Arg293 en la cara  $\beta$  del azúcar. A pesar de desfavorecer las especies intermedias de este mecanismo, la Arg293 es clave en la etapa de ataque frontal del GlcA pues por una parte facilita la formación de la interacción  $O_P-HO4_{GlcA}$  (en caso de que sea necesario), y por otro lado mantiene al  $O4_{GlcA}$  bien posicionado para el ataque nucleofílico a través de su interacción con  $COO^-_{GlcA}$ .

Las simulaciones de MD de las mutantes permitieron ver, por ejemplo, que el Glu246 no sería adecuado para reemplazar al Asp246 en el mecanismo de ataque frontal. Esto se debe a que el Glu246 modifica la orientación relativa de GlcA con respecto al sustrato dador desfavoreciendo la reactividad o incluso cambiando su regioespecificidad. Los resultados mecánicos de la mutante R293A indican que la formación del CGE sería un proceso factible, sin embargo la MD sugiere que la afinidad del sustrato aceptor por la enzima sería menor si la Arg293 no formara parte del sistema, hecho que podría explicar la escasa actividad transferasa del mutante determinado experimentalmente. Los resultados de la mutante D246A, la simulación *in silico* del mecanismo y la MD, muestran que la falta de Asp246 sube enormemente la barrera de energía (32.3 kcal/mol), lo cual posiblemente se deba a que el efecto desfavorable de la Arg293 no se ve compensado. Además, la MD muestra como la Arg293 sería capaz de interactuar fuertemente con  $O4_{GlcA}$  en ausencia del Asp246, lo cual deja ver que la mutación de cualquiera de estos residuos podría inactivar a la enzima, en acuerdo con los resultados experimentales. Por último, el mecanismo de ataque frontal en la doble mutante alcanza una barrera de energía de 35.6 kcal/mol, confirmando la importancia de esta pareja de residuos en la cara  $\beta$  del azúcar y también indicando que el gran aumento de energía se relaciona principalmente con la mutación del Asp246 más que con la mutación de la Arg293 (la barrera para la mutante D246A es de 32.3 kcal/mol). Para el doble mutante no se han realizado MDs aquí, pero es de esperar que la unión del aceptor se vea afectada como en el caso de los mutantes simples.

Finalmente, los resultados acumulados hasta ahora permiten notar la importancia que podría tener el sustrato aceptor en el tipo de estrategia que utiliza una GT-ret para catalizar el mecanismo de ataque frontal. Esto se nota principalmente en el residuo que ocupa la cara  $\beta$  azúcar, el cual puede tener la habilidad de estabilizar el desarrollo de carga en el azúcar y a la vez ejercer un empujón nucleofílico, catálisis asistida por nucleófilo. Adicionalmente, el residuo de la cara  $\beta$  puede interactuar con el sustrato aceptor y de este modo, podría estar involucrado en la correcta orientación del aceptor para la catálisis. Esta estrategia, se complementa con la catálisis asistida por sustrato y podría ser necesaria

cuando la disociación UDP-azúcar es más difícil. Los factores relacionados con el grado de dificultad de la disociación UDP-azúcar se discutirán en el siguiente capítulo.

# 7 Comparaciones Adicionales y Discusión General

## 7.1 Introducción

Los mecanismos catalíticos de GpgS y EXTL2 se han descrito en detalle en los capítulos anteriores. GpgS sigue un mecanismo de ataque frontal con la participación clave del grupo carboxilato de PGA ( $\text{COO}^-_{\text{PGA}}$ ), y de un grupo de moléculas de agua localizadas en el sitio activo. Estas aguas, además de interactuar con PGA, pueden contribuir a la estabilización de las especies oxocarbenio (OC) (Modelo 1). Debido a que las densidades electrónicas que daban cuenta de las aguas no pudieron ser asignadas con certeza, se generó un modelo *in silico* (Modelo 2) con el objetivo de investigar el mecanismo catalítico bajo esta nueva configuración del sistema. Nuevamente, se caracteriza un mecanismo de ataque frontal pero esta vez, como es de esperar, sólo el  $\text{COO}^-_{\text{PGA}}$  contribuye a la estabilización de las especies OC. De este modo, el hecho de que las barreras energéticas del Modelo 2 resultan ser un poco más altas que las del Modelo 1, apunta a que las aguas podrían contribuir a la estabilización de las especies OC a través de sus interacciones con la Glucosa, además de formar una pseudo-esfera de solvatación que acompaña los movimientos del  $\text{COO}^-_{\text{PGA}}$ , lo cual favorece las interacciones inter-sustratos. Por otro lado, la Leu209 ubicada en la cara  $\beta$  del azúcar es capaz de estabilizar discretamente las especies UDP----OC logrando una contribución máxima de -3.70 kcal/mol cuando la distancia  $\text{C1}_{\text{Glc}}-\text{O}_p$  es de 2.74 Å. Además, en el IP la Leu209 está a 2.98 Å del carbono anomérico, 1.1 Å más cerca que en el reactivo. Por último, también se evaluó el ataque nucleofílico de la  $\text{W}_2$ , lo cual nos permitió comprobar que la hidrólisis de UDP-Glc es un proceso muy poco favorecido cuya barrera supera las 30 kcal/mol.

Al contrario de GpgS, en EXTL2 se observa un residuo nucleofílico fuerte ocupando la cara  $\beta$  de GalNAc o GlcNAc en el complejo binario 1OMZ y 1ON6, respectivamente; aunque el residuo se encuentra a 5.2 y 5.1 Å del carbono anomérico, respectivamente. Además, en esta misma cara la Arg293 está a una distancia más corta del carbono anomérico en comparación al Asp246. En cuanto a los mecanismos de reacción, en el caso de EXTL2-GalNAc se encontró un intermedio poco profundo UDP----OC ( $\text{IP}_\beta$ ) en el que Asp246 y GalNAc se han acercado bastante, desde 5.43 Å en el reactivo hasta 3.16 Å en el  $\text{IP}_\beta$ . Este notable acercamiento motivó la investigación del mecanismo de formación del CGE usando al Asp246 como nucleófilo. De esta manera, se caracterizaron dos mecanismos. El primero es el mecanismo de ataque frontal, en el que el Asp246 tiene un rol clave en la estabilización de las especies oxocarbenio; y un segundo mecanismo, en el que el Asp246 actúa como nucleófilo. En ambos casos la disociación UDP-GalNAc genera especies estructural y energéticamente muy similares, posiblemente porque en esta etapa ambas coordenadas de reacción están describiendo el mismo proceso químico, la disociación UDP-GalNAc. Distinta es la etapa de ataque nucleofílico, en la que el ataque del GlcA resulta 5.5 kcal/mol más favorecido que el ataque del Asp246; una diferencia de energía



que se traduce en una  $k_{cat}$  ( $s^{-1}$ ) 4 órdenes de magnitud más rápida para el ataque del nucleófilo GlcA a 27 °C. Probablemente, esto se debe a que mientras el GlcA es activado por el UDP que actúa como base catalítica, el Asp246 interacciona con la Arg293, lo cual disminuye su habilidad nucleofílica. Por último, la contribución electrostática de la Arg293 desfavorece el mecanismo de reacción, sin embargo el Asp246 compensa este efecto desfavorable a través de sus interacciones cercanas con las especies oxocarbenio.

En el caso de EXTL2-GlcNAc, también se caracteriza un mecanismo de ataque frontal en el que Asp246 tiene un rol importante en la estabilización de las especies oxocarbenio a lo largo del mecanismo. Curiosamente, aunque el Asp246 estaba un poco más cerca del carbono anomérico ( 4.5 Å en los reactivos), las barreras para su ataque nucleofílico resultan bastante menos favorecidas que las del mecanismo de ataque frontal, 12.8 y 9.5 kcal/mol más altas en R-495 y R-350, respectivamente. Adicionalmente, los resultados obtenidos en base a los modelos de EXTL2-GlcNAc nos permiten concluir que el Asp246 no sólo es capaz de estabilizar las especies oxocarbenio sino también, de ejercer un empujón nucleofílico facilitando un mecanismo más disociado, esto es, especies  $UDP^{\delta-}$  y  $OC^{\delta+}$  más separadas. Esto es muy importante pues a mayor grado de disociación, menos impedida estará la entrada del GlcA y, por consiguiente, más fácilmente se producirá el ataque nucleofílico.

Por otra parte, las cuatro enzimas mutantes que han sido estudiadas nos permiten consolidar las hipótesis ya planteadas acerca del rol del Asp246, y también permiten evaluar más detalladamente el papel de la Arg293 en el mecanismo catalítico. Además, la mutante R293A muestra que Arg293 desfavorece la habilidad nucleofílica del Asp246, dado que éste ataca fácilmente en esta mutante. Además, la mayor energía que se necesita para la etapa de ataque nucleofílico (Figura 93), deja ver que la Arg293 puede ayudar a mantener al sustrato aceptor bien posicionado para el ataque a través de sus interacciones con  $COO^-_{GlcA}$ , encontrado así una explicación para los resultados experimentales de mutagénesis que indican que EXTL2 sólo mantiene una actividad transferasa mínima cuando Arg293 es mutada por Ala. Evidencia adicional de la importancia de la Arg293 la brinda la simulación MD de la mutante R293A, cuyos resultados apuntan a que la Arg293 es importante para la unión del sustrato aceptor y por consiguiente, para la correcta orientación y ataque nucleofílico de GlcA.

En resumen, EXTL2 cataliza un mecanismo de ataque frontal con la participación clave de dos residuos ubicados en la cara  $\beta$  del azúcar, el Asp246 y la Arg293 en la etapa de disociación UDP-azúcar y ataque nucleofílico, respectivamente. GpgS también cataliza un mecanismo de ataque frontal con la participación clave del  $COO^-_{PGA}$  y un grupo de moléculas de agua en la cara  $\beta$  del azúcar (en el caso de estar presentes). Se debe tener en cuenta que la interacción entre hidroxilo nucleofílico (Nuc) y  $O_P$ , de aquí en adelante  $O_P-Nuc$ , está presente en todos los casos estudiados en esta tesis (catálisis asistida por sustratos), sin embargo nos referiremos a ella más tarde en esta discusión ya que es una estrategia transversal para todas las GTs-ret y ahora nos enfocamos en las diferencias.

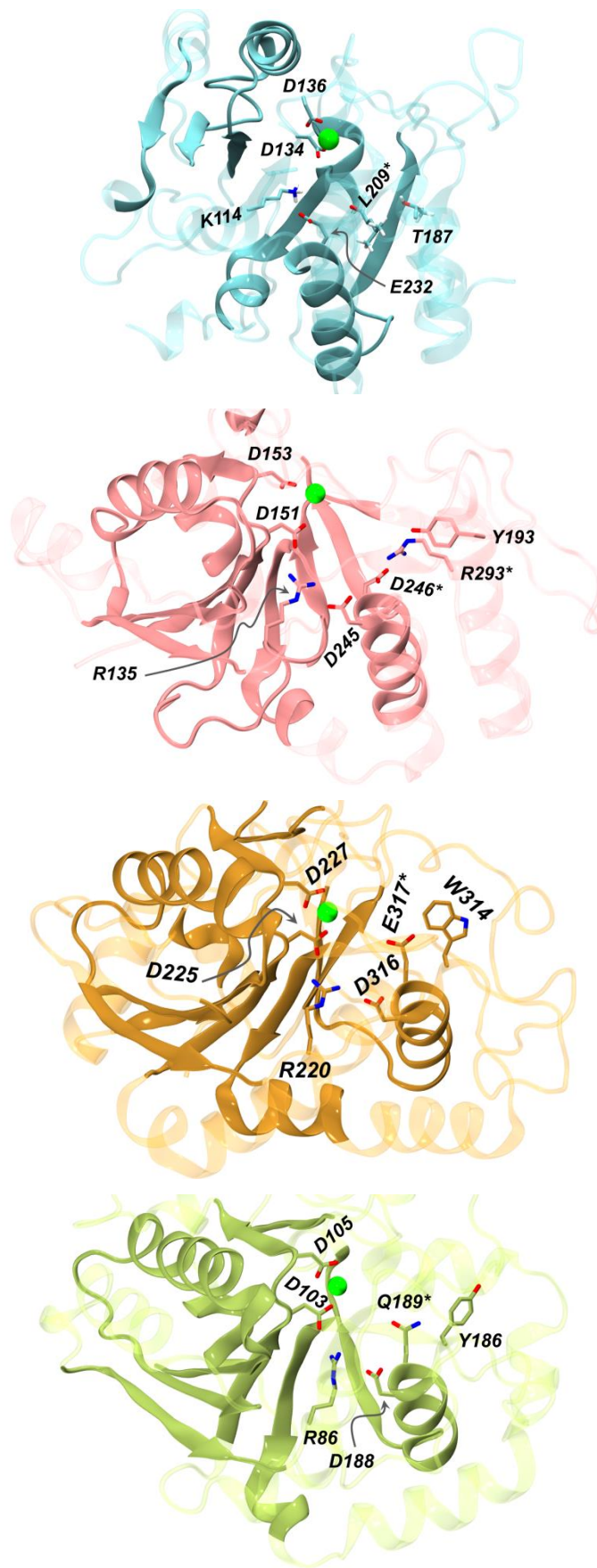
## 7.2 Metodología y Modelos

Las disociaciones de los distintos sustratos dadores se llevaron a cabo usando a EXTL2, GpgS,  $\alpha$ 3GalT<sup>124</sup> y LgtC.<sup>125</sup> Las primeras dos enzimas se basaron en los modelos correspondientes a R1<sub>A</sub> y al Modelo 1, respectivamente. La construcción de estos modelos así como los detalles de los métodos QMMM utilizados han sido descritos en el capítulo 4 (GpgS) y 5 (EXTL2-GalNAc).<sup>285</sup> También se utilizaron los modelos de  $\alpha$ 3GalT<sup>124</sup> y LgtC<sup>125</sup>, los cuales se usaron como puntos de partida para la optimización a nivel QM(BP86/TZVP)/MM(CHARMM) manteniendo la misma partición QMMM de los trabajos originales. La coordenada de reacción para la disociación de los sustratos dadores corresponde a  $RC = d(C1_{azúcar}-O_P)$  y se realizó en pasos de 0.1 Å hasta alcanzar los 3.0 Å. También se recalcularon las energías al nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM.

La manera de realizar el cálculo de las contribuciones energéticas para residuos de interés también ha sido descrita previamente en las secciones metodológicas de los capítulos 4, 5 y 6, por lo que no es necesario indicar metodologías nuevas respecto a estos cálculos.

## 7.3 Resultados y Discusión

Además de las enzimas estudiadas en esta tesis, es interesante comentar los resultados mecánicos de dos GTs-ret estudiadas en una tesis anterior del grupo. Una de ellas es LgtC, la cual cataliza un mecanismo de ataque frontal y la formación del CGE no parece ser una posibilidad en la enzima nativa. Distinto es el caso de  $\alpha$ 3GalT, para la cual se caracterizó un mecanismo de ataque frontal y también un mecanismo de doble desplazamiento, los cuales podrían estar compitiendo en la catálisis, o incluso podría actuar únicamente el mecanismo de doble desplazamiento según el trabajo de Rojas-Cervellera *et al.* Los residuos que se observan en la cara  $\beta$  de estas enzimas corresponden a Gln189 en el caso de LgtC y a Glu317 en el caso de  $\alpha$ 3GalT. Todas estas GTs-ret poseen un plegamiento GT-A y comparten algunas características dentro del sitio activo como el cofactor metálico, el motivo DXD y la pareja de residuos Lys-Glu o Arg-Asp a la que nos hemos referido a lo largo de la tesis. Debido a estas características, se decidió superponer los complejos de Michaelis que han sido modelizados en nuestro grupo. Para hacer esto, se usó la matriz BLOSUM62 del módulo de alineamiento MatchMaker de Chimera USCF (Figura 99, 100 y 101).



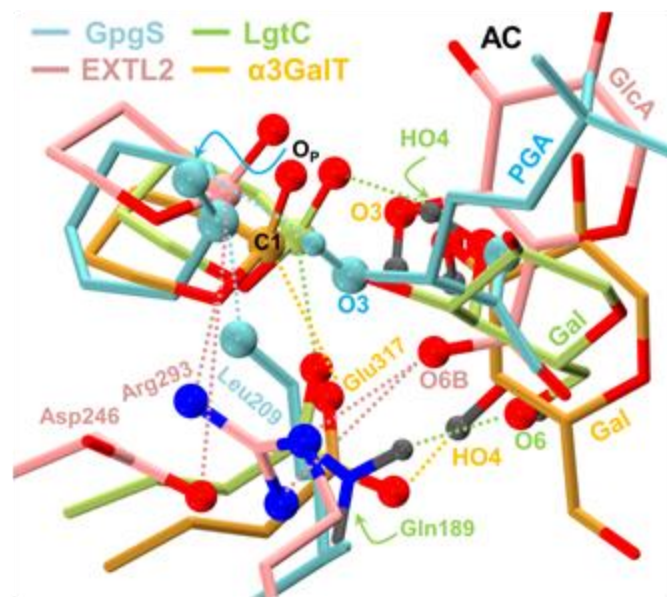
**Figura 99.** Representación de las estructuras tridimensionales de GpgS (celeste), EXTL2 (rosado),  $\alpha$ 3-GalT (naranja) y LgtC (verde); el metal se indica a través de una esfera verde. Se resaltan las zonas estructuralmente análogas de las 4 enzimas y también algunos residuos (barras) importantes: el primer y tercer residuo del motivo

D<sub>1</sub>XD<sub>3</sub>, relacionado con la unión del metal (D<sub>3</sub>) y la unión del azúcar (D<sub>1</sub>); la pareja de residuos conservados que participan de la unión del azúcar y que en GpgS corresponden a la pareja K-L y al par R-D en EXTL2, α3-GalT y LgtC; un residuo relacionado sólo con la unión del sustrato aceptor, T187 (GpgS), Y193 (EXTL2), W314 (α3GalT) y Y186 (LgtC); y los residuos que se marcan con un asterisco corresponden a los que ocupan la cara β del azúcar. Las estructuras son complejos de Michaelis modelados en esta tesis (GpgS) y EXTL2 (Mendoza *et.al.*, 2016); y en el caso de α3-GalT y LgtC ver referencias (Gómez *et.al.*, 2013) y (Gómez *et.al.*, 2012), respectivamente.

Como puede observarse en la Figura 99, la pareja Asp245/**Asp246** en EXTL2 y Asp316/**Glu317** en α3GalT se superponen perfectamente con la pareja Asp188/**Gln189** en LgtC (el residuo en *negrita* ocupa la cara β del azúcar). En el caso de GpgS, el Glu232 ocupa una posición análoga a la del Asp245, Asp316 y Asp188, y todos son residuos que interactúan con el hidroxilo 4 y/o 6 de los correspondientes azúcares; aunque no se encuentra un residuo que coincida con Asp246, Glu317 o Gln189. Por otra parte, a pesar de que la Lys114 en GpgS, Arg135 en EXTL2, Arg220 en α3GalT y Arg86 en LgtC ocupan posiciones análogas interactuando con el hidroxilo 3 del azúcar en todas estas GTs-ret, no son residuos que el algoritmo pueda alinear pero se indican igualmente. Además, todas las enzimas poseen el motivo DXD y hojas β paralelas típicas de dominios tipo-Rossmann. En EXTL2 y GpgS, la Arg293 y la Leu209 son residuos que pertenecen a una región estructuralmente no conservada entre estas enzimas, sin embargo ocupan parte de la cara β del azúcar. A diferencia de las otras enzimas, en EXTL2 hay dos residuos en la cara β del azúcar, el Asp246 que es análogo a los residuos nucleofílicos de LgtC (Gln189) y α3GalT (Glu317), y la Arg293 que interactúa fuertemente con el Asp246. La presencia de Arg293 se asocia a la naturaleza del sustrato aceptor y tanto Asp246 como Arg293, son residuos altamente conservados a lo largo de la familia GT64 a la que pertenece EXTL2. (Mendoza *et.al.*, 2016)

Adicionalmente, al ubicarse en la zona de interacción del sustrato aceptor, el segmento base de la Arg293 se puede relacionar con los segmentos base de la Thr187 en GpgS, Trp314 en α3GalT y Tyr186 en LgtC. Notar que aunque en las tres últimas enzimas estos residuos interactúan únicamente con el aceptor, en EXTL2 la Arg293 también interactúa con el azúcar y con el Asp46 en la cara β del azúcar.

Si ahora comparamos más de cerca los sustratos y los residuos en el sitio activo de las estructuras superpuestas, se pueden comentar algunas diferencias importantes respecto a las características que modulan la cara β de estas GTs-ret (Figura 100).



**Figura 100.** Representación de barras y esferas para GpgS (celeste), EXTL2 (rosado), LgtC (verde) y  $\alpha$ 3GalT (naranja). En cada caso se representan: el azúcar del sustrato dador UDP-azúcar (sólo el anillo), parte del sustrato aceptor (el fragmento reactivo), y los residuos estructuralmente relacionados que ocupan la cara  $\beta$  del azúcar. Los hidrógenos se colorean de gris y se señalan las interacciones en las que participan los residuos de la cara  $\beta$ . Notar que los nombres de átomos y residuos, así como las líneas discontinuas siguen los códigos de color de cada enzima, con la excepción del  $O_p$ , el C1 y el sustrato aceptor (AC) que se indican de manera general.

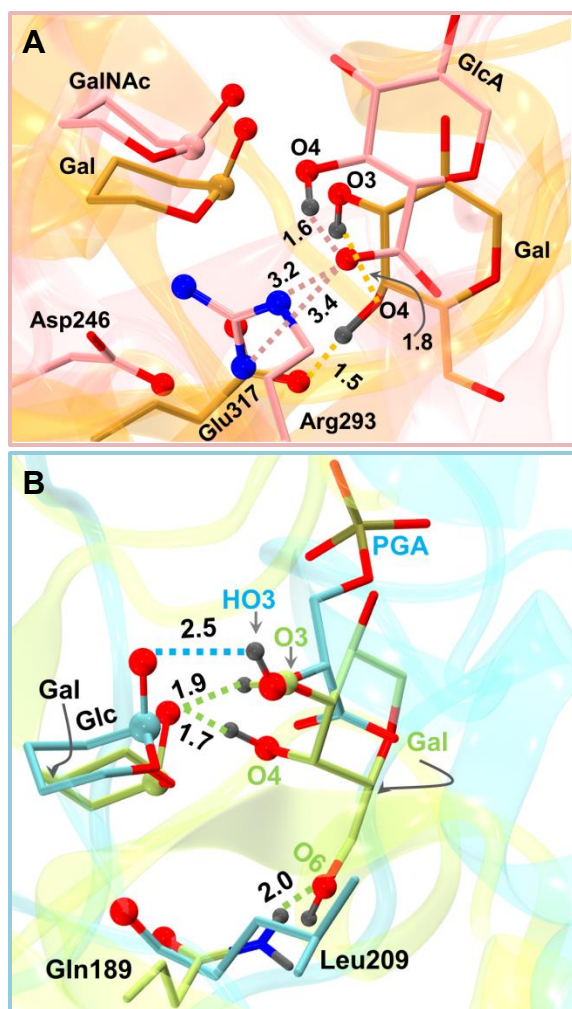
En LgtC, la Gln189 interacciona con el  $O6_{Gal}$  del sustrato aceptor (AC, 2.0 Å) y a la vez se ubica a 3.4 Å del carbono anomérico (C1), una distancia que disminuye hasta 2.9 Å en el TS2. (Gómez *et.al.*, 2012) En el caso de  $\alpha$ 3GalT, el Glu317 puede interaccionar con el  $HO4_{Gal}$  del AC (1.5 Å) y a la vez, se ubica a 4.3 Å del C1 del azúcar; una distancia que disminuye hasta 2.5 Å en el IP. (Gómez *et.al.*, 2013). En GpgS, la Leu209 se ubica a 4.3 Å del C1 en el Modelo 1 (4.1 Å en Modelo 2), sin embargo no interacciona con el AC. Es importante señalar que nuestros resultados del Modelo 1 sugieren que las posibles aguas podrían hacer de puente entre ambos sustratos, especialmente la  $W_2$  y la  $W_4$ . Por último, en EXTL2 hay dos residuos ocupando la cara  $\beta$  del azúcar; por una parte está el Asp246 que se ubica a 5.4 Å del  $C1_{GalNAc}$  (Mendoza *et.al.*, 2016) (4.6 Å de  $C1_{GlcNAc}$  en R-495), y como puede verse en la Figura 100, su cadena lateral posee un carbono menos que las cadenas laterales de Gln189 (LgtC) o Glu317 ( $\alpha$ 3GalT). Además está la Arg293 que se ubica a 3.0 y 3.5 Å del  $O5$  y el C1 de GalNAc, respectivamente (3.2 y 3.9 Å, respectivamente en EXTL2-GlcNAc, R-495). Simultáneamente, la Arg293 tiene a dos nitrógenos del grupo guanidinio a 3.4 y 3.2 Å del  $O6B_{GlcA}$  del AC (3.8 y 3.9 Å para los nitrógenos análogos en EXTL2-GlcNAc, R-495). Estas características nos permiten concluir que las cadenas laterales de la pareja Asp246-Arg293 juegan un papel análogo al de las cadenas laterales de Gln189 y Glu317 en LgtC y  $\alpha$ 3GalT, respectivamente. La posición que ocupa Gln189 y Glu317 en la cara  $\beta$  del correspondiente azúcar les permite interaccionar con ambos sustratos, y a la vez contribuir a la estabilización de las especies OC. En EXTL2 la pareja Asp246-Arg293 puede llevar a cabo esta doble tarea durante el mecanismo, es decir, la Arg293 interacciona con ambos sustratos y el Asp246 estabiliza las especies oxocarbenio. En resumen, EXTL2 usa una

pareja de residuos para llevar a cabo las tareas que en LgtC y  $\alpha$ 3-GalT realizan Gln189 y Glu317, respectivamente. Esto es:

1. Mantener el sustrato dador y el sustrato aceptor en posiciones cercanas y por consiguiente, adecuadas para la catálisis.
2. Contribuir a la estabilización de las especies oxocarbenio durante la reacción de transferencia.

En el caso de GpgS, aunque la Leu209 es capaz de estabilizar discretamente las especies oxocarbenio (al igual que Gln189 en LgtC), no se encuentra en una posición que le permita interactuar con el aceptor. En este caso, el sustrato aceptor no es un carbohidrato y posiblemente los 7 residuos con los que interacciona el PGA en el sitio activo (Arg185, Arg167, Thr187, Gly183, Gly184, Asn246 y His258) le permiten mantenerse en la cavidad en una posición cercana al sustrato dador. Cabe mencionar que el número de residuos con los que interactúan la lactosa en LgtC y  $\alpha$ 3-GalT, y GlcA $\beta$ 1-3Gal (EXTL2) es más pequeño; 5 residuos en LgtC (Tyr186, Phe132, Asn130, Gln189 e His78); 6 residuos en  $\alpha$ 3GalT (Trp314, Glu317, Tyr250, Thr259, Trp249, Gln247); y 5 residuos en EXTL2 (Trp284, Arg181, Tyr193, Arg293, Phe290). Además, en los modelos de Reactivo de GpgS está la interacción HO<sub>3</sub>PGA-O<sub>P</sub> (catálisis asistida por sustratos), lo cual también ha de contribuir a la correcta orientación del sustrato aceptor (esta conclusión aplica tanto al Modelo 2 como al Modelo 1).

La pregunta que surge en este punto es: ¿por qué en algunas GTs-ret se encuentra un residuo nucleofílico más fuerte en la cara  $\beta$  y en otras no? Esto está relacionado con la facilidad con la que ocurre la disociación UDP-azúcar. Creemos que esto se relaciona básicamente con dos aspectos: i) la facilidad con que el nucleófilo de la reacción (el aceptor) pueda establecer la interacción con el O<sub>P</sub> (catálisis asistida por sustratos); y ii) la propia naturaleza química del aceptor. Los modelos proporcionados por LgtC,  $\alpha$ 3GalT, y EXTL2 nos permiten llegar a esta conclusión pues las tres enzimas poseen un residuo estructuralmente análogo en la cara  $\beta$  (Figura 99, ver \*) que puede interactuar al mismo tiempo con el azúcar y con el aceptor. En el caso de LgtC y  $\alpha$ 3GalT, este residuo también es capaz de interactuar favorablemente con las especies oxocarbenio. Contrariamente, en EXTL2 el residuo en cuestión no puede estabilizar las especies oxocarbenio porque se trata de la Arg293 y es por eso que el Asp246 también es necesario. Además, como se ha comentado antes, es esencial porque compensa el efecto desfavorable de la Arg293. En otras palabras, parece ser que un residuo nucleofílico fuerte como Asp o Glu sería necesario en la cara  $\beta$  del azúcar cuando hay factores que hacen más difícil la disociación UDP-azúcar. Un factor que afecta tal proceso (aspecto i) es la habilidad que un determinado Nuc tenga para formar la interacción con O<sub>P</sub> ya en los reactivos (O<sub>P</sub>-Nuc). En EXTL2 y  $\alpha$ 3GalT el hidroxilo nucleofílico (Nuc) puede formar una interacción intra-molecular con el O6B de GlcA y el O4 de Gal, respectivamente (Figura 101.A), por lo que la interacción O<sub>P</sub>-Nuc podría ser más difícil de formarse en el reactivo. Además, en EXTL2 está la Arg293 desfavoreciendo sustancialmente la formación de las UDP----OC. El hecho de que justamente en estas dos enzimas se encuentre un residuo nucleofílico fuerte en la cara  $\beta$  del azúcar, apoya nuestra hipótesis acerca de que el grado de dificultad de la disociación UDP-azúcar, modulará el tipo de residuo que aparece justo debajo del azúcar.

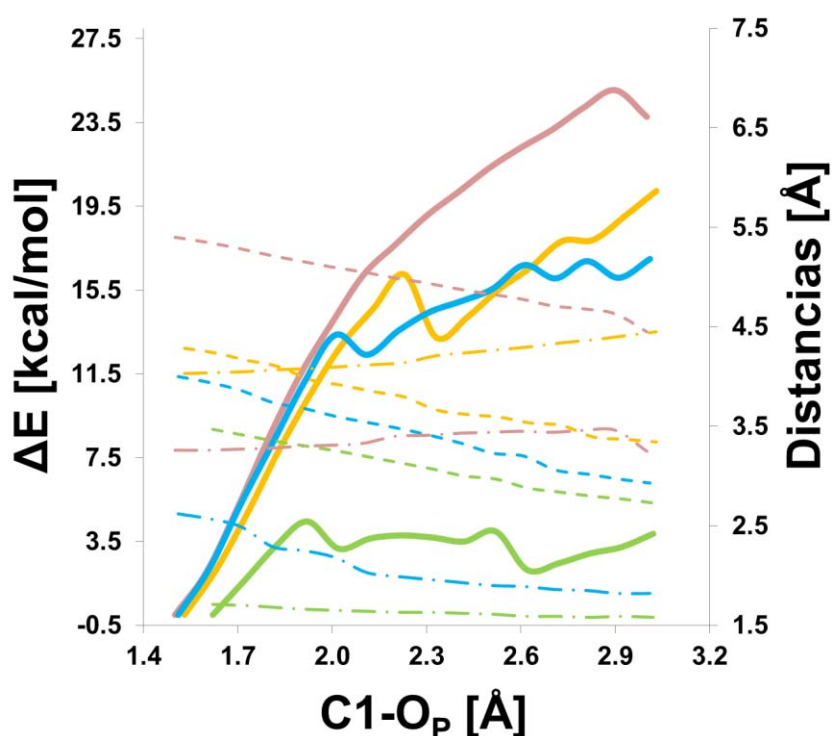


**Figura 101.** Representación de los sustratos, los residuos de la cara  $\beta$  del azúcar y las interacciones en las que participan (líneas discontinuas) para A: EXTL2 (rosado) y  $\alpha$ 3GalT (naranja) y B: GpgS (celeste) y LgtC (verde). Se indican las interacciones intra-GlcA (EXTL2) o intra-Gal ( $\alpha$ 3GalT), junto a la interacción  $\text{HO4}_{\text{Gal}}/\text{HO3}_{\text{Gal}}\text{-OP}$  (LgtC) y  $\text{HO3}_{\text{PGA}}\text{-OP}$  (GpgS). El  $\text{O3}_{\text{Gal}}$  de LgtC es de color verde para diferenciarlo del  $\text{O3}_{\text{PGA}}$  (rojo). También se incluyen las interacciones en las que participa el residuo nucleofílico en cada caso.

En EXTL2 la Arg293 puede interaccionar con el carboxilato de GlcA, y en  $\alpha$ 3GalT el Glu317 puede interaccionar con  $\text{HO4}_{\text{Gal}}$ . El Glu317 y el Asp246 son residuos nucleofílicos fuertes, aunque su habilidad como nucleófilo debería verse reducida por la fuerte interacción que forma con el aceptor en el caso de Glu317, y con la Arg293 en el caso de Asp246. Sin embargo, sus capacidades para estabilizar las especies oxocarbenio o ejercer el empujón nucleofílico no deberían verse perjudicadas por esto. Notar que en  $\alpha$ 3GalT y EXTL2 no se establece la interacción  $\text{O}_\text{P}\text{-Nuc}$  en el reactivo debido a que el Nuc forma una interacción intra-molecular:  $\text{O4}_{\text{Gal}}\text{-HO3}_{\text{Gal}}$  y  $\text{O6B}_{\text{GlcA}}\text{-HO4}_{\text{GlcA}}$ , respectivamente. Aunque se debe recordar que en EXTL2.GlcNAc esta interacción está presente, lo cual indica que en esta enzima el principal factor que dificulta la disociación UDP-azúcar está relacionado con la Arg293. Distinto es el caso de GpgS y LgtC, en las que la interacción  $\text{O}_\text{P}\text{-Nuc}$  está presente desde el principio, es decir, se encuentra en el reactivo. En estas dos enzimas también se encuentra

un residuo que podría tener cierto carácter nucleofílico (aunque mucho más débil) en la cara  $\beta$  del azúcar, la Leu209 y la Gln189 en GpgS y LgtC, respectivamente, sin embargo la estabilización que estos residuos nucleofílicos pueden dar es menor porque son neutros. Es presumible que la disociación UDP-azúcar será más fácil o energéticamente menos costosa en el caso de GpgS y LgtC, pues ambas poseen un residuo nucleofílico que puede estabilizar discretamente las especies OC y además, tienen la interacción  $O_P$ -Nuc que contribuirá a la estabilización del UDP durante la disociación; además, en LgtC también se forma la interacción inter-sustratos  $HO3_{Gal}-O_P$  (1.9 Å, Figura 101.B). Para poder verificar cualitativamente estas ideas, se han calculado las contribuciones electrostáticas de los residuos de la cara  $\beta$  del azúcar y del aceptor a lo largo de la disociación UDP-azúcar para cada enzima. (Figura 102 y 103). Antes de entrar en la discusión con respecto a lo anterior, vale la pena recordar que Gln189 interacciona cercanamente con el  $O6_{Gal}$ , en cambio la Leu209 no interacciona con PGA. Posiblemente, esto es una consecuencia de la distinta naturaleza de los sustratos aceptores en LgtC y GpgS (Figura 101.B). De hecho, de los sistemas aquí comparados, vemos que en los casos en que el sustrato aceptor es un azúcar (varios grupos hidroxilos que podrían atacar al carbono anomérico), los residuos de la cara  $\beta$  del monosacárido transferido interaccionan con el aceptor. Esto nos puede llevar a sugerir que esta interacción ayuda al aceptor a adoptar la buena orientación para la regioespecificidad mostrada por la enzima.

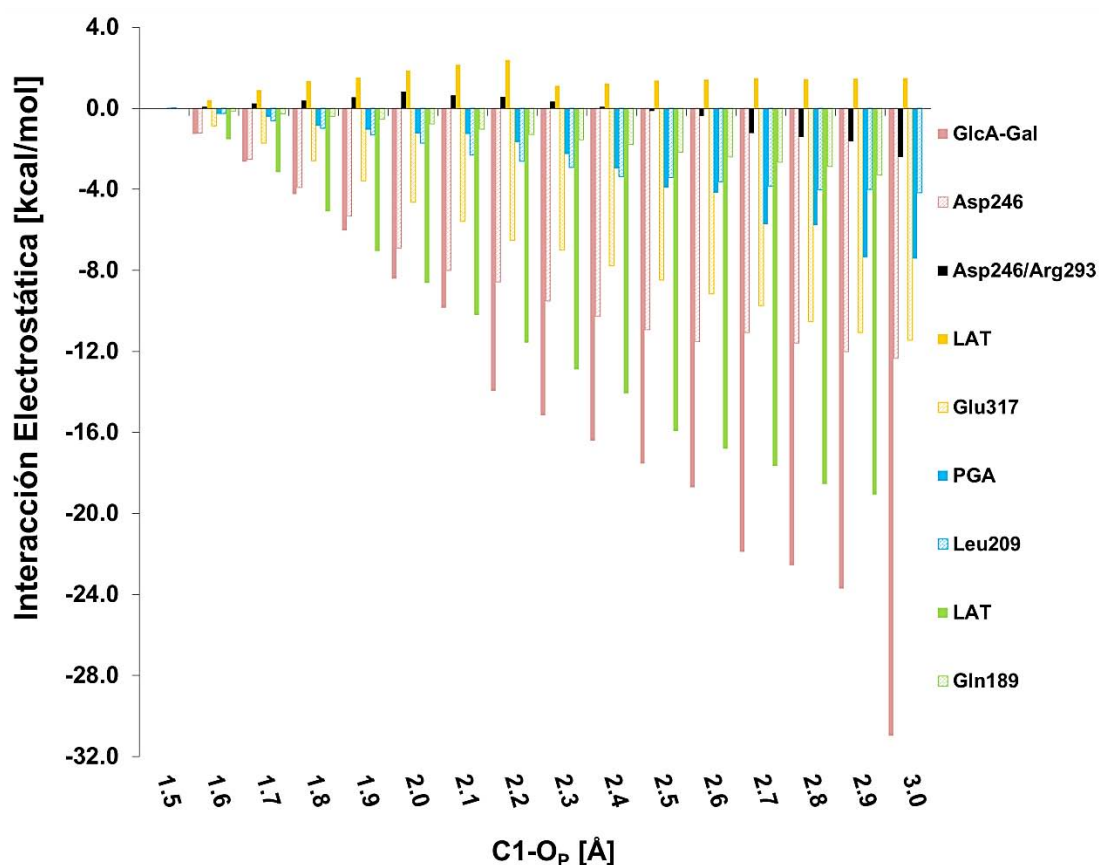
La disociación de EXTL2 resulta ser energéticamente la más costosa, lo cual es lógico porque hay una Arg293 en la cara  $\beta$  que desfavorece sustancialmente la formación de las especies oxocarbenio y además no está presente la interacción  $O_P$ -Nuc. Hay que tener en cuenta que en la reacción de ataque frontal la interacción  $O_P$ -Nuc se forma muy tempranamente (ver capítulo de EXTL2:GalNAc), pero que aquí, utilizando solo como coordenada la distancia UDP-GalNAc, esta interacción no se forma.





**Figura 102.** Curvas de energía potencial a nivel QM(M05-2X/TZVP)/MM/QM(BP86/TZVP)/MM para la disociación de UDP-GalNAc en EXTL2 (línea rosada), UDP-Glc en GpgS (línea celeste), UDP-Gal en  $\alpha$ 3GalT (línea amarilla) y UDP-Gal en LgtC (línea verde). La coordenada de reacción se define como  $RC = d(C1-O_P)$ . También se indican las distancias  $O_P$ -Nuc (líneas discontinuas de rayas y puntos), y las distancias de los residuos nucleofílicos Asp246 (EXTL2), Leu209 (GpgS), Glu317 ( $\alpha$ 3GalT) y Gln189 (LgtC), con respecto al carbono anomérico (líneas discontinuas de rayas).

En este caso, el mutuo acercamiento de 1 Å entre C1 y el Asp246 puede generar una contribución electrostática importante que estabiliza las especies oxocarbenio (catálisis asistida por nucleófilo) (Figura 103). En el caso de  $\alpha$ 3GalT y GpgS, las curvas de energía potencial para la disociación del sustrato dador son semejantes, alcanzando valores máximos de 20 y 17 kcal/mol, respectivamente. Esto es curioso pues en GpgS la interacción  $HO3_{PGA}-O_P$  estabiliza la carga de UDP (Figura 102), pero no en  $\alpha$ 3GalT pues no se llega a formar la interacción  $HO3_{Gal}-O_P$  y tampoco hay otra interacción entre sustrato dador y aceptor. Además, la distancia  $HO3_{Gal}-O_P$  incluso aumenta ligeramente a lo largo de la disociación desde 4.05 Å hasta 4.45 Å (Figura 102). Nuevamente, hay que tener en cuenta que en la coordenada de reacción por ataque frontal en  $\alpha$ 3GalT esta interacción sí aparece, pero no cuando, como aquí, se utiliza solo la coordenada de disociación UDP-Gal. Al considerar el mutuo acercamiento entre Glu317 y carbono anomérico, se observa que la distancia entre ellos se reduce en 1 Å al igual que C1 y el Asp246 en EXTL2, por lo que es de esperar que la catálisis asistida por nucleófilo esté estabilizando las especies oxocarbenio y de esta manera, compensando el hecho de que el UDP no pueda ser estabilizado por el aceptor durante la disociación. En GpgS también actúa la catálisis asistida por nucleófilo pero su efecto ha de ser menor que el del Glu317 (estas contribuciones se analizarán en detalle a continuación). Por otra parte, en LgtC la disociación es sorprendentemente fácil y dos cosas pueden explicar este resultado. Primero, la interacción  $HO4_{Gal}-O_P$  es muy buena (muy corta, Figura 102) desde el principio de la disociación y segundo, el Gln189 se acerca rápidamente al carbono anomérico y es la distancia más corta que se observa entre cualquiera de los residuos nucleofílicos y el respectivo azúcar, se reduce hasta 2.7 Å. Además, en LgtC hay un segundo hidroxilo de la lactosa aceptor que interacciona con el fosfato del UDP, estabilizando la carga negativa creciente. Así, la contribución electrostática del aceptor lactosa y el residuo nucleofílico Gln189 debería de ser buena.



**Figura 103.** Contribuciones electrostáticas de los residuos de la cara  $\beta$  de cada enzima junto a la contribución del sustrato aceptor durante la disociación del sustrato dador. A EXTL2 le corresponde el color rosado, a  $\alpha$ 3GalT el color amarillo, a GpgS el color celeste y a LgtC el color verde. Los cálculos se realizan con el método QM(M05-2X/TZVP//BP86/TZVP)/CHARM(MM).

Observando la Figura 103, se ve que en el caso de LgtC la mayor contribución la brinda el aceptor (LAT), la cual alcanza los -19 kcal/mol mientras que la contribución de Gln189 es menor que 4 kcal/mol durante toda la disociación. En el caso de  $\alpha$ 3GalT, la mayor contribución proviene del Glu317 (-12 kcal/mol en el último punto), aunque el sustrato aceptor incluso desfavorece ligeramente a las especies OC, lo cual podría deberse a que no existe ninguna interacción inter-sustratos. En GpgS, las contribuciones de la Leu209 y el sustrato aceptor (PGA) son similares, aunque la última es un poco más grande, sobre todo en el último punto de la disociación en el que la contribución de PGA es de -7 kcal/mol, y la de la Leu209 es de -4 kcal/mol. De esta manera, es posible que las curvas de energía potencial para la disociación de UDP-azúcar en  $\alpha$ 3GalT y GpgS sean similares porque las contribuciones de PGA y Leu209 prácticamente igualan a la contribución de Glu317 en la cara  $\beta$  de Galactosa. Por supuesto, hay otros residuos que estabilizan a UDP y que en este cálculo no se consideran, sin embargo esto deja en evidencia que un residuo nucleofílico fuerte como Glu o Asp puede llegar a contribuir sustancialmente durante la disociación del sustrato dador a través de la estabilización de las especies oxocarbenio.

Por último, la curva de disociación en EXTL2 resulta ser la más alta, lo cual se relaciona con el efecto desfavorable de la Arg293 y por supuesto con la ausencia de la interacción  $\text{HO}_{4\text{GlcA}}\text{-O}_{\text{P}}$ . A pesar de que esta última interacción no se ha formado aquí, la contribución electrostática del sustrato aceptor es bastante alta, lo cual ha de relacionarse con las otras

dos interacciones inter-sustratos que se establecen en EXTL2:  $\text{HN2}'_{\text{GalNAc}}-\text{O3}_{\text{GlcA}}$  y  $\text{O1}\beta\text{-HO3}_{\text{GlcA}}$ . Por otra parte, el Asp246 también brinda una estabilización sustancial que correspondería casi a la mitad de la contribución del aceptor. Teniendo en cuenta la gran contribución que pueden brindar el aceptor y el Asp246, -30 y -12 kcal/mol, respectivamente, en el último punto, resulta curioso que la disociación de UDP-GalNAc sea energéticamente la más costosa. Sin embargo, sabemos que en la cara  $\beta$  de GalNAc está la Arg293 a corta distancia del centro anomérico; antes se ha mostrado que la Arg293 desfavorece a las especies oxocarbenio y es por eso que en la Figura 103 también se incluye la contribución conjunta de la pareja Asp246-Arg293. Ésta resulta en una estabilización muy escasa a lo largo de prácticamente todo el proceso, y tal vez esta es la razón por la que la disociación del sustrato dador es la más costosa energéticamente, a pesar del gran efecto favorable del sustrato aceptor.

## 7.4 Conclusiones

En este capítulo se han estudiado en detalle las disociaciones de los sustratos dadores de 4 GTs-ret pertenecientes a diferentes familias: EXTL2 GT64, GpgS GT81,  $\alpha$ 3GalT GT6 y LgtC GT8, todas del clan GT-A. También se han calculado las contribuciones electrostáticas del sustrato aceptor y de los residuos de la cara  $\beta$  del respectivo azúcar. Las cuatro enzimas muestran la importancia de la estrategia de catálisis asistida por sustratos y también indican que en un sistema en el que se establecen pocas o ninguna interacción inter-sustratos en el reactivo (disociación UDP-azúcar energéticamente más costosa), el residuo nucleofílico de la cara  $\beta$  del azúcar puede llevar a cabo la catálisis asistida por nucleófilo de una forma muy eficiente. Este es el caso de  $\alpha$ 3GalT, en la que el sustrato aceptor lactosa presenta un puente de hidrógeno intramolecular que no le permite interactuar con el UDP, requiriendo un residuo nucleofílico fuerte que pueda estabilizar sustancialmente las especies oxocarbenio durante la disociación UDP-Gal. Distinto es el caso de LgtC y GpgS, en las que la interacción  $\text{O}_P\text{-Nuc}$  está presente en el reactivo y la catálisis asistida por sustrato se puede llevar a cabo. Además, en LgtC está la interacción  $\text{O}_P\text{-HO3}_{\text{Gal}}$ , una interacción inter-sustratos adicional que debe hacer aún más eficiente la catálisis asistida por sustratos. Finalmente, EXTL2 representa un caso especial porque tiene una pareja Asp246-Arg293 en la cara  $\beta$  del azúcar. Esto ha permitido identificar que el residuo de la cara  $\beta$  del azúcar puede estabilizar las especies OC y/o interactuar con ambos sustratos manteniéndolos en posiciones adecuadas para la catálisis. Este es el caso de Gln189 en LgtC y Glu317 en  $\alpha$ 3GalT. En EXTL2 es la pareja Asp246-Arg293 la que puede llevar a cabo esta tarea, usando a la Arg293 para mantener las interacciones inter-sustratos y Asp246 para estabilizar las especies oxocarbenio. A pesar de que en EXTL2 el sustrato aceptor puede estabilizar bastante durante la disociación del sustrato dador, el residuo nucleofílico Asp246 se necesita igualmente debido a que la Arg293 desfavorece a las especies oxocarbenio. En base a estos análisis, concluimos que la naturaleza del sustrato aceptor modula la cara  $\beta$  del azúcar y esta podría ser la razón por la cual no se observa una arquitectura estructural conservada en esta región a través de todas las familias de GTs-ret.

## 8 Conclusiones Generales

El mecanismo a través del cual las GTs catalizan la formación del enlace glicosídico con retención de la configuración del carbono anomérico del azúcar que se transfiere, es un área de investigación complicada y a pesar de los avances durante las últimas décadas, sigue siendo existiendo algunas incógnitas. A nivel experimental ha sido difícil obtener evidencia para cualquiera de las dos propuestas de mecanismo, la de doble desplazamiento a través de un intermedio CGE, o el mecanismo de ataque frontal o sustitución nucleofílica tipo  $S_Ni$ . El primero ha ido perdiendo sustento a lo largo de los años, por lo menos como mecanismo general, debido a que sólo en una familia de GTs-ret se observa un residuo nucleofílico fuerte en la cara  $\beta$  del azúcar que pueda actuar como nucleófilo para formar el CGE. Nos referimos a la familia GT6, a la cual pertenece  $\alpha 3GalT$ , aunque esto puede ser considerado una excepción pues la mayoría de las GTs-ret no tienen una arquitectura conservada en la cara  $\beta$  del azúcar y por consiguiente, el residuo nucleofílico fuerte que se esperaría encontrar en esta cara no aparece. Esto es completamente distinto a lo que se encuentra en las GHs-ret, en las que el nucleófilo que forma el CGE antes del ataque del agua es conservado en todas las GHs-ret. En vista de esta información se propuso un mecanismo alternativo, el mecanismo de ataque frontal, en el que el sustrato aceptor ataca por la misma cara por la que UDP sale. La característica de ataque frontal en esta propuesta supone una repulsión estérica grande entre grupo saliente (UDP) y nucleófilo (sustrato aceptor), aunque se debe tener presente que el UDP actuará como base catalítica activando el nucleófilo de la reacción y a la vez, estabilizando el desarrollo de carga negativa que se genera por la rotura heterolítica UDP-azúcar. Así, el grupo saliente y el nucleófilo entrante forman una interacción clave que permite que ambos grupos actúen por la misma cara.

En esta tesis se ha utilizado la aproximación QM(DFT)/MM(CHARMM) para estudiar la transferencia de tres azúcares diferentes en dos enzimas. Una de ellas es GpgS de *Mycobacterium Tuberculosis* (GT81), cuyo sustrato dador y aceptor corresponden a UDP-Glc y PGA, respectivamente. La otra enzima corresponde a EXTL2 (GT64), que es catalíticamente activa para dos azúcares que forman parte de los sustratos dadores: UDP-GalNAc y UDP-GlcNAc. El sustrato aceptor es GlcA $\beta$ 1-3Gal, un análogo del núcleo de unión común GAG-proteína a partir del cual se inicia la biosíntesis de varios GAGs.

En GpgS se caracteriza un mecanismo de ataque frontal con la participación clave del sustrato aceptor PGA a través de su grupo carboxilato ( $COO^-_{PGA}$ ), y también a través de la interacción  $O_P-HO3_{PGA}$ ; además, la Leu209 en la cara  $\beta$  de la Glucosa provee una estabilización discreta a las especies oxocarbenio. En el caso de considerarse las posibles aguas en el sitio activo, se comprueba que éstas son capaces de facilitar el mecanismo de ataque frontal interaccionando con ambos sustratos: el grupo  $COO^-_{PGA}$ , y con la Glucosa a través de O2 y C1-H1 (Modelo 1). Independiente de si hay aguas o no en el sitio activo, el mecanismo que sigue GpgS corresponde a un ataque frontal, y aunque a nivel experimental es difícil estar seguro de que las densidades electrónicas corresponden a las aguas por la movilidad que están tienen, es muy probable que entren en el sitio activo.

EXTL2 cataliza un mecanismo de ataque frontal con la participación clave de una pareja de residuos en la cara  $\beta$  GalNAc o GlcNAc, Asp246-Arg293. Esta pareja de residuos convierte a EXTL2 en una enzima muy peculiar pues un residuo con carga neta positiva en la cara  $\beta$  no parece adecuado por la carga positiva que se desarrollará en el azúcar que se transfiere. Los diferentes análisis que hemos llevado a cabo nos han permitido identificar el rol que esta pareja de residuos lleva a cabo en la enzima durante la catálisis del mecanismo de ataque frontal. A pesar de que la Arg293 desfavorece a las especies oxocarbenio, es fundamental para mantener el sustrato aceptor en una posición adecuada para la reacción, además de que facilita la formación de la interacción  $O_P-HO4_{GlcA}$ . El Asp246 también es clave en el mecanismo porque es capaz de estabilizar sustancialmente las especies oxocarbenio y a la vez, compensar el efecto desfavorable de la Arg293 en la cara  $\beta$ .

En el capítulo 7 de esta tesis, se han investigado y comparado las disociaciones de los sustratos dadores de LgtC (UDP-Gal),  $\alpha$ 3GalT (UDP-Gal), GpgS (UDP-Glc) y EXTL2, para la que se ha utilizado el sustrato dador UDP-GalNAc. Además, se calcularon las contribuciones electrostáticas del sustrato aceptor y del o los residuos de la cara  $\beta$  de los azúcares. Los resultados de estos estudios nos permiten generalizar las implicaciones mecanísticas de los residuos de la cara  $\beta$  del azúcar transferible y aportar nueva información respecto a esta zona estructuralmente no conservada en las GTs-ret. En este sentido, las principales conclusiones de nuestros resultados computacionales son:

1. El mecanismo de ataque frontal parece ser el mecanismo preferido o general de las GTs-ret (por lo menos para las estudiadas hasta la fecha). A pesar de que en algunas enzimas, como las de la familia GT6 ( $\alpha$ 3GalT) o GT64 (EXTL2), se encuentra un residuo nucleofílico fuerte en la cara  $\beta$  del azúcar, el rol de éste no sería participar en la formación del intermedio CGE, sino contribuir a la estabilización de las especies oxocarbenio que se generan antes del ataque nucleofílico. Como límite de máxima estabilización, es posible que haya configuraciones del sistema enzima:sustratos en que este residuo nucleofílico fuerte y la especie oxocarbenio puedan llegar a colapsar y formar un CGE. Pero no parece ser una estrategia necesaria, y menos teniendo en cuenta la reorganización nuclear y electrónica posterior necesaria para llegar al producto final.
2. La habilidad nucleofílica del residuo de la cara  $\beta$  del azúcar que se transfiere, en el caso de que esté presente, es disminuida por las interacciones que forma con el sustrato aceptor, y en EXTL2, por su fuerte interacción con la Arg293 que interacciona con el sustrato aceptor.
3. El o los residuos que se encuentran en la cara  $\beta$  del azúcar parecen estar principalmente determinados por el grado de dificultad con que se romperá el enlace UDP-azúcar. Esto a su vez está fuertemente relacionado con la naturaleza química del sustrato aceptor. Teniendo en cuenta que hay un grupo de residuos que estabilizan al UDP en todas las GTs-ret, si el aceptor también puede estabilizar la carga negativa que se desarrolla sobre el UDP, el residuo nucleofílico que se encuentre en la cara  $\beta$  de la GT-ret no será tan fuerte como un Asp o un Glu pero incluso así, podrá brindar una estabilización discreta a las especies oxocarbenio. De lo contrario, el residuo tendrá una carga negativa y podrá estabilizar sustancialmente el desarrollo de carga positiva en el azúcar durante el mecanismo (catálisis asistida por nucleófilo).

4. Además, la cadena lateral de estos residuos de la cara  $\beta$  interacciona con el sustrato aceptor. Eso es así, por lo menos, para todos los sistemas estudiados hasta ahora en los que el aceptor es otro azúcar, pero podría darse también para otros tipos de sustratos.
5. Por lo dicho en las conclusiones 3 y 4, los residuos Gln189 en LgtC y Glu317 en  $\alpha$ 3GalT pueden ejercer su doble función (estabilización de la carga positiva naciente en el carbono anomérico y anclaje del aceptor). Por el contrario, la presencia del grupo carboxilato en el GlcA de EXTL2 obliga a que el anclaje del aceptor lo haga un residuo con carga positiva, la Arg293, la cual necesitará el Asp246 para contrarrestar su efecto desestabilizante sobre el carbono anomérico.
6. La estabilización de la carga negativa que adquiere el grupo saliente (UDP) es fundamental en la catálisis. Esta estabilización es realizada por diferentes residuos enzimáticos, en algunos casos moléculas de agua, y también por el sustrato aceptor, que dependiendo de su estructura establecerá sólo la interacción  $O_p$ -Nuc, o más interacciones inter-sustratos.
7. La interacción inter-sustratos más importante es el enlace de hidrógeno entre UDP y el hidroxilo nucleofílico ( $O_p$ -Nuc). A pesar de que la abstracción del protón no ocurre hasta que la especie UDP---OC está altamente disociada, esta interacción aparece tempranamente en el camino de reacción para estabilizar la salida del grupo saliente y favorecer la cercanía entre los sustratos, reduciendo las repulsiones estéricas inherentes a este mecanismo (catálisis asistida por sustratos).
8. Además de reducir las repulsiones estéricas, la necesidad de las interacciones inter-sustratos en la catálisis podría estar relacionada con evitar la hidrólisis del sustrato dador activado UDP-azúcar.

De esta manera, concluimos que el mecanismo que catalizan las GTs-ret está influenciado por diferentes factores que incluyen la naturaleza del sustrato aceptor y la especificidad del enlace glicosídico que se formará. El tipo de sustrato aceptor influenciará los residuos que se encuentren en la cara  $\beta$  del azúcar que se transferirá, lo cual se debe a que si las interacciones inter-sustratos no son suficientes para estabilizar al UDP y compensar las repulsiones estéricas entre los sustratos, será necesaria una estabilización adicional que favorecerá las especies oxocarbenio. La estabilización de estas especies será llevada a cabo por un residuo nucleofílico en la cara  $\beta$  del azúcar, que puede ser neutro o con carga neta negativa. Cabe mencionar que este residuo nucleofílico, en caso de estar presente, no implica necesariamente la formación del intermedio covalente CGE. La propuesta inicial de que el mecanismo que retiene la configuración del carbono anomérico en las GTs-ret podría ser análogo al mecanismo de retención que catalizan las GHs-ret, ha llevado a asumir que la presencia de un residuo nucleofílico en la cara  $\beta$  del azúcar supone la formación del CGE. Sin embargo, nuestros estudios indican que este residuo es esencial cuando la catálisis asistida por sustratos falla (Glu317 en  $\alpha$ 3GalT), o existe un factor desfavorable pero necesario como la Arg293 (el Asp246 en EXTL2).

# REFERENCIAS

---

- (1) Walt, D. A.-K., K.; Bertozzi, C.; Bendiak, B.; Boons, G.; Darvill, A.; Hart, G.; Kiessling, L.; Lowe, J.; Moon, R.; Paulson, J.; Sasisekharan, R.; Varki, A.; Wong, C. *Transforming Glycoscience: A Roadmap for Future*; Committee on Assessing the Importance and Impact of Glycomics and Glycosciences ed.; The National Academies Press: Washington DC 20001, 2012.
- (2) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochimica Et Biophysica Acta-General Subjects* **1999**, *1473*, 4-8.
- (3) Coutinho, P. M.; Deleury, E.; Davies, G. J.; Henrissat, B. *Journal of Molecular Biology* **2003**, *328*, 307-317.
- (4) Cortes, A.; Munoz-Antoli, C.; Sotillo, J.; Fried, B.; Esteban, J. G.; Toledo, R. *Parasite Immunology* **2015**, *37*, 32-42.
- (5) Varki, A. *Cell* **2006**, *126*, 841-845.
- (6) Bishop, J. R.; Gagneux, P. *Glycobiology* **2007**, *17*, 23R-34R.
- (7) Liang, D. M.; Liu, J. H.; Wu, H.; Wang, B. B.; Zhu, H. J.; Qiao, J. J. *Chemical Society Reviews* **2015**, *44*, 8350-8374.
- (8) In *Carbohydrate Active Enzymes database*.
- (9) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Research* **2009**, *37*, D233-D238.
- (10) Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Research* **2014**, *42*, D490-D495.
- (11) Kiessling, L. L.; Splain, R. A. In *Annual Review of Biochemistry, Vol 79*; Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., Thorner, J. W., Eds., 2010; Vol. 79; pp 619-653.
- (12) Henrissat, B. *Biochemical Journal* **1991**, *280*, 309-316.
- (13) Chothia, C.; Lesk, A. M. *The EMBO Journal* **1986**, *5*, 823-826.
- (14) Gaboriaud, C.; Bissery, V.; Benchetrit, T.; Mornon, J. P. *Febs Letters* **1987**, *224*, 149-155.
- (15) Lemeslevarloot, L.; Henrissat, B.; Gaboriaud, C.; Bissery, V.; Morgat, A.; Mornon, J. P. *Biochimie* **1990**, *72*, 555-574.
- (16) Kötztler, M. P.; Hancock, S. M.; Withers, S. G. In: *eLS. John Wiley & Sons, Ltd: Chichester*. **2014**, July, 1-14.
- (17) Henrissat, B.; Bairoch, A. *Biochemical Journal* **1993**, *293*, 781-788.
- (18) Henrissat, B.; Davies, G. *Current Opinion in Structural Biology* **1997**, *7*, 637-644.
- (19) Johnson, L. N.; Phillips, D. C. *Nature* **1965**, *206*, 761-&.
- (20) KOSHLAND, D. E. J. *Biological Reviews* **1953**, *28*, 416-436.
- (21) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853-859.
- (22) Phillips, D. C. *Proceedings of the National Academy of Sciences of the United States of America* **1967**, *57*, 484-495.
- (23) Kirby, A. J. *Nature Structural Biology* **2001**, *8*, 737-739.
- (24) Sinnott, M. L. *Chemical Reviews* **1990**, *90*, 1171-1202.
- (25) Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. *Nature* **2001**, *412*, 835-838.
- (26) Bowman, A. L.; Grant, I. M.; Mulholland, A. J. *Chemical Communications* **2008**, 4425-4427.
- (27) Bottoni, A.; Miscione, G. P.; De Vivo, M. *Proteins-Structure Function and Bioinformatics* **2005**, *59*, 118-130.
- (28) Zechel, D. L.; Withers, S. G. *Accounts of Chemical Research* **2000**, *33*, 11-18.

- (29) Ardevol, A.; Rovira, C. *Journal of the American Chemical Society* **2015**, *137*, 7528-7547.
- (30) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280-15287.
- (31) Tews, I.; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E. *Nature Structural Biology* **1996**, *3*, 638-648.
- (32) Espinosa, J. F.; Montero, E.; Vian, A.; Garcia, J. L.; Dietrich, H.; Schmidt, R. R.; Martin-Lomas, M.; Imberty, A.; Canada, F. J.; Jimenez-Barbero, J. *Journal of the American Chemical Society* **1998**, *120*, 1309-1318.
- (33) Kirby, A. J. *Accounts of Chemical Research* **1984**, *17*, 305-311.
- (34) Biarnes, X.; Nieto, J.; Planas, A.; Rovira, C. *Journal of Biological Chemistry* **2006**, *281*, 1432-1441.
- (35) Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B. *Biochemical Journal* **1997**, *326*, 929-939.
- (36) Breton, C.; Snajdrova, L.; Jeanneau, C.; Koca, J.; Imberty, A. *Glycobiology* **2006**, *16*, 29R-37R.
- (37) Hu, Y. N.; Walker, S. *Chemistry & Biology* **2002**, *9*, 1287-1296.
- (38) Qasba, P. K.; Ramakrishnan, B.; Boeggeman, E. *Trends in Biochemical Sciences* **2005**, *30*, 53-62.
- (39) Lovering, A. L.; de Castro, L. H.; Lim, D.; Strynadka, N. C. J. *Science* **2007**, *315*, 1402-1405.
- (40) Igura, M.; Maita, N.; Kamishikiryō, J.; Yamada, M.; Obita, T.; Maenaka, K.; Kohda, D. *Embo Journal* **2008**, *27*, 234-243.
- (41) Liu, J.; Mushegian, A. *Protein Science* **2003**, *12*, 1418-1431.
- (42) Vetting, M. W.; Frantom, P. A.; Blanchard, J. S. *Journal of Biological Chemistry* **2008**, *283*, 15834-15844.
- (43) Chiu, C. P. C.; Watts, A. G.; Lairson, L. L.; Gilbert, M.; Lim, D.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. J. *Nature Structural & Molecular Biology* **2004**, *11*, 163-170.
- (44) Pak, J. E.; Arnoux, P.; Zhou, S.; Sivarajah, P.; Satkunarajah, M.; Xing, X.; Rini, J. M. *Journal of Biological Chemistry* **2006**, *281*, 26693-26701.
- (45) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. In *Annual Review of Biochemistry*, 2008; Vol. 77; pp 521-555.
- (46) Davies, G. J. *Nature Structural Biology* **2001**, *8*, 98-100.
- (47) Geremia, R. A.; Petroni, E. A.; Ielpi, L.; Henrissat, B. *Biochemical Journal* **1996**, *318*, 133-138.
- (48) Saxena, I. M.; Brown, R. M.; Fevre, M.; Geremia, R. A.; Henrissat, B. *Journal of Bacteriology* **1995**, *177*, 1419-1424.
- (49) In *CAZypedia*.
- (50) Henrissat, B.; Bairoch, A. *Biochemical Journal* **1996**, *316*, 695-696.
- (51) Franco, O. L.; Rigden, D. J. *Glycobiology* **2003**, *13*, 707-712.
- (52) Bennett-Lovsey, R. M.; Herbert, A. D.; Sternberg, M. J. E.; Kelley, L. A. *Proteins-Structure Function and Bioinformatics* **2008**, *70*, 611-625.
- (53) Breton, C.; Fournel-Gigleux, S.; Palcic, M. M. *Current Opinion in Structural Biology* **2012**, *22*, 540-549.
- (54) Unligil, U. M.; Rini, J. M. *Current Opinion in Structural Biology* **2000**, *10*, 510-517.
- (55) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *Journal of Molecular Biology* **1995**, *247*, 536-540.
- (56) Chiu, C. P. C.; Lairson, L. L.; Gilbert, M.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. J. *Biochemistry* **2007**, *46*, 7196-7204.
- (57) Vrieland, A.; Ruger, W.; Driessen, H. P. C.; Freemont, P. S. *Embo Journal* **1994**, *13*, 3413-3422.



- (58) Ihara, H.; Ikeda, Y.; Toma, S.; Wang, X.; Suzuki, T.; Gu, J.; Miyoshi, E.; Tsukihara, T.; Honke, K.; Matsumoto, A.; Nakagawa, A.; Taniguchi, N. *Glycobiology* **2007**, *17*, 455-466.
- (59) Maeda, Y.; Watanabe, R.; Harris, C. L.; Hong, J. J.; Ohishi, K.; Kinoshita, K.; Kinoshita, T. *Embo Journal* **2001**, *20*, 250-261.
- (60) Strahlbolsinger, S.; Immervoll, T.; Deutzmann, R.; Tanner, W. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 8164-8168.
- (61) Takahashi, M.; Inoue, N.; Ohishi, K.; Maeda, Y.; Nakamura, N.; Endo, Y.; Fujita, T.; Takeda, J.; Kinoshita, T. *Embo Journal* **1996**, *15*, 4254-4261.
- (62) Gastinel, L. N.; Cambillau, C.; Bourne, Y. *Embo Journal* **1999**, *18*, 3546-3557.
- (63) Ramakrishnan, B.; Qasba, P. K. *Journal of Molecular Biology* **2001**, *310*, 205-218.
- (64) Ramasamy, V.; Ramakrishnan, B.; Boeggeman, E.; Qasba, P. K. *Journal of Molecular Biology* **2003**, *331*, 1065-1076.
- (65) Boix, E.; Swaminathan, G. J.; Zhang, Y. N.; Natesh, R.; Brew, K.; Acharya, K. R. *Journal of Biological Chemistry* **2001**, *276*, 48608-48614.
- (66) Boix, E.; Zhang, Y. N.; Swaminathan, G. J.; Brew, K.; Acharya, K. R. *Journal of Biological Chemistry* **2002**, *277*, 28310-28318.
- (67) Persson, K.; Ly, H. D.; Dieckelmann, M.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. J. *Nature Structural Biology* **2001**, *8*, 166-175.
- (68) Unligil, U. M.; Zhou, S. H.; Yuwaraj, S.; Sarkar, M.; Schachter, H.; Rini, J. M. *Embo Journal* **2000**, *19*, 5269-5280.
- (69) Milac, A. L.; Buchete, N. V.; Fritz, T. A.; Hummer, G.; Tabak, L. A. *Journal of Molecular Biology* **2007**, *373*, 439-451.
- (70) Pedersen, L. C.; Dong, J.; Taniguchi, F.; Kitagawa, H.; Krahn, J. M.; Pedersen, L. G.; Sugahara, K.; Negishi, M. *Journal of Biological Chemistry* **2003**, *278*, 14420-14428.
- (71) Urresti, S.; Albesa-Jove, D.; Schaeffer, F.; Pham, H. T.; Kaur, D.; Gest, P.; van der Woerd, M. J.; Carreras-Gonzalez, A.; Lopez-Fernandez, S.; Alzari, P. M.; Brennan, P. J.; Jackson, M.; Guerin, M. E. *Journal of Biological Chemistry* **2012**, *287*, 24649-24661.
- (72) Charnock, S. J.; Davies, G. J. *Biochemistry* **1999**, *38*, 6380-6385.
- (73) Tarbouriech, N.; Charnock, S. J.; Davies, G. J. *Journal of Molecular Biology* **2001**, *314*, 655-661.
- (74) Garinot-Schneider, C.; Lellouch, A. C.; Geremia, R. A. *Journal of Biological Chemistry* **2000**, *275*, 31407-31413.
- (75) Pedersen, L. C.; Darden, T. A.; Negishi, M. *Journal of Biological Chemistry* **2002**, *277*, 21869-21873.
- (76) Ramakrishnan, B.; Balaji, P. V.; Qasba, P. K. *Journal of Molecular Biology* **2002**, *318*, 491-502.
- (77) Ramakrishnan, B.; Ramasamy, V.; Qasba, P. K. *Journal of Molecular Biology* **2006**, *357*, 1619-1633.
- (78) Ramakrishnan, B.; Shah, P. S.; Qasba, P. K. *Journal of Biological Chemistry* **2001**, *276*, 37665-37671.
- (79) Ramasamy, V.; Ramakrishnan, B.; Boeggeman, E.; Ratner, D. M.; Seeberger, P. H.; Qasba, P. K. *Journal of Molecular Biology* **2005**, *353*, 53-67.
- (80) Gordon, R. D.; Sivarajah, P.; Satkunarajah, M.; Ma, D.; Tarling, C. A.; Vizitiu, D.; Withers, S. G.; Rini, J. M. *Journal of Molecular Biology* **2006**, *360*, 67-79.
- (81) Jinek, M.; Chen, Y.-W.; Clausen, H.; Cohen, S. M.; Conti, E. *Nature Structural & Molecular Biology* **2006**, *13*, 945-946.
- (82) Ohtsubo, K.; Imajo, S.; Ishiguro, M.; Nakatani, T.; Oka, S.; Kawasaki, T. *Journal of Biochemistry* **2000**, *128*, 283-291.
- (83) Kakuda, S.; Shiba, T.; Ishiguro, M.; Tagawa, H.; Oka, S.; Kajihara, Y.; Kawasaki, T.; Wakatsuki, S.; Kato, R. *Journal of Biological Chemistry* **2004**, *279*, 22693-22703.

- (84) Ramakrishnan, B.; Qasba, P. K. *Journal of Biological Chemistry* **2010**, *285*, 15619-15626.
- (85) Ha, S.; Walker, D.; Shi, Y. G.; Walker, S. *Protein Science* **2000**, *9*, 1045-1052.
- (86) Hu, Y. N.; Chen, L.; Ha, S.; Gross, B.; Falcone, B.; Walker, D.; Mokhtarzadeh, M.; Walker, S. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 845-849.
- (87) Bolam, D. N.; Roberts, S.; Proctor, M. R.; Turkenburg, J. P.; Dodson, E. J.; Martinez-Fleites, C.; Yang, M.; Davis, B. G.; Davies, G. J.; Gilbert, H. J. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 9911-9911.
- (88) Lazarus, M. B.; Nam, Y.; Jiang, J.; Sliz, P.; Walker, S. *Nature* **2011**, *469*, 564-U168.
- (89) Chang, A.; Singh, S.; Helmich, K. E.; Goff, R. D.; Bingman, C. A.; Thorson, J. S.; Phillips, G. N., Jr. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 17649-17654.
- (90) Shao, H.; He, X. Z.; Achnine, L.; Blount, J. W.; Dixon, R. A.; Wang, X. Q. *Plant Cell* **2005**, *17*, 3141-3154.
- (91) Offen, W.; Martinez-Fleites, C.; Yang, M.; Kiat-Lim, E.; Davis, B. G.; Tarling, C. A.; Ford, C. M.; Bowles, D. J.; Davies, G. J. *Embo Journal* **2006**, *25*, 1396-1405.
- (92) Bolam, D. N.; Roberts, S.; Proctor, M. R.; Turkenburg, J. P.; Dodson, E. J.; Martinez-Fleites, C.; Yang, M.; Davis, B. G.; Davies, G. J.; Gilbert, H. J. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 5336-5341.
- (93) Miley, M. J.; Zielinska, A. K.; Keenan, J. E.; Bratton, S. M.; Radomska-Pandya, A.; Redinbo, M. R. *Journal of Molecular Biology* **2007**, *369*, 498-511.
- (94) Brazier-Hicks, M.; Offen, W. A.; Gershater, M. C.; Revett, T. J.; Lim, E.-K.; Bowles, D. J.; Davies, G. J.; Edwards, R. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 20238-20243.
- (95) Mulichak, A. M.; Losey, H. C.; Lu, W.; Wawrzak, Z.; Walsh, C. T.; Garavito, R. M. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 9238-9243.
- (96) Mulichak, A. M.; Lu, W.; Losey, H. C.; Walsh, C. T.; Garavito, R. M. *Biochemistry* **2004**, *43*, 5170-5180.
- (97) Mulichak, A. M.; Losey, H. C.; Walsh, C. T.; Garavito, R. M. *Structure* **2001**, *9*, 547-557.
- (98) Morera, S.; Lariviere, L.; Kurzeck, J.; Aschke-Sonnenborn, U.; Freemont, P. S.; Janin, J.; Ruger, W. *Journal of Molecular Biology* **2001**, *311*, 569-577.
- (99) Lariviere, L.; Gueguen-Chaignon, V.; Morera, S. *Journal of Molecular Biology* **2003**, *330*, 1077-1086.
- (100) Grizot, S.; Salem, M.; Vongsouthi, V.; Durand, L.; Moreau, F.; Dohi, H.; Vincent, S.; Escaich, S.; Ducruix, A. *Journal of Molecular Biology* **2006**, *363*, 383-394.
- (101) Sun, H.-Y.; Lin, S.-W.; Ko, T.-P.; Pan, J.-F.; Liu, C.-L.; Lin, C.-N.; Wang, A. H. J.; Lin, C.-H. *Journal of Biological Chemistry* **2007**, *282*, 9973-9982.
- (102) Murray, B. W.; Wittmann, V.; Burkart, M. D.; Hung, S. C.; Wong, C. H. *Biochemistry* **1997**, *36*, 823-831.
- (103) Ihara, H.; Ikeda, Y.; Taniguchi, N. *Glycobiology* **2006**, *16*, 333-342.
- (104) Lira-Navarrete, E.; Valero-Gonzalez, J.; Villanueva, R.; Martinez-Julvez, M.; Tejero, T.; Merino, P.; Panjikar, S.; Hurtado-Guerrero, R. *Plos One* **2011**, *6*.
- (105) Gibson, R. P.; Turkenburg, J. P.; Charnock, S. J.; Lloyd, R.; Davies, G. J. *Chemistry & Biology* **2002**, *9*, 1337-1346.
- (106) Gibson, R. P.; Tarling, C. A.; Roberts, S.; Withers, S. G.; Davies, G. J. *Journal of Biological Chemistry* **2004**, *279*, 1950-1955.
- (107) Lairson, L. L.; Chiu, C. P. C.; Ly, H. D.; He, S. M.; Wakarchuk, W. W.; Strynadka, N. C. J.; Withers, S. G. *Journal of Biological Chemistry* **2004**, *279*, 28339-28344.
- (108) Gibbons, B. J.; Roach, P. J.; Hurley, T. D. *Journal of Molecular Biology* **2002**, *319*, 463-477.

- (109) Kubota, T.; Shiba, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H. *Journal of Molecular Biology* **2006**, *359*, 708-727.
- (110) Fritz, T. A.; Raman, J.; Tabak, L. A. *Journal of Biological Chemistry* **2006**, *281*, 8613-8619.
- (111) Gomez, H.; Rojas, R.; Patel, D.; Tabak, L. A.; Lluch, J. M.; Masgrau, L. *Organic & Biomolecular Chemistry* **2014**, *12*, 2645-2655.
- (112) Lira-Navarrete, E.; Iglesias-Fernandez, J.; Zandberg, W. F.; Companon, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocado, D. J.; Rovira, C.; Hurtado-Guerrero, R. *Angewandte Chemie-International Edition* **2014**, *53*, 8206-8210.
- (113) Reinert, D. J.; Jank, T.; Aktories, K.; Schulz, G. E. *Journal of Molecular Biology* **2005**, *351*, 973-981.
- (114) Flint, J.; Taylor, E.; Yang, M.; Bolam, D. N.; Tailford, L. E.; Martinez-Fleites, C.; Dodson, E. J.; Davis, B. G.; Gilbert, H. J.; Davies, G. J. *Nature Structural & Molecular Biology* **2005**, *12*, 608-614.
- (115) Lobsanov, Y. D.; Romero, P. A.; Sleno, B.; Yu, B. M.; Yip, P.; Herscovics, A.; Howell, P. L. *Journal of Biological Chemistry* **2004**, *279*, 17921-17931.
- (116) Bobovska, A.; Tvaroska, I.; Kona, J. *Organic & Biomolecular Chemistry* **2014**, *12*, 4201-4210.
- (117) Sinnott, M. L.; Jencks, W. P. *Journal of the American Chemical Society* **1980**, *102*, 2026-2032.
- (118) Klein, H. W.; Im, M. J.; Palm, D. *European Journal of Biochemistry* **1986**, *157*, 107-114.
- (119) Hosfield, D. J.; Zhang, Y. M.; Dougan, D. R.; Broun, A.; Tari, L. W.; Swanson, R. V.; Finn, J. *Journal of Biological Chemistry* **2004**, *279*, 8526-8529.
- (120) Gabelli, S. B.; McLellan, J. S.; Montalvetti, A.; Oldfield, E.; Docampo, R.; Amzel, L. M. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 80-88.
- (121) Tvaroska, I. *Carbohydrate Research* **2004**, *339*, 1007-1014.
- (122) Andre, I.; Tvaroska, I.; Carver, J. P. *Carbohydrate Research* **2003**, *338*, 865-877.
- (123) Ardevol, A.; Rovira, C. *Angewandte Chemie-International Edition* **2011**, *50*, 10897-10901.
- (124) Gomez, H.; Lluch, J. M.; Masgrau, L. *Journal of the American Chemical Society* **2013**, *135*, 7053-7063.
- (125) Gomez, H.; Polyak, I.; Thiel, W.; Lluch, J. M.; Masgrau, L. *Journal of the American Chemical Society* **2012**, *134*, 4743-4752.
- (126) Rojas-Cervellera, V.; Ardevol, A.; Boero, M.; Planas, A.; Rovira, C. *Chemistry-a European Journal* **2013**, *19*, 14018-14023.
- (127) Errey, J. C.; Lee, S. S.; Gibson, R. P.; Fleites, C. M.; Barry, C. S.; Jung, P. M. J.; O'Sullivan, A. C.; Davis, B. G.; Davies, G. J. *Angewandte Chemie-International Edition* **2010**, *49*, 1234-1237.
- (128) Lee, S. S.; Hong, S. Y.; Errey, J. C.; Izumi, A.; Davies, G. J.; Davis, B. G. *Nature Chemical Biology* **2011**, *7*, 631-638.
- (129) Frantom, P. A.; Coward, J. K.; Blanchard, J. S. *Journal of the American Chemical Society* **2010**, *132*, 6626-+.
- (130) Lairson, L. L.; Withers, S. G. *Chemical Communications* **2004**, 2243-2248.
- (131) Gomez, H.; Lluch, J. M.; Masgrau, L. *Carbohydrate Research* **2012**, *356*, 204-208.
- (132) Bobovska, A.; Tvaroska, I.; Kona, J. *Glycobiology* **2015**, *25*, 3-7.
- (133) Zhang, Y. N.; Swaminathan, G. J.; Deshpande, A.; Boix, E.; Natesh, R.; Xie, Z. H.; Acharya, K. R.; Brew, K. *Biochemistry* **2003**, *42*, 13512-13521.
- (134) Trnka, T.; Kozmon, S.; Tvaroska, I.; Koca, J. *Plos Computational Biology* **2015**, *11*.
- (135) Cramer, C. J. *Theories and Models*; Second Edition ed.; John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2004.

- (136) Leach, A. *Molecular Modelling: Principles and Applications* Second Edition ed.; Pearson, 2001.
- (137) Jensen, F. *Introduction to Computational Chemistry*; Second Edition ed.; John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2007.
- (138) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *Journal of Computational Chemistry* **1986**, *7*, 230-252.
- (139) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179-5197.
- (140) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187-217.
- (141) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* **1998**, *102*, 3586-3616.
- (142) Jorgensen, W. L.; Tiradorives, J. *Journal of the American Chemical Society* **1988**, *110*, 1657-1666.
- (143) Damm, W.; Frontera, A.; TiradoRives, J.; Jorgensen, W. L. *Journal of Computational Chemistry* **1997**, *18*, 1955-1970.
- (144) Hermans, J.; Berendsen, H. J. C.; Vangunsteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, *23*, 1513-1518.
- (145) Ott, K. H.; Meyer, B. *Journal of Computational Chemistry* **1996**, *17*, 1068-1084.
- (146) Mackerell, A. D. *Journal of Computational Chemistry* **2004**, *25*, 1584-1604.
- (147) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089-10092.
- (148) Rokhlin, V. *Journal of Computational Physics* **1985**, *60*, 187-207.
- (149) Born, M.; Oppenheimer, R. *Annalen Der Physik* **1927**, *84*, 0457-0484.
- (150) Nebot-Gil, I.; Tel, L. M.; Martín, F.; Sordo, T.; Ugalde, J. M.; Vázquez, S. A.; Caballol, R.; Yáñez Montero, M.; Solà, M.; Ugalde, J. M.; Novoa, J. J.; Largo, A.; Barrientos, C.; Besalú, E.; Bader, R. F. W. *THEORETICAL AND COMPUTATIONAL CHEMISTRY: FOUNDATIONS METHODS AND TECHNIQUES*; Publicacions de la Universitat Jaume I, Castelló de la Plana 2007.
- (151) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, 1133-&.
- (152) Slater, J. C. *Physical Review* **1951**, *81*, 385-390.
- (153) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Journal of Physical Chemistry A* **2007**, *111*, 10439-10452.
- (154) Becke, A. D. *Physical Review A* **1988**, *38*, 3098-3100.
- (155) Levy, M.; Perdew, J. P. *Journal of Chemical Physics* **1986**, *84*, 4519-4523.
- (156) Handy, N. C.; Cohen, A. J. *Molecular Physics* **2001**, *99*, 403-412.
- (157) Xu, X.; Goddard, W. A. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 2673-2677.
- (158) Kurth, S.; Perdew, J. P.; Blaha, P. *International Journal of Quantum Chemistry* **1999**, *75*, 889-909.
- (159) Wang, Y.; Perdew, J. P. *Physical Review B* **1991**, *43*, 8911-8916.
- (160) Becke, A. D. *Journal of Chemical Physics* **1986**, *84*, 4524-4529.
- (161) Perdew, J. P.; Levy, M. *Physical Review B* **1985**, *31*, 6264-6272.
- (162) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Physical Review Letters* **1996**, *77*, 3865-3868.
- (163) Adamo, C.; Barone, V. *Journal of Chemical Physics* **2002**, *116*, 5933-5940.
- (164) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Physical Review Letters* **2003**, *91*.

- (165) Van Voorhis, T.; Scuseria, G. E. *Journal of Chemical Physics* **1998**, *109*, 400-410.
- (166) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *2*, 364-382.
- (167) Perdew, J. P.; Tao, J. M.; Staroverov, V. N.; Scuseria, G. E. *Journal of Chemical Physics* **2004**, *120*, 6898-6911.
- (168) Dewar, M. J. S.; Thiel, W. *Journal of the American Chemical Society* **1977**, *99*, 4899-4907.
- (169) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *Journal of the American Chemical Society* **1985**, *107*, 3902-3909.
- (170) Stewart, J. J. P. *Journal of Computational Chemistry* **1989**, *10*, 221-264.
- (171) Elstner, M. *Theoretical Chemistry Accounts* **2006**, *116*, 316-325.
- (172) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Physical Review B* **1998**, *58*, 7260-7268.
- (173) Cai, Z.-L.; Lopez, P.; Reimers, J. R.; Cui, Q.; Elstner, M. *Journal of Physical Chemistry A* **2007**, *111*, 5743-5750.
- (174) Slater, J. C.; Koster, G. F. *Physical Review* **1954**, *94*, 1498-1524.
- (175) Friesner, R. A.; Guallar, V. In *Annual Review of Physical Chemistry*, 2005; Vol. 56; pp 389-427.
- (176) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. *Chemical Society Reviews* **2012**, *41*, 3025-3038.
- (177) Senn, H. M.; Thiel, W. *Angewandte Chemie-International Edition* **2009**, *48*, 1198-1229.
- (178) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. *Journal of Chemical Theory and Computation* **2006**, *2*, 815-826.
- (179) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. *Journal of Physical Chemistry B* **1999**, *103*, 6133-6141.
- (180) Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discussions* **1997**, *106*, 79-92.
- (181) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schafer, A.; Lennartz, C. *Journal of Molecular Structure-Theochem* **2003**, *632*, 1-28.
- (182) Antes, I.; Thiel, W. *Journal of Physical Chemistry A* **1999**, *103*, 9290-9295.
- (183) Xiao, C.; Zhang, Y. *Journal of Chemical Physics* **2007**, *127*.
- (184) Zhang, Y. *Theoretical Chemistry Accounts* **2006**, *116*, 43-50.
- (185) Thery, V.; Rinaldi, D.; Rivail, J. L.; Maignet, B.; Ferenczy, G. G. *Journal of Computational Chemistry* **1994**, *15*, 269-282.
- (186) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *Journal of Computational Chemistry* **2000**, *21*, 1442-1457.
- (187) Garcia-Viloca, M.; Gao, J. L. *Theoretical Chemistry Accounts* **2004**, *111*, 280-286.
- (188) Freindorf, M.; Shao, Y. H.; Furlani, T. R.; Kong, J. *Journal of Computational Chemistry* **2005**, *26*, 1270-1278.
- (189) Riccardi, D.; Li, G. H.; Cui, Q. *Journal of Physical Chemistry B* **2004**, *108*, 6467-6478.
- (190) Verlet, L. *Physical Review* **1967**, *159*, 98-+.
- (191) Hockney, R. *Methods in Computational Physics* **1970**, *9*, 135-211.
- (192) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327-341.
- (193) Krautler, V.; Van Gunsteren, W. F.; Hunenberger, P. H. *Journal of Computational Chemistry* **2001**, *22*, 501-508.
- (194) Andersen, H. C. *Journal of Computational Physics* **1983**, *52*, 24-34.

- (195) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Second ed.; Academic Press, San Diego, USA, 2002.
- (196) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695-1697.
- (197) Nose, S. *Journal of Chemical Physics* **1984**, *81*, 511-519.
- (198) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684-3690.
- (199) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *Journal of Chemical Physics* **1995**, *103*, 4613-4621.
- (200) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *Journal of Chemical Physics* **1994**, *101*, 4177-4189.
- (201) Kirkwood, J. G. *Journal of Chemical Physics* **1935**, *3*, 300-313.
- (202) Torrie, G. M.; Valleau, J. P. *Chemical Physics Letters* **1974**, *28*, 578-581.
- (203) Torrie, G. M.; Valleau, J. P. *Journal of Computational Physics* **1977**, *23*, 187-199.
- (204) Roux, B. *Computer Physics Communications* **1995**, *91*, 275-282.
- (205) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *Journal of Computational Chemistry* **1992**, *13*, 1011-1021.
- (206) Kastner, J.; Thiel, W. *Journal of Chemical Physics* **2005**, *123*.
- (207) Kastner, J.; Thiel, W. *Journal of Chemical Physics* **2006**, *124*.
- (208) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nature Reviews Drug Discovery* **2004**, *3*, 935-949.
- (209) Goodsell, D. S.; Lauble, H.; Stout, C. D.; Olson, A. J. *Proteins-Structure Function and Genetics* **1993**, *17*, 1-10.
- (210) Hart, T. N.; Read, R. J. *Proteins-Structure Function and Genetics* **1992**, *13*, 206-222.
- (211) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *Journal of Computational Chemistry* **1998**, *19*, 1639-1662.
- (212) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. *Journal of Computer-Aided Molecular Design* **1995**, *9*, 113-130.
- (213) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *Journal of Molecular Biology* **1997**, *267*, 727-748.
- (214) Desmet, J.; Demaeyer, M.; Hazes, B.; Lasters, I. *Nature* **1992**, *356*, 539-542.
- (215) Leach, A. R. *Journal of Molecular Biology* **1994**, *235*, 345-356.
- (216) Kramer, B.; Rarey, M.; Lengauer, T. *Proteins-Structure Function and Genetics* **1999**, *37*, 228-241.
- (217) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins-Structure Function and Genetics* **2003**, *52*, 609-623.
- (218) Gest, P.; Kaur, D.; Pham, H. T.; van der Woerd, M.; Hansen, E.; Brennan, P. J.; Jackson, M.; Guerin, M. E. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* **2008**, *64*, 1121-1124.
- (219) Stadthagen, G.; Sambou, T.; Guerin, M.; Barilone, N.; Boudou, F.; Kordulakova, J.; Charles, P.; Alzari, P. M.; Lemassu, A.; Daffe, M.; Puzo, G.; Gicquel, B.; Riviere, M.; Jackson, M. *Journal of Biological Chemistry* **2007**, *282*, 27270-27276.
- (220) Albesa-Jove, D.; Mendoza, F.; Rodrigo-Unzueta, A.; Gomollon-Bel, F.; Cifuentes, J. O.; Urresti, S.; Comino, N.; Gomez, H.; Romero-Garcia, J.; Lluch, J. M.; Sancho-Vaello, E.; Biarnes, X.; Planas, A.; Merino, P.; Masgrau, L.; Guerin, M. E. *Angewandte Chemie-International Edition* **2015**, *54*, 9898-9902.
- (221) Fiser, A.; Do, R. K. G.; Sali, A. *Protein Science* **2000**, *9*, 1753-1773.
- (222) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins-Structure Function and Bioinformatics* **2005**, *61*, 704-721.
- (223) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 525-537.
- (224) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 2284-2295.
- (225) In PDB2PQR.

- (226) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Research* **2007**, *35*, W522-W525.
- (227) MacKerell, A. D.; Feig, M.; Brooks, C. L. *Journal of the American Chemical Society* **2004**, *126*, 698-699.
- (228) Zhu, X.; Lopes, P. E. M.; MacKerell, A. D., Jr. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2012**, *2*, 167-185.
- (229) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *Journal of Chemical Theory and Computation* **2009**, *5*, 2353-2370.
- (230) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W., II; MacKerell, A. D., Jr. *Journal of Chemical Theory and Computation* **2011**, *7*, 3162-3180.
- (231) Mallajosyula, S. S.; Guvench, O.; Hatcher, E.; MacKerell, A. D., Jr. *Journal of Chemical Theory and Computation* **2012**, *8*, 759-776.
- (232) In *CGenFF interface at paramchem.org*.
- (233) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. *Journal of Computational Chemistry* **2010**, *31*, 671-690.
- (234) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D., Jr. *Journal of Computational Chemistry* **2012**, *33*, 2451-2468.
- (235) Vanommeslaeghe, K.; MacKerell, A. D., Jr. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144-3154.
- (236) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. *Journal of Chemical Information and Modeling* **2012**, *52*, 3155-3168.
- (237) Mahoney, M. W.; Jorgensen, W. L. *Journal of Chemical Physics* **2000**, *112*, 8910-8922.
- (238) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *Journal of Computational Physics* **1999**, *151*, 283-312.
- (239) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781-1802.
- (240) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics & Modelling* **1996**, *14*, 33-38.
- (241) Field, M. J.; Bash, P. A.; Karplus, M. *Journal of Computational Chemistry* **1990**, *11*, 700-733.
- (242) Singh, U. C.; Kollman, P. A. *Journal of Computational Chemistry* **1986**, *7*, 718-730.
- (243) Bakowies, D.; Thiel, W. *Journal of Physical Chemistry* **1996**, *100*, 10580-10594.
- (244) Perdew, J. P. *Physical Review B* **1986**, *33*, 8822-8824.
- (245) Perdew, J. P. *Physical Review B* **1986**, *34*, 7406-7406.
- (246) Vosko, S. H.; Wilk, L.; Nusair, M. *Canadian Journal of Physics* **1980**, *58*, 1200-1211.
- (247) Schafer, A.; Horn, H.; Ahlrichs, R. *Journal of Chemical Physics* **1992**, *97*, 2571-2577.
- (248) Schafer, A.; Huber, C.; Ahlrichs, R. *Journal of Chemical Physics* **1994**, *100*, 5829-5835.
- (249) Kona, J.; Tvaroska, I. *Chemical Papers* **2009**, *63*, 598-607.
- (250) In *ChemShell, a Computational Chemistry Shell*.
- (251) Smith, W.; Forester, T. R. *Journal of Molecular Graphics* **1996**, *14*, 136-141.
- (252) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chemical Physics Letters* **1989**, *162*, 165-169.
- (253) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J.

- L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr, J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; J. M. Millam; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; J. Cioslowski; Fox, D. J.; Gaussian 09, revision B.01; Gaussian, Inc: Wallingford CT, 2010.
- (254) Thiel, W.; 7.0 ed.: Muelheim, Germany, 2005.
- (255) Billeter, S. R.; Turner, A. J.; Thiel, W. *Physical Chemistry Chemical Physics* **2000**, *2*, 2177-2186.
- (256) Turner, A. J.; Moliner, V.; Williams, I. H. *Physical Chemistry Chemical Physics* **1999**, *1*, 1323-1331.
- (257) Broyden, C. G. *Mathematics of Computation* **1970**, *24*, 365-&.
- (258) Fletcher, R. *Computer Journal* **1970**, *13*, 317-&.
- (259) Baker, J. *Journal of Computational Chemistry* **1986**, *7*, 385-395.
- (260) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *Journal of Physical Chemistry* **1985**, *89*, 52-57.
- (261) Claeysens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schuetz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angewandte Chemie-International Edition* **2006**, *45*, 6856-6859.
- (262) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. *Journal of Physical Chemistry B* **2010**, *114*, 1156-1162.
- (263) Lonsdale, R.; Houghton, K. T.; Zurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J. *Journal of the American Chemical Society* **2013**, *135*, 8001-8015.
- (264) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *Journal of Chemical Physics* **1985**, *83*, 735-746.
- (265) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chemical Reviews* **1988**, *88*, 899-926.
- (266) Reed, A. E.; Weinhold, F. *Journal of Chemical Physics* **1983**, *78*, 4066-4073.
- (267) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F., NBO Program 3.1.
- (268) Gomez, H.; Mendoza, F.; Lluch, J. M.; Masgrau, L. *Advances in protein chemistry and structural biology* **2015**, *100*, 225-254.
- (269) Kitagawa, H.; Shimakawa, H.; Sugahara, K. *Journal of Biological Chemistry* **1999**, *274*, 13933-13937.
- (270) Iozzo, R. V. *Annual Review of Biochemistry* **1998**, *67*, 609-652.
- (271) Lindahl, U.; Kusche-Gullberg, M.; Kjellen, L. *Journal of Biological Chemistry* **1998**, *273*, 24979-24982.
- (272) Perrimon, N.; Bernfield, M. *Nature* **2000**, *404*, 725-728.
- (273) Nadanaka, S.; Kitagawa, H. *Journal of Biochemistry* **2008**, *144*, 7-14.
- (274) van Wijk, X. M.; Thijssen, V. L.; Lawrence, R.; van den Broek, S. A.; Dona, M.; Naidu, N.; Oosterhof, A.; van de Westerlo, E. M.; Kusters, L. J.; Khaled, Y.; Jokela, T. A.; Nowak-Sliwinska, P.; Kremer, H.; Stringer, S. E.; Griffioen, A. W.; van Wijk, E.; van Delft, F. L.; van Kuppevelt, T. H. *Acs Chemical Biology* **2013**, *8*, 2331-2338.
- (275) Wise, C. A.; Clines, G. A.; Massa, H.; Trask, B. J.; Lovett, M. *Genome Research* **1997**, *7*, 10-16.
- (276) Lind, T.; Tufaro, F.; McCormick, C.; Lindahl, U.; Lidholt, K. *Journal of Biological Chemistry* **1998**, *273*, 26265-26268.
- (277) Nadanaka, S.; Zhou, S.; Kagiyama, S.; Shoji, N.; Sugahara, K.; Sugihara, K.; Asano, M.; Kitagawa, H. *Journal of Biological Chemistry* **2013**, *288*, 9321-9333.
- (278) Fritz, T. A.; Gabb, M. M.; Wei, G.; Esko, J. D. *Journal of Biological Chemistry* **1994**, *269*, 28809-28814.



- (279) Kim, B. T.; Kitagawa, H.; Tamura, J.; Saito, T.; Kusche-Gullberg, M.; Lindahl, U.; Sugahara, K. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 7176-7181.
- (280) Kitagawa, H.; Tanaka, Y.; Tsuchida, K.; Goto, F.; Ogawa, T.; Lidholt, K.; Lindahl, U.; Sugahara, K. *Journal of Biological Chemistry* **1995**, *270*, 22190-22195.
- (281) Koike, T.; Izumikawa, T.; Tamura, J.-I.; Kitagawa, H. *Biochemical Journal* **2009**, *421*, 157-162.
- (282) Nadanaka, S.; Kagiya, S.; Kitagawa, H. *Biochemical Journal* **2013**, *454*, 133-145.
- (283) Busse-Wicher, M.; Wicher, K. B.; Kusche-Gullberg, M. *Matrix Biology* **2014**, *35*, 25-33.
- (284) Andre, I.; Tvaroska, I.; Carver, J. P. *Journal of Physical Chemistry A* **2000**, *104*, 4609-4617.
- (285) Mendoza, F.; Gomez, H.; Lluch, J. M.; Masgrau, L. *Acs Catalysis* **2016**, *6*, 2577-2589.
- (286) Jones, G.; Willett, P.; Glen, R. C. *Journal of Molecular Biology* **1995**, *245*, 43-53.
- (287) Ribeiro, A. J. M.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. *Acs Catalysis* **2015**, *5*, 5617-5626.