TESIS DOCTORAL

Modelado Molecular de la Interacción de Fármacos Antitumorales y Nucleasas con el ADN

Juan Antonio Bueren Calabuig

Alcalá de Henares, 2011

Departamento de Farmacología

Universidad de Alcalá

Universidad de Alcalá
DEPARTAMENTO DE FARMACOLOGÍA

FACULTAD DE MEDICINA
Campus Universitario
28871 Alcalá de Henares
Teléfono: 91 885 45 14
Fax: 91 885 45 91
federico.gago@uah.es

UNIVERSIDAD DE ALCALÁ. PATRIMONIO DE LA HUMANIDAD

Federico Gago Badenas, Catedrático y Director del Departamento de Farmacología de la Universidad de Alcalá,

INFORMA:

Que el trabajo titulado "Modelado Molecular de la Interacción de Fármacos Antitumorales y Nucleasas con el ADN" ha sido realizado en este Departamento, por **D. Juan Antonio Bueren Calabuig** y, a su juicio, cumple todos los requisitos para proceder a su defensa pública como Tesis Doctoral.

Alcalá de Henares, a 22 de Septiembre de 2011

Universidad de Alcalá

DEPARTAMENTO DE FARMACOLOGÍA

LABORATORIO DE MODELADO MOLECULAR

FACULTAD DE MEDICINA
Campus Universitario
28871 Alcalá de Henares
Teléfono: 91 885 45 14
Fax: 91 885 45 91
federico.gago@uah.es

UNIVERSIDAD DE ALCALÁ. PATRIMONIO DE LA HUMANIDAD

Federico Gago Badenas, Catedrático y Director del Departamento de Farmacología de la Universidad de Alcalá,

INFORMA:

Que el trabajo titulado "Modelado Molecular de la Interacción de Fármacos Antitumorales y Nucleasas con el ADN" ha sido realizado en este Departamento, por **D. Juan Antonio Bueren Calabuig** bajo su dirección y, a su juicio, cumple todos los requisitos para proceder a su defensa pública como Tesis Doctoral.

Alcalá de Henares, a 22 de Septiembre de 2011

AGRADECIMIENTOS

Mil gracias a Federico, por confiar en mí. Por enseñarme algo nuevo todos los días, por demostrarme como se debe trabajar. Gracias por ser tan buen maestro y estar siempre dispuesto a pasar la antorcha a tus alumnos. Por ayudarme con los artículos y por integrarme en todo el mundo investigador que te rodea y que te admira, gracias. Gracias a Antonio Jiménez por todas tus preguntas y más preguntas sobre modelado molecular que tanto me ayudaron y mil gracias por haber podido trabajar con vuestro grupo; ha sido una colaboración preciosa. Gracias a todas las personas de PharmaMar con las que he colaborado. Especialmente a Carlos Galmarini, Carmen Cuevas, Luis Francisco García y José María Fernández. Gracias por esta magnifica experiencia, por haberme enseñado vuestros grandes avances y haber trabajado con vosotros. Merci beaucoup à Jean Marc Égly et à Christophe. Je suis très heureux d'avoir travaillé avec vous.

Gracias Ana. Gracias por todo, como siempre. Me harían falta diez tesis como esta para darte las gracias. Por tus notas, tus ánimos, tu ayuda y tu confianza. Gracias por tus consejos y por ayudarme a solucionar todas mis cosas raras raras. Gracias NYC. Ha sido un regalo para mi haber estado contigo.

Muchas gracias Claire; gracias por estar conmigo durante todos estos años. Gracias por ser tan buena compañera (y amiga). Gracias por tus ganas de ayudarme siempre y por tus esfuerzos constantes. Gracias Alberto Gómez Ramos (de flores, ¡qué bonito!) por acompañarme en esta última etapa que me ha sido más fácil gracias a ti. Gracias Marta. Sobretodo, gracias por tu entusiasmo, tus ánimos y tu confianza en mí; por apoyarnos en la RANN de Jaca que nos trae tantos buenos recuerdos. Gracias Irene. Por tu alegría, tu ayuda y por los ratos con el moviemaker que tanto disfrutamos. Gracias a Antonio "el gallego". Por esos paseos con Marta por la Puerta del Sol currándonos el mejor amigo invisible de la historia. Gracias Pedro, por contagiarme tus ganas de aprender y por tus ratos en el labo. Grazie mille Antonellina; your short time here made us very happy. Gracias a Obis y a Luisma por vuestros buenos consejos y vuestras ganas de ayudar y colaborar.

Gracias David. Por ganarme siempre en la carreras a las 9 de la mañana (¡¿POR QUÉ?!) y por dejarte ganar en el pádel; gracias por tus ánimos, y tu ayuda desde el primer día que llegué aquí. Gracias Irene, porque eres un ejemplo para todos nosotros; tus esfuerzos, tus consejos y tu valor no se olvidarán. Gracias Pablo, por crear siempre buen ambiente a tu alrededor y por la alegría que trasmites. Gracias Raúl, por ser tan buen amigo y por entenderme tan bien. Y por supuesto, por ese viaje inolvidable a Valencia. Gracias al "Señor Doctor Profesor" Carlos; porque tus esfuerzos me animan a seguir trabajando día a día. Gracias Kilian, por tener todos y cada uno de los días una sonrisa y una palabra amable hacia mí. Gracias Vero por haber estado conmigo y haberme ayudado tanto los primeros días en Alcalá. Gracias Ana por estar siempre de buen humor conmigo y contagiarme tu alegría. Gracias Borja por tus visitas al labo que nunca te devolvía y que nos animaban tanto. Gracias a todos los compis de los distintos labos, Cris, Irene, Miguel, Nadia, Julia, Gisela, Andrés, Mali, Lauralicia (¡lo siento!), Omar, Ariel y todos los que han estado conmigo y que crearon el buen ambiente con el que he vivido todos estos años. Gracias a todos mis compañeros del Máster de Dianas. Gracias a Nieves, a Antonio, a Santiago, a Begoña, a Jose Antonio, a Maica, a Luis, a Inés y a todos los profesores e investigadores de la UAH que tanto me han animado. Gracias a Alberto Domingo por la colaboración con los experimentos de *melting*. Gracias a Angélica y a Rosa por vuestra ayuda.

Gracias a todos los colegas del CBM. Gracias por el congreso en memoria de Ángel. Gracias Antonio, por ayudarnos sin dudar todos los días y estar pendiente de nosotros como si estuviéramos allí. Mil gracias. Gracias a todos: a Álvaro, a Rubén, a David, a Alfonso (casi nos os hemos dado trabajo, ¿verdad?), a Helena, a Ugo, a Raúl, a Gonzalo, a Alberto y a Javi. Gracias Almu por tus consejos y ánimos. Soy muy afortunado de haber coincidido con vosotros. Gracias a Anita, Esthercita, Clairecita, Juanito, Ciemat, Bacchus, Medusa, Marenostrum y especialmente a Adelita.

Gracias a toda la gente de Barcelona. Gracias a Alberto Pérez por sus consejos y mil y una dudas que me resolvió con Curves y a Agnès por su ayuda. Gracias a Carlos y a Ignacio, aunque discrepemos en algún asuntillo. Gracias a Modesto Orozco y a Javier Luque, por sus consejos, su ayuda, su amabilidad y por recibirme tan bien siempre. Gracias también a Adrian Roitberg por sus consejos sobre QM/MM.

Gracias a mis alumnos de los Masters. Gracias por ayudarme y hacerme feliz mientras trataba de enseñar lo poco que sé. Gracias a Gonzalo por ayudarnos con las clases. Gracias al equipo de River, del Liceo, Alkaska, al de los martes y al los de la fila de detrás en el campo. Gracias a mis compañeros del camino de Santiago: a los Juanes y a Antonio que me enseñaron que siempre hay que seguir hacia delante. ¡Buen camino! Gracias a Sergio por ser mi compañero inseparable desde la facultad. Gracias por animarme siempre. Gracias Paula, Patri y Susana por vuestros ánimos y confianza. Gracias a los Caamaño y a todo Esteiro por ayudarme a coger fuerzas año tras año. Gracias a los chicos y a las chicas de Coruña. Gracias a Alcalá de Henares, a Nuestra Señora del Val y a la calle Mayor por acogerme. Y gracias también a Elena, Luis y Laura por hacerlo más fácil.

Gracias Eva; porque has conseguido que llegue hasta el final. Gracias por creer tanto en mí, por animarme, ayudarme y estar siempre a mi lado. Gracias por entenderme y hacerme feliz.

Gracias a toda mi familia. A María, a Yosune, a Teresa, a Anita, a Diegol y a Maca. Gracias a Almu y a Laurent por estar siempre conmigo. Gracias a Pula, a Buyo, a Rafa, a Ricardo y a todos y cada uno de mis primos. Gracias Tía Pepa, por echarme una mano cada vez que lo necesito. Gracias a todos mis tíos porque sé que siempre estáis ahí.

Gracias a Memé, a Caco, a la Yaya, a la Amoña y al abuelo Mariano.

Gracias Jesús porque pones siempre cordura en todo y me ayudas cada día. "*De hermano a hermano, vuestro en vida y muerte"*. Gracias a Ana y a Papá por animarme, empujarme y confiar en mí. Gracias por ser la roca que me sostiene para seguir adelante.

A Eva,

“Todo se consigue con ímprobo trabajo y tenaz perseverancia”
Mariano Calabuig

A Ana, a Papá y a Jesús

A la Yaya

ÍNDICE

ABREVIATURAS Y ACRÓNIMOS

ADN	Ácido desoxirribonucleico
AMBER	*Assisted model building with energy refinment;* Construcción asistida de modelos con refinamiento de energía
ARN	Ácido Ribonucleico
ARNm	ARN mensajero
ARNr	ARN ribosómico
ARNt	ARN de transferencia
BLAST	*Basic local alignment search tool;* Herramienta básica de búsqueda de alineamientos locales
DSBs	*Double strand breaks;* Roturas de doble cadena
ESP	*Electrostatic surface potential*; Potencial electrostático de superficie
FA	Anemia de Fanconi
HR	*Homologous recombination;* Recombinación homóloga
ICL	*Interstrand crosslink;* Entrecruzamiento intercatenario
MD	*Molecular dynamics,* Dinámica molecular
MEP	*Molecular electrostatic potential*; Potencial electróstatico molecular
MM	*Molecular mechanics*; Mecánica molecular
MMC	Mitomicina C
MoA	*Mechanism of action*; Mecanismo de acción
NER	*Nucleotide Excision repair;* Reparación por escisión de nucleótidos
NOE	*Nuclear Overhauser Effect*
PBC	*Periodic boundary conditions;* Condiciones de límite periódico
PME	*Particle mesh ewald*
QM	*Quantum mechanics*; Mecánica cuántica
QM/MM	*Quantum mechanics/Molecular mechanics*; Mecánica cuántica/Mecánica molecular
RESP	*Restrained electrostatic surface potential*; Potencial electrostático de superficie restringido
RMN	Resonancia magnética nuclear
RMSD	*Root mean square deviation;* Desviación cuadrática media
Vvn	Nucleasa de *Vibrio vulnificus*

RESUMEN / SUMMARY

Computational methods have become increasingly important in a number of areas such as characterization of ligand-binding sites in proteins, docking of small molecules into protein and DNA binding sites, and molecular dynamics simulations. The results obtained yield information that sometimes is beyond current experimental possibilities and can be used to guide and improve a vast array of experiments. The purpose of this memory is to study by means of molecular modelling techniques the interaction between DNA and different ligands including several antitumour drugs, different families of nucleases such as XPF and the nuclease of *Vibrio vulnificus*, and RNA polymerase II. The investigations were carried out in collaboration with different experimental groups from *Universidad de Alcalá* , Prof. Egly's group at *Institut de Génétique et de Biologie Moléculaire et Cellulaire in Strasbourg* and *PharmaMar* researchers and are divided in the following sections: **i)** Understanding the interaction of DNA with the new antitumour drug Zalypsis® and analysis of the sequence specificity using molecular modelling techniques. **ii)** Understanding the interaction of DNA with the new antitumour drug PM01183 and analysis of the sequence specificity using molecular modelling techniques. **iii)** Study of the interaction between DNA and the enzymes XPF and RNA polymerase II. Effects of the antitumour drugs Yondelis® and Zalypsis®. **iv)** Rationale for the opposite stereochemistry of the major monoadducts and interstrand crosslinks formed by mitomycin C and its decarbamoylated analogue at CpG steps in DNA and the effect of cytosine modification on reactivity **v)** Temperature-induced melting of double-stranded DNA in the absence and presence of covalently bonded antitumour drugs: insight from molecular dynamics simulations. **vi)** Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: *Vibrio vulnificus* periplasmic nuclease as a test case.

1. INTRODUCCIÓN

"*I am an optimist. It does not seem too much use being anything else.*"

Winston Churchill

1.1. Aspectos Generales del Modelado Molecular

El ácido desoxirribonucleico (ADN), responsable del almacenamiento de la información genética, constituye una de las dianas más clásicas en la terapia contra el cáncer (1-3). A pesar de la aparente sencillez de su estructura (cuatro tipos de pares de bases unidas por enlaces de hidrógeno en una doble hélice), es imprescindible emplear técnicas experimentales como la difracción de rayos X (sobre fibras o cristales) o la espectroscopía de resonancia magnética nuclear (RMN) para poder estudiar sus propiedades físicas y estructurales. Aunque la mera visualización de una única estructura molecular puede servir para responder algunas de las cuestiones más simples en cuanto a su función, ésta no es capaz de resolver problemas más complejos como, por ejemplo, el movimiento de una molécula o su unión a un ligando. En estos casos es necesario acudir a la simulación de tales procesos. El principal objetivo del modelado molecular por ordenador, como aproximación *in silico* a la bioquímica y a la farmacología, es la descripción de las interacciones biomoleculares en base a las leyes generales de la química y de la física. En la mayoría de los sistemas, las herramientas elegidas para estudiar grandes macromoléculas biológicas son la mecánica clásica, junto con los métodos de simulación de dinámica molecular (*Molecular Dynamics)* (MD) (4-6).

Los avances en modelado molecular y en las técnicas de simulación han demostrado ser de gran valor en numerosos campos de la investigación incluyendo el estudio de sistemas binarios y terciarios, como los complejos ADN-ligando, ADN-proteína y proteína-ADN-ligando. A la hora de desarrollar estos trabajos es imprescindible una acción interdisciplinar en la cual experimentos teóricos complementen a los experimentales y viceversa. Gracias a esta mutua colaboración se enriquecen las investigaciones y se generan resultados de mayor relevancia.

1.2. Bases Teóricas del Modelado Molecular

El término modelado molecular hace referencia al estudio de moléculas mediante modelos físicos que, al combinarse con un conjunto de reglas teóricas, permiten racionalizar datos experimentales, planificar futuras investigaciones y proporcionar una información que no es posible obtener con las herramientas puramente experimentales.

Hoy en día el modelado molecular está íntimamente ligado al uso de computadoras y gráficos interactivos. El espectacular desarrollo de los ordenadores y de las técnicas computacionales ha permitido la incorporación de modelos teóricos para el tratamiento de todos aquellos problemas relacionados con la geometría y la energía de las moléculas, hasta tal punto que, con suficiente poder de cálculo y un nivel de teoría elevado, se podría modelar cualquier sistema biológico. Sin embargo, en la práctica, esto no es siempre posible debido a limitaciones tanto de tiempo como de recursos materiales. Por este motivo, es necesario introducir aproximaciones para mantener la viabilidad del experimento aplicando distintos niveles de teoría en función del sistema que va a ser analizado. Como norma general, a medida que aumenta el nivel teórico aumenta la calidad de los resultados obtenidos, pero también se incrementa notablemente el coste computacional. La descripción más completa de un sistema viene definida por la mecánica cuántica (Q*uantum Mechanics*, QM) que tiene en cuenta en el cálculo tanto a los núcleos como a los electrones en el cálculo, haciendo posible el estudio de la estructura, las propiedades que dependen de la distribución electrónica y la reactividad química (formación y ruptura de enlaces) (7). Sin embargo, su aplicabilidad está restringida a sistemas con centenares de átomos como máximo, siendo inviable para aquellos constituidos por miles de átomos en ausencia de recursos de supercomputación. Por otro lado, la mecánica molecular (*Molecular Mechanics*) (MM) ignora los movimientos electrónicos y calcula la energía de una molécula o conjunto de moléculas únicamente en función de la disposición de los núcleos atómicos, por lo que resulta el nivel de teoría generalmente escogido para estudiar macromoléculas biológicas como el ADN o las proteínas. Ambas técnicas, no obstante, se pueden

combinar en aproximaciones de mecánica cuántica/mecánica molecular (QM/MM). Con este método híbrido, por ejemplo una pequeña parte del sistema, implicada en una reacción química, se trata mediante QM mientras que el resto de los átomos se considera mediante MM.

A continuación se describirán los distintitos niveles de teoría utilizados en la presente memoria: métodos cuánticos, métodos clásicos de MM y métodos híbridos de QM/MM.

1.2.1. Métodos Cuánticos

Dentro de la QM pueden diferenciarse, entre otros, los métodos *ab initio* y los métodos semiempíricos. *Ab initio* significa "desde el principio" o "a partir de primeros principios", lo que implica que un cálculo aplicando esta aproximación requiere únicamente constantes físicas sin que haya necesidad de añadir parámetros ajustados. Una de las aplicaciones prácticas de los métodos *ab initio* es la obtención del potencial electrostático molecular (*Molecular Electrostatic Potential*) (MEP). El MEP contiene toda la información sobre la distribución de carga nuclear y electrónica de una molécula y, por tanto, es de gran utilidad para dar una idea de la reactividad del compuesto (**Figura 1**).

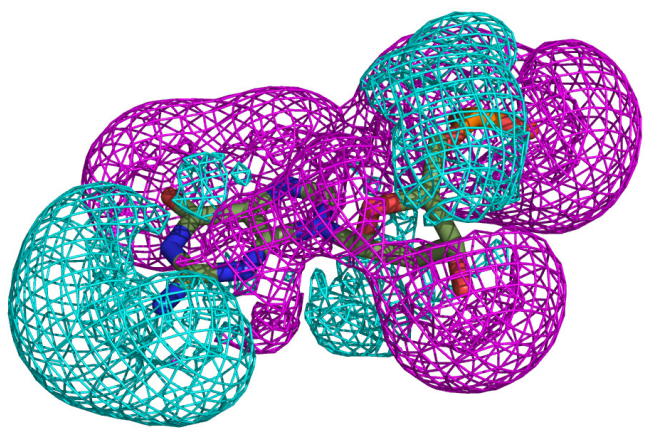


Figura 1. Potencial electrostático molecular (MEP) de una guanosina monofosfato. En rosa, valores negativos y en azul valores positivos.

Los métodos cuánticos *ab initio* permiten igualmente calcular cargas parciales que se asignan a cada átomo presente en la molécula. Son las llamadas cargas atómicas. En los métodos clásicos, que no tratan los electrones explícitamente, estas cargas atómicas parciales se utilizan para derivar las interacciones electrostáticas, generalmente mediante un potencial coulómbico. La calidad de estas cargas, por lo tanto, es crítica para la validez de los cálculos clásicos. Los métodos generalmente utilizados para derivar dichas cargas, por su mayor precisión, las obtienen a partir del MEP o ESP (*electrostatic surface potencial)*, el cual deben reproducir. El proceso se refinó para el campo de fuerzas de AMBER (*Assisted Model Building with Energy Refinment)* dando lugar a la metodología RESP (ajuste mediante *restrained electrostatic surface potencial*) (8-10).

Los cálculos *ab initio* pueden llegar a ser extremadamente costosos en cuanto a recursos computacionales. Por este motivo se emplean métodos de semiempíricos aproximados que, al incorporar parámetros derivados de datos experimentales, pueden calcular propiedades del sistema de una forma relativamente precisa con un gasto computacional mucho menor. Estos métodos se emplean normalmente a la hora de simular reacciones químicas de ruptura y formación de enlaces y poseen a su vez distintos niveles de teoría.

1.2.2. Métodos Clásicos

Los métodos de campo de fuerzas o métodos clásicos se basan en la aproximación de Born-Oppenheimer, que considera la energía potencial de los átomos de manera explícita mientras que el movimiento de los electrones se trata implícitamente.

1.2.2.1. El Campo de Fuerzas

Un campo de fuerzas está formado por dos componentes: las funciones de energía potencial, definidas por un conjunto de ecuaciones que engloban una serie de contribuciones energéticas enlazadas y no enlazadas, y por otro lado, los

parámetros empíricos usados por cada uno de los términos. Una de las características principales del campo de fuerzas es la transferencia de parámetros, de manera que haciendo la parametrización sobre un pequeño número de casos, se puede extrapolar a un amplio espectro de moléculas reproduciendo de manera fiable los resultados con un coste mucho menor. La mayoría de campos de fuerzas actuales tienen formas muy similares pero en esta memoria nos centraremos principalmente en el campo de fuerzas de AMBER (11).

La expresión matemática que define la energía potencial de un sistema (E_{sis}) en un campo de fuerzas se define como el sumatorio de las energías de interacción entre pares de átomos enlazados entre sí y también entre aquellos que no se encuentran unidos covalentemente. Se puede definir mediante la siguiente ecuación:

$$E_{sis} = E_{términos\ enlazados} + E_{términos\ no\text{-}enlazados} \qquad (1.1)$$

Los términos enlazados incluyen contribuciones debidas a los enlaces covalentes, ángulos de valencia y ángulos torsionales propios e impropios. Los términos no-enlazados se definen por un término de atracción-repulsión de tipo Lennard-Jones para las fuerzas de van der Waals y un término Coulómbico para las interacciones electrostáticas.

i) Términos Enlazados

a) Término de Enlace (*Bond Stretching*)

La energía potencial para un enlace covalente entre dos átomos se define mediante una función de Morse que incluye tres parámetros: energía de disociación, valor de referencia y constante de fuerza del enlace. Sin embargo, el potencial de Morse no se emplea en campos de fuerza de mecánica molecular dado que raras veces los enlaces se desvían significativamente de sus valores de equilibrio. Por lo tanto se consideran expresiones computacionalmente más eficientes, como el potencial de Hooke:

$$E_{enlace} = (k_i /2)(l_i - l_{eq})^2 \qquad (1.2)$$

donde l_i corresponde a la distancia de enlace, l_{eq} al valor de equilibrio y k_i a la constante de fuerza del enlace **(Figuras 2 y 3).**

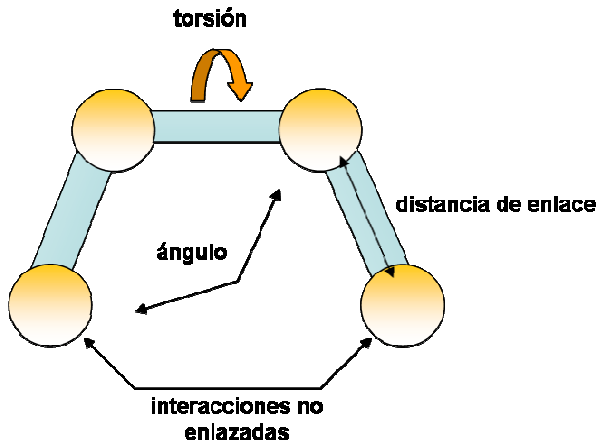


Figura 2. Esquema de las interacciones enlazadas y no enlazadas que se tienen en cuenta en un campo de fuerzas de mecánica molecular.

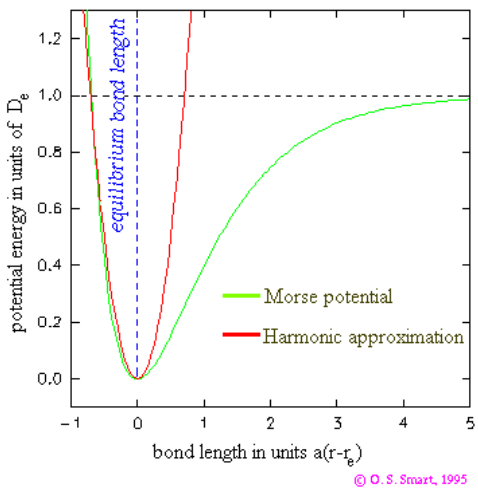


Figura 3. Forma del potencial de enlace según una representación de Morse y según una representación armónica.

b) Término de Ángulo (*Angle Bending*)

La desviación de los ángulos de sus valores de referencia también se describe mediante una un potencial armónico de Hooke:

$$E_{ángulo} = (k_i/2)(\theta i - \theta_{eq})^2 \qquad (1.3)$$

donde k_i corresponde a la constante de fuerza angular y θ_{eq} constituye el valor de equilibrio para ese ángulo.

c) Términos de Torsión

El término de torsión describe las barreras energéticas de rotación alrededor de un enlace. A diferencia de los anteriores, no tiene un comportamiento armónico basal, sino que presenta una periodicidad que hace necesario expresar su perfil energético como una serie de Fourier:

$$E_{tor} = \sum_{tor}\sum_{n} \frac{V_n}{2} [1+\cos (n\Phi - n)] \qquad (1.4)$$

donde V_n es la altura de la barrera de torsión, n la multiplicidad (el número de mínimos en la función cuando se rota 360º), Φ el ángulo diedro y n el ángulo de fase (indica en qué punto pasa la torsión por el mínimo energético). Este tipo de torsión se denomina *torsión propia* para diferenciarla de la *torsión impropia* que se utiliza para mantener la quiralidad y la planaridad

ii) Términos No Enlazados

a) Término Electrostático

La distribución electrónica de una molécula se puede representar a partir de las cargas puntuales fraccionales centradas en los núcleos. Estas cargas reproducirán la energía electrostática del sistema que se representa habitualmente usando la ley de Coulomb:

$$E_{ele} = \sum_{i} \sum_{j>i} \frac{q_i q_j}{4 \pi \varepsilon_0 r_{ij}} \qquad \textbf{(1.5)}$$

donde q_i y q_j son las cargas puntuales de cada átomo, r_{ij} la distancia entre ellos y ε la constante dieléctrica del medio que las separa. Si las cargas puntuales de dos átomos son contrarias éstos se atraerán entre sí, pero si las cargas son del mismo signo se repelerán **(Figura 4).**

b) Término de van der Waals

La interacción de van der Waals entre dos átomos se origina a partir de un balance entre fuerzas atractivas y repulsivas. Los núcleos atómicos tienden a repelerse unos a otros cuando están muy cerca y a atraerse cuando están a una distancia internuclear óptima (**Figura 4)**. Este comportamiento se puede describir mediante un potencial de Lennard-Jones, cuyas exponenciales son típicamente 6 y 12:

$$E_{Lennard\ Jones} = \sum_{i} \sum_{j>i} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \qquad \textbf{(1.6)}$$

donde A_{ij} y B_{ij} son constantes para cada par de átomos relacionadas con sus radios de van der Waals y r_{ij} es la distancia entre ellos.

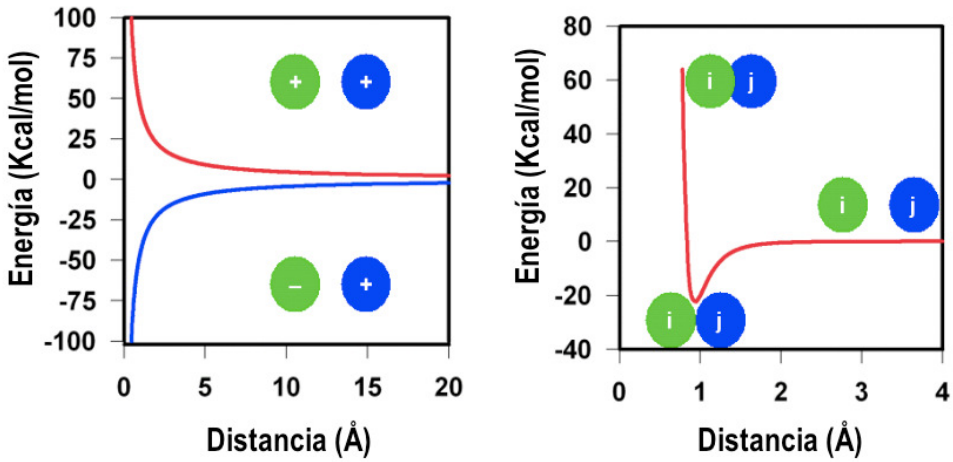


Figura 4. Energías de los términos no enlazados

iii) Parametrización

Una correcta parametrización de los términos mencionados anteriormente (enlazados y no enlazados) es imprescindible para aplicar un campo de fuerzas óptimo que asegure la calidad y veracidad de los cálculos posteriores. Los campos de fuerzas actuales han sido cuidadosamente parametrizados usando datos experimentales y *ab initio* en sistemas modelo utilizados como referencia.

1.2.2.2. Métodos Basados en Campos de Fuerza

Podemos utilizar un campo de fuerzas parametrizado para poder explorar la superficie de energía potencial de un sistema asociada a la configuración de sus átomos. Para ello se pueden aplicar distintas metodologías destacando la MM y la MD.

i) Mecánica Molecular: Minimización de Energía

La superficie de energía potencial de un sistema viene definida por el modo en el que la energía de las moléculas varía en función de sus coordenadas. En modelado molecular, es necesario conocer los puntos mínimos de la superficie energética que corresponden a los estados estables del sistema. El proceso que permite identificar las geometrías del sistema que corresponden a los puntos de mínima energía potencial se denomina minimización **(Figura 5).** Dado que la superficie de energía potencial de los sistemas macromoleculares es compleja, encontraremos un mínimo global y varios mínimos locales. La estructura minimizada suele corresponder a un mínimo energético relativo y no al global. Esta técnica es idónea para reducir las interacciones desfavorables que tenga el sistema de partida, preparándola para la exploración de la superficie de potencial con otros métodos como la MD.

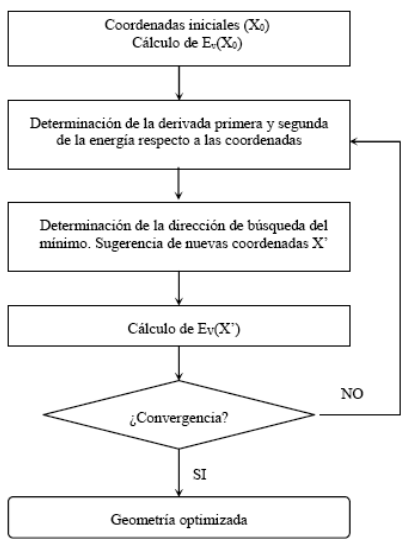


Figura 5. Algoritmo básico de los métodos de mecánica molecular

Matemáticamente, un mínimo puede ser descrito de la siguiente forma:

$$\frac{\delta f}{\delta x_i} = 0 \qquad \frac{\delta^2 f}{\delta x_i^2} > 0 \tag{1.7}$$

siendo f la función de potencial y x_i las coordenadas cartesianas (o internas) del sistema. Por lo tanto, en un punto mínimo la primera derivada de la función con respecto a las coordenadas será igual a cero mientras que las segundas derivadas serán todas positivas.

Los métodos más empleados para la minimización de energía son justamente los basados en las derivadas de la función de potencial, especialmente el de "descenso más pronunciado" (*steepest descent*) y el de "gradiente conjugado" (*conjugate gradient*). Ambos modifican las coordenadas de los átomos mientras desplazan el sistema más y más cerca del punto de mínima energía. El primero utiliza una dirección de búsqueda basado en el máximo gradiente (o máxima pendiente hacia el mínimo), siendo idóneo cuando la posición de partida está alejada del mínimo, mientras que el segundo se basa en el movimiento en una dirección que conjuga la del gradiente con las direcciones previas de búsqueda, por lo que resulta más eficiente cuando la búsqueda está más cercana al mínimo. En la práctica, suele realizarse una combinación de ambos, usando primero *steepest descent* y *conjugate gradient* después.

La minimización de energía genera configuraciones individuales de mínima energía que, en muchos casos, pueden ser suficientes para predecir ciertas propiedades de un sistema. Sin embargo, el conocimiento de la configuración de una sola estructura, incluso si ésta pertenece al mínimo global, puede ser insuficiente para llevar a cabo algunos estudios. Por ello, las técnicas de minimización son idóneas para preparar el sistema para una exploración más exhaustiva de la superficie de potencial empleando otros métodos como la MD.

ii) Dinámica Molecular

La aplicación de la MD permite generar configuraciones sucesivas del sistema integrando las leyes del movimiento de Newton. El resultado es una trayectoria que especifica cómo las posiciones y las velocidades varían con el tiempo. Se trata por lo tanto de un método determinista, es decir, las posiciones de los átomos se suceden en la escala temporal.

La trayectoria se obtiene resolviendo la ecuación diferencial contenida en la segunda ley de Newton (*F=ma)*:

$$\frac{d^2x_i}{dt^2} = \frac{F_i}{m_i} \qquad \textbf{(1.8)}$$

que describe el movimiento de una partícula i de masa m_i a lo largo de la coordenada x_i, siendo F_i la fuerza aplicada y t el tiempo.

A partir de unas coordenadas (x_o) y velocidades iniciales (v_0) se determina la energía potencial y la energía cinética del sistema, siendo la energía total la suma de ambas:

$$\boldsymbol{Etot = Epot\ (x_0) + E_k\ (v_0)} \qquad \textbf{(1.9)}$$

Para seguir la evolución temporal del sistema se recurre a métodos de integración numérica. De este modo se obtienen pequeñas etapas sucesivas separadas en el tiempo por un intervalo fijo δt (tiempo de integración). La unión de todos estos fragmentos genera la trayectoria del sistema en el tiempo total de la simulación.

La fuerza que actúa sobre cada núcleo en un instante de tiempo *t* se determina mediante la derivada de la energía potencial con respecto a las coordenadas.

$$\boldsymbol{F = -\delta_{Epot}/\delta x} \qquad \textbf{(1.10)}$$

Sumando sobre cada núcleo las fuerzas de interacción con las demás partículas del sistema se obtiene la fuerza total a partir de la cual podemos determinar las aceleraciones de las partículas (ecuación 1.8) que se combinan a continuación con las posiciones y velocidades a tiempo *t* para calcular las posiciones

y velocidades a tiempo $t + \delta t$. Se asume que la fuerza es constante durante cada paso de integración. Para cada uno de ellos se computan las fuerzas que actúan sobre los átomos y se combinan con las velocidades actuales para generar nuevas posiciones y nuevas velocidades.

Todos los algoritmos empleados en MD que integran las ecuaciones de movimiento asumen que las posiciones y las propiedades dinámicas pueden ser aproximadas mediante series de Taylor:

$$\begin{aligned} xi(t+\delta t) &= xi(t) + \delta t vi(t) + 1/2\ \delta t2ai(t) + \ldots \\ vi(t+\delta t) &= vi(t) + \delta t ai(t) + 1/2\ \delta t2bi(t) + \ldots \\ ai(t+\delta t) &= ai(t) + \delta t bi(t) + 1/2\ \delta t2ci(t) + \ldots \end{aligned} \qquad \textbf{1.11}$$

donde x es la posición, v la velocidad (primera derivada) y a la aceleración (segunda derivada). El algoritmo de Verlet (12) es el método más utilizado para integrar las ecuaciones de movimiento en simulaciones de MD. Éste usa las posiciones y aceleraciones a tiempo t y las posiciones en la etapa previa, $x(t-\delta t)$ para calcular las nuevas posiciones a $t+\delta t$, $x(t+\delta)$.

La elección del tiempo de integración, δt, es crucial, ya que durante este tiempo las fuerzas no se recalculan y el sistema tiende a seguir en el mismo estado. Un tiempo de integración demasiado grande daría lugar a inestabilidades por solapamiento de altas energías entre átomos mientras que un tiempo demasiado pequeño cubriría solamente una limitada porción del espacio conformacional **(Figura 6)**

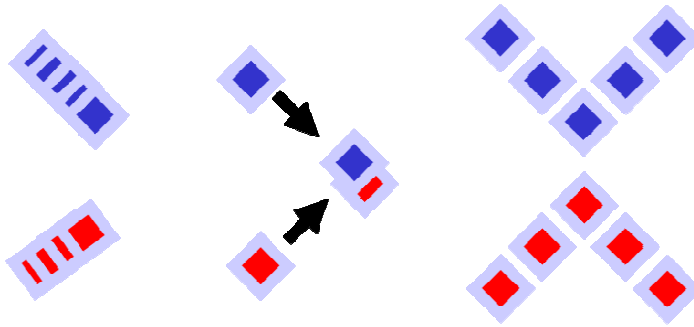


Figura 6. Efecto de diferentes tiempos de integración. Los rombos de distinto color representan dos átomos en un paso de la trayectoria. Con un tiempo de integración muy pequeño (izquierda), la capacidad para explorar el espacio conformacional queda muy limitada; un tiempo de integración demasiado grande (centro) da lugar a inestabilidades; con un tiempo de integración adecuado (derecha) el espacio conformacional se cubre de manera eficiente.

En el caso de las biomoléculas, los movimientos de frecuencia más rápida son los debidos a las vibraciones de los enlaces, que tienen lugar en la escala del femtosegundo (fs). Sin embargo estos movimientos rápidos son de poco interés ya que contribuyen poco al movimiento global. Por lo tanto se suele fijar la longitud de los enlaces covalentes a sus valores de equilibrio mediante el algoritmo *SHAKE* (13). Al evitar estos rápidos movimientos, se puede incrementar el tiempo de integración a 2 fs sin introducir inestabilidades y reduciendo el coste computacional.

Por otro lado, el cálculo de los términos no enlazados en una MD es posiblemente el proceso más costoso computacionalmente. Para acelerar este proceso, se suele aplicar otra aproximación en la cual las interacciones entre dos átomos separados por una distancia mayor a una distancia previamente definida (*cutoff*) no serán tenidas en cuenta. Sin embargo, el uso de distancias de corte afecta fundamentalmente al término electrostático ya que esta interacción es de largo alcance, y la omisión de parte de los pares de átomos del sistema en una simulación

de MD puede producir artefactos y conducir a propiedades del sistema no realistas. Por ello es recomendable el uso de métodos como el sumatorio de *Ewald* (PME) que permite corregir el sistema al sumar las interacciones con infinitos sistemas réplicas del original. El método se basa en dividir la suma de las interacciones en dos sumatorios: el espacio directo y el espacio recíproco.

También hay que tener muy presentes las anomalías que pueden surgir con las moléculas que se encuentran en el borde del área de simulación. Para evitar que estas partículas experimenten un comportamiento diferente a las demás debido a la naturaleza finita del sistema, se aplican *condiciones periódicas de contorno* o *condiciones de límite periódico* (*Periodic boundary conditions o PBC)* que permiten el cálculo de la simulación como si el sistema fuera infinito en el espacio.

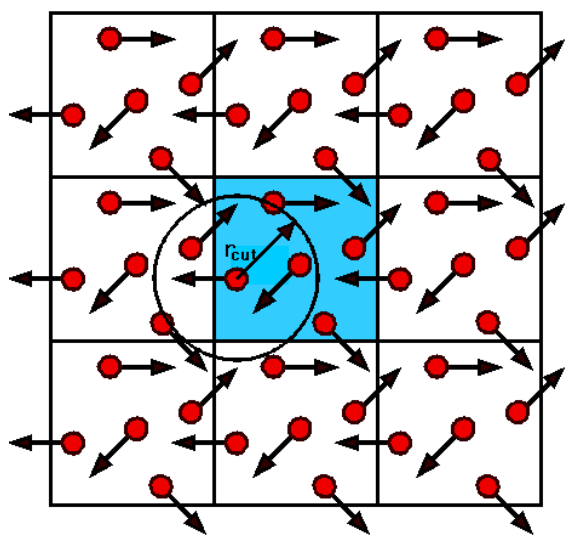


Figura 7. Condiciones de límite periódico en dos dimensiones (imagen del *UK's Science and Technology Facilities Council*).

La **Figura 7** ilustra el concepto de las PBC. La caja azul representa el sistema que estamos simulando mientras que las cajas que lo rodean son copias exactas con todo detalle: cada partícula en la simulación tiene un duplicado exacto en cada celda que la rodea. Incluso las velocidades (flechas) son las mismas. El resultado es que aunque

un átomo se vaya de la celda de la simulación, éste se reemplaza por otro que entra por la cara contraria, exactamente con la misma velocidad. De esta manera el número de átomos en la celda se conserva y además ninguna partícula sufre los efectos de las fuerzas del límite de la simulación. La forma y tamaño de las cajas depende de la geometría del sistema, usándose cajas octaédricas y rectangulares para el estudio de ácidos nucleicos y proteínas (**Figura 8**).

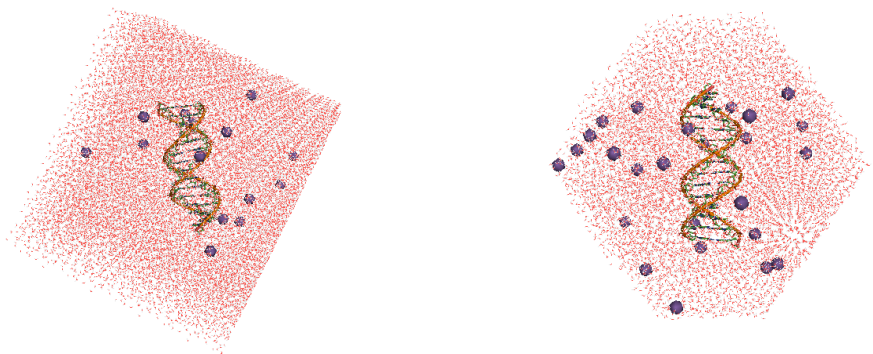


Figura 8. Caja rectangular y octaedro truncado

Otra consideración importante es el tamaño de la caja que debe ser tal que el soluto no pueda interaccionar con una copia de sí mismo representada en el espacio, por lo cual debe haber suficiente disolvente en cada dirección del soluto para que la distancia desde cualquier punto de éste hasta cualquiera de sus copias vecinas sea mayor que la distancia de corte (*cutoff*) establecida para las interacciones no enlazadas.

Preparación y ejecución de una MD

En esta sección discutiremos algunas etapas del desarrollo de las simulaciones de MD bajo las condiciones y aproximaciones que han sido descritas anteriormente. En primer lugar, es necesario establecer una configuración inicial del sistema. Dicha estructura puede provenir de datos experimentales de rayos X, RMN,

de un modelo teórico o de una combinación de datos experimentales y teóricos. La posterior asignación de velocidades iniciales a los átomos de esta estructura (previamente minimizada) normalmente se realiza al azar a partir de una distribución Maxwell-Boltzmann a la temperatura elegida. Finalmente, es preciso mantener fijas algunas de las condiciones de simulación: número de partículas (N), volumen (V), temperatura (T), presión (P) o energía total del sistema (E). En función de sus combinaciones se puede distinguir entre simulaciones microcanónicas (NVE), isotérmico-isobáricas (NPT) y canónicas (NVT) siendo NPT y NVT las más empleadas en MD de proteínas y ácidos nucleicos.

Si se utiliza un formalismo NPT, es necesario ajustar tanto la temperatura como la presión. Existen distintos métodos para ajustar la temperatura a un determinado valor. La manera más fácil es aplicar un escalado de velocidades (14) en el cual las velocidades se multiplican por un factor constante en cada paso de calentamiento. Un método alternativo para mantener la temperatura es acoplar el sistema a un baño térmico externo que se fija a la temperatura deseada (15). El baño actúa como fuente de energía térmica añadiendo o quitando calor al sistema según sea necesario. Las velocidades se escalan en cada paso en función de la diferencia de temperatura entre el baño y el sistema.

Muchos de los métodos usados para el control de la temperatura se emplean también a la hora de controlar la presión. Por lo tanto, la presión puede ser mantenida a un valor constante simplemente por un escalado del volumen. Y la alternativa es acoplar el sistema a un "baño de presión" análogo al baño de temperatura.

Fases de la simulación

Las simulaciones de MD se componen de dos etapas: una fase de equilibrado y una fase de producción. El objetivo del equilibrado es llevar al sistema a un estado de equilibrio a partir de la configuración inicial. Durante esta fase se monitorizan varios parámetros como la energía potencial, la temperatura y la densidad hasta que llegan a un valor estable. Una buena técnica para conseguir un equilibrado óptimo es

aplicar restricciones al sistema, liberándolas después lentamente para permitir su adaptación a las condiciones deseadas. Este proceso suele durar entre 200 y 500 picosegundos aunque en ciertos casos conviene equilibrar el sistema durante varios nanosegundos.

Después de un correcto equilibrado comienza la fase de producción en la cual permitimos la evolución del sistema que queda recogida en forma de coordenadas y velocidades. Éstas se almacenan como una trayectoria que será utilizada para el análisis y la obtención de información. Cuanto mayor sea el tiempo de simulación, más calidad tendrán los resultados ya que se explorará más espacio conformacional. En general se estima que el tiempo de simulación debería ser al menos 10 veces más largo que la escala temporal del proceso a estudiar. En la práctica esto significa que con los ordenadores actuales sólo se pueden simular procesos que tienen lugar en la escala de unos pocos nanosegundos aunque gracias a la mejora en los algoritmos de paralelización, la capacidad de almacenamiento en discos y el uso de supercomputadores se ha podido pasar de simulaciones de 10 ns a las de 1 microsegundo (16).

Tipos de simulaciones

El espacio conformacional accesible a las macromoléculas biológicas es enorme y no puede ser cubierto por ninguna de las simulaciones actuales. Los datos experimentales (rayos X, RMN, etc.) indican que las biomoléculas tienden a adoptar conformaciones bastante definidas. Esta información experimental se puede utilizar en una simulación para restringir el espacio conformacional al que puede acceder el sistema, fundamentalmente dificultando la adopción de configuraciones que resultan incompatibles con la información experimental mediante una función de penalización. Cuanto más se aleje de los datos experimentales la configuración simulada del sistema, mayor será el valor de la función de penalización. Entre los tipos de parámetros que se pueden incorporar para forzar la trayectoria de modo que se satisfaga la información experimental se encuentran (i) los límites superiores de distancias interatómicas, por ejemplo, provenientes de medidas NOE (*Nuclear*

Overhauser Effect), (ii) los límites para los valores que pueden adoptar ciertos ángulos torsionales, a partir de datos de acoplamiento J obtenidos por espectroscopía de RMN, o (iii) las amplitudes de los factores de estructura medidos por difracción de rayos X. Además de las simulaciones con restricciones (*restrained*) y sin restricciones (*unrestrained*), se pueden considerar otra serie de posibilidades igualmente utilizadas, como por ejemplo:

(1) Templado simulado (*simulated annealing*) o dinámica molecular "amortiguada" (*quenched* molecular dynamics): se lleva a cabo una búsqueda grosera a una temperatura elevada para después "enfriar" lentamente en búsqueda del mínimo local más próximo. La temperatura del sistema no hay que considerarla como una temperatura física sino más bien como un parámetro de control que determina si el sistema puede escapar de ciertos mínimos locales. Este protocolo resulta particularmente útil para refinar estructuras procedentes de espectroscopía de RMN o de difracción de rayos X.

(2) Muestreo aumentado localmente (*Locally Enhanced Sampling*, LES): esta técnica permite aumentar el muestreo de la región de interés (por ejemplo, un asa o bucle de una proteína, un ligando en su sitio de unión, etc.) mediante la construcción de múltiples copias de esa región o grupo de átomos. Las copias no interaccionan entre sí e interaccionan con el sistema de una forma promediada. Este tipo de protocolo reduce las barreras o las transiciones conformacionales y permite obtener múltiples trayectorias de esta región llevando a cabo una única simulación.

(3) Dinámica molecular "activada", "sesgada" (*biased*) o "guiada" (*steered*): la adición de fuerzas externas predeterminadas reduce las barreras energéticas y aumenta la probabilidad de obtener configuraciones poco probables de observar en una simulación convencional. Este procedimiento resulta adecuado para estudiar procesos que son intrínsecamente rápidos pero que constituyen acontecimientos raros (frecuencia promedio << 10^{11} s^{-1}) porque están limitados por una o más barreras de energía.

(4) Forzado sobre un molde o "plantilla" (*template forcing*) o MD dirigida a un blanco (*targeted molecular dynamics*): a la función de energía potencial se añade un término adicional basado en la desviación cuadrática media, ponderada por la masa, de un conjunto de átomos en la estructura simulada con respecto a los mismo átomos en la estructura de referencia (el "blanco" o "diana"). En cada paso de MD, el algoritmo realiza un ajuste de mínimos cuadrados para los átomos seleccionados en las dos estructuras. Este método puede resultar idóneo para simular los grandes cambios conformacionales que ocurren en una escala de tiempos de milisegundos.

Análisis de resultados

En la mayoría de las simulaciones de MD realizadas en esta tesis se guardaron las coordenadas cada 2 ps durante unos tiempos de simulación entre 10 y 200 ns, lo que nos ha permitido obtener una representación muy exhaustiva de los sistemas a estudiar. En cada trayectoria podemos analizar distintos parámetros que nos aportan información acerca de los cambios estructurales que afectan a las moléculas, como son distancias entre átomos, ángulos, cambios en torsionales, parámetros helicoidales del ADN, amplitud de los surcos, ángulos diedros del esqueleto azúcar-fosfato, análisis de superficie accesible al disolvente (SAS), etc…

Otro parámetro importante a la hora de analizar trayectorias de MD es la desviación cuadrática media (*root mean square deviation*, RMSD) respecto a una estructura de referencia, normalmente la estructura promedio del colectivo o la estructura de partida. La función que representa la evolución del RMSD con respecto al tiempo es uno de los mejores indicadores del comportamiento del sistema durante la simulación. Si éste se halla equilibrado, los sucesivos valores del RMSD se mantienen aproximadamente constantes, fluctuando en torno a una o varias conformaciones de equilibrio.

1.2.3.3. Cálculo de Diferencias de Energía Libre de Unión (MM-GBSA)

La energía libre de unión que describe la asociación proteína–ligando o proteína–proteína es la diferencia entre la energía libre, G, del complejo y la de la suma de sus respectivas parejas:

$$\Delta G_{unión} = G_{complejo} - (G_{receptor} + G_{ligando}) \quad (1.12a)$$

$$\Delta G_{unión} = G_{complejo} - (G_{proteína} + G_{proteína}) \quad (1.12b)$$

En estas ecuaciones los tres valores de G normalmente indican promedios sobre varios pasos de las trayectorias de MD independientes del sistema molecular en equilibrio. Pero, si queremos estudiar la evolución a lo largo del tiempo de esta energía de interacción, cada uno de los pasos de la trayectoria puede ser analizado individualmente.

$$\Delta G_{unión} = G_{complejo\ (i)} - (G_{proteína\ (i)} + G_{ligando\ (i)}) \quad (1.13)$$

donde i indica cada "snapshot" de la trayectoria de MD. El valor de G para cada especie puede descomponerse en los siguientes términos:

$$G = Egas + Gsol - TS$$
$$Egas = Eint + Eele + Evdw \quad (1.14)$$
$$Eint = Eenlace + Eángulo + Etorsión$$

$$Gsol = GGB + Gno\text{-}polar \quad (1.15)$$

$$Gno\text{-}polar = \gamma\ SASA \quad (1.16)$$

donde Egas es la energía en fase gas calculada usando el campo de fuerzas de AMBER como la suma de la energía interna (Eint), que a su vez es la suma de la energía de

enlace (Eenlace), ángulo (Eángulo) y torsión (Etorsión), y las energías no enlazadas culómbicas (Eele) y de van der Waals (Evdw). La energía de solvatación (Gsol) se descompone en contribuciones polares y no polares. La contribución polar (GGB) se calcula resolviendo la ecuación del generalizado de Born (17) mientras que la contribución apolar se estima de la superficie accessible al disolvente (SASA) (18) determinada con una sonda de una molécula de agua de radio 1.4 Å. γ es la constante de tensión de superficie. T y S son la temperatura y la entropía, respectivamente. S se calcula haciendo uso de la estadística clásica y análisis de modos normales de "snapshots" representativos y previamente minimizados de una trayectoria de MD. Estas mismas ecuaciones se pueden aplicar para estudiar la interacción de ligandos con el ADN

1.2.3. Métodos Híbridos: QM/MM

Esta técnica permite combinar potenciales mecánico-cuánticos (QM) y mecánico-moleculares (MM) en un potencial híbrido QM/MM (7,19-21). En este tipo de aproximación las regiones de la proteína (o ADN) y el sustrato directamente implicados en una reacción química se analizan mediante una función QM mientras que el potencial de los demás átomos del sistema se examina por cálculos clásicos de MM. Este método combina la simplicidad y velocidad del tratamiento MM con el potencial de la química cuántica que permite el modelado de la formación y rotura de enlaces, así como la inclusión de la polarización electrónica. Al dividir el sistema, somos capaces de realizar cálculos significativamente más grandes que los que hubieran sido posibles únicamente con una aproximación QM pura y al mismo tiempo podemos estudiar reacciones químicas que no son susceptibles de un modelado riguroso mediante MM. En el caso de un complejo fármaco-ADN, la región QM sería naturalmente el fármaco y la(s) bases covalentemente unida(s) a él mientras que la influencia mecánica y electrostática de los demás nucleótidos y del disolvente se tratan con métodos de campos de fuerza **(Figura 9)**.

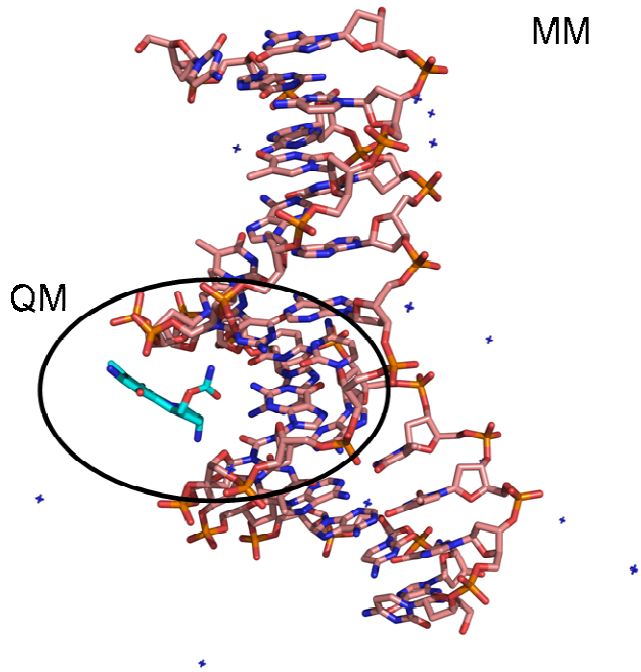


Figura 9. Ejemplo de un esquema QM/MM: el fármaco y los residuos que interaccionan con él pertenecen a la región QM; el resto del ADN, contraiones (cruces azules) y el disolvente (no mostrado) pertenecen a la región MM.

La energía total *Etot* para este tipo de sistemas se puede escribir de la siguiente forma:

$$Etot = E_{QM} + E_{MM} + E_{QM/MM} \qquad (1.17)$$

donde E_{QM} y E_{MM} corresponden a la energía de aquellas partes del sistema tratadas exclusivamente con mecánica cuántica y mecánica molecular, respectivamente. $E_{QM/MM}$ es la energía de interacción entre las partes mecánico-cuánticas y mecánico-moleculares.

Se han implementado distintos métodos con el fin de estudiar un sistema mediante QM/MM, los cuales difieren entre sí por el nivel de teoría utilizado para la

mecánica cuántica (semiempírico, *ab initio,* enlace de valencia o funcional de densidad), por el modelo de mecánica molecular y por el modo de representar el disolvente (explícitamente o usando un modelo simplificado). Otra diferencia importante es el modo de tratar la unión entre las regiones QM/MM. En general, es preferible cortar enlaces no polares (como enlaces sencillos C–C) que cortar enlaces insaturados o polares. Existen dos métodos para abordar este problema. En una primera aproximación, se establece un orbital híbrido sp^2 que contiene un electrón a lo largo de la región QM/MM (22). El método alternativo y más utilizado incluye simplemente un *"link" atom* o átomo enlazante (normalmente hidrógeno) que asegura la conservación de la valencia.

Por lo tanto, el empleo de cálculos mixtos QM/MM es una herramienta adecuada para estudiar reacciones enzimáticas, unión covalente de fármacos a proteínas y a ADN.

1.3 Parámetros Estructurales del ADN Estudiados por Modelado Molecular

1.3.1. Introducción

Los ácidos nucleicos son macromoléculas formadas por unidades monoméricas, los nucleótidos, y se clasifican en ácidos desoxirribonucleicos (ADN) y en ácidos ribonucleicos (ARN). El papel principal del ADN es almacenar la información genética que se transmitirá de una generación a la siguiente. Se distinguen distintos tipos de ARN: i) el ARN mensajero (ARNm) transporta la información genética procedente del ADN al ribosoma para sintetizar la proteína que se encuentra en él codificada, ii) el ARN ribosómico (ARNr) es un componente de los ribosomas y se encarga de unirse al ARNm para realizar la síntesis proteica, y iii) el ARN de transferencia (ARNt) transfiere los aminoácidos a los ribosomas durante la elongación de la cadena polipeptídica.

El ADN fue descubierto en 1868 por Johan Friedrich Miescher aunque no fue considerado como el material genético hasta casi un siglo después cuando, en 1952,

Martha Chase y Alfred Hershey lo demostraron utilizando bacteriófagos para transferir el ADN a bacterias (23), (24). En el mismo año Rosalind Franklin, como resultado de la difracción de rayos X sobre el ADN, consiguió la famosa "Fotografía 51", que sería fundamental para que en 1953 James Watson y Francis Crick publicaran la estructura del ADN (25).

Para poder comprender las funciones biológicas del ADN, así como su interacción con proteínas o pequeños ligandos, es necesario conocer sus características estructurales. Hoy en día esto se puede realizar con un nivel de detalle atomístico gracias al desarrollo de técnicas como la espectroscopía de RMN o la cristalografía de rayos X. A continuación se describirá la estructura interna del ADN, sus posibles conformaciones y sus propiedades de unión a pequeños ligandos y proteínas.

1.3.2. Estructura Interna del ADN

La unidad básica y repetida en el ADN y en el ARN es el nucleótido (**Figura 10**). Cada nucleótido está formado por un azúcar con un anillo furanósido (β-D-2'-desoxirribosa en el ADN o β-D-ribosa en el ARN) sustituido en el C1' por una base nitrogenada y un grupo fosfato en la posición 5'. Las bases nitrogenadas son estructuras heterocíclicas que se clasifican en bases púricas, Adenina (A) y Guanina (G), y pirimidínicas, Citosina (C) y Timina (T). Esta última se sustituye por Uracilo (U) en el ARN. Los nucleótidos están unidos entre sí mediante enlaces 3',5' fosfodiester, lo que confiere una direccionalidad a la cadena de nucleótidos que por convención se considera en el sentido 5´ $\rightarrow$ 3'.

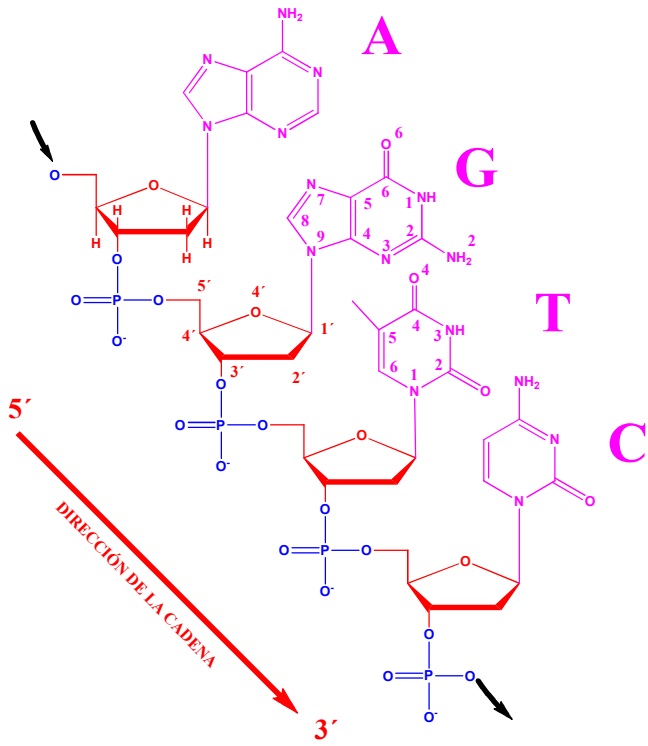


Figura 10. Fragmento de ácido desoxirribonucleico (ADN) de secuencia adenina (A), guanina (G), timina (T) y citosina (C) unidos por enlaces 3´,5´-fosfodiéster.

1.3.3. La Doble Hélice

La interacción entre las distintas cadenas marcará el tipo de estructura (*dúplex, tríplex* o *cuádruplex*). En el desarrollo de esta tesis, únicamente se han estudiado moléculas de ADN de doble cadena. El dúplex se organiza en forma de doble hélice, de manera que el esqueleto hidrofílico azúcar–fosfato de cada hebra se dispone en el exterior mientras que las bases nitrogenadas (hidrofóbicas) se disponen en el interior de la doble hélice (26). En la doble hélice de ADN se diferencian dos surcos: uno ancho y más profundo, denominado surco mayor, y otro estrecho y menos profundo, denominado surco menor (**Figura 11**)

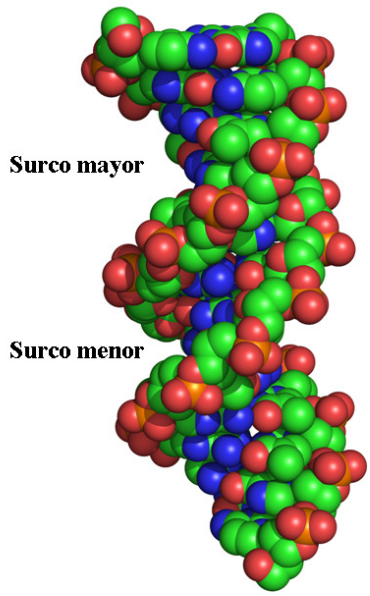


Figura 11. Surco mayor y menor del ADN

Las dos cadenas de polinucleótidos quedan estabilizadas por enlaces de hidrógeno entre bases complementarias y por interacciones de apilamiento o *stacking*. Se generan de esta manera apareamientos canónicos de tipo Watson-Crick A:T y C:G, cumpliendo las leyes de Chargaff (27), que explican la capacidad de la estructura del ADN de permitir la duplicación fiel de la información genética.

1.3.4. Interacción entre Bases

La estabilidad a nivel de la estructura primaria viene determinada por los enlaces covalentes fosfodiéster creados entre nucleótidos. A nivel de la doble hélice la estabilidad viene determinada en gran parte por la interacción entre sus bases, que incluyen enlaces de hidrógeno e interacciones de apilamiento (*stacking*).

1.3.4.1 Enlaces de Hidrógeno

Las bases poseen gran capacidad para formar enlaces de hidrógeno al presentar un gran número de grupos dadores y aceptores de hidrógeno que permiten la formación de distintos apareamientos (**Figura 12**), siendo el de Watson-Crick el mayoritario.

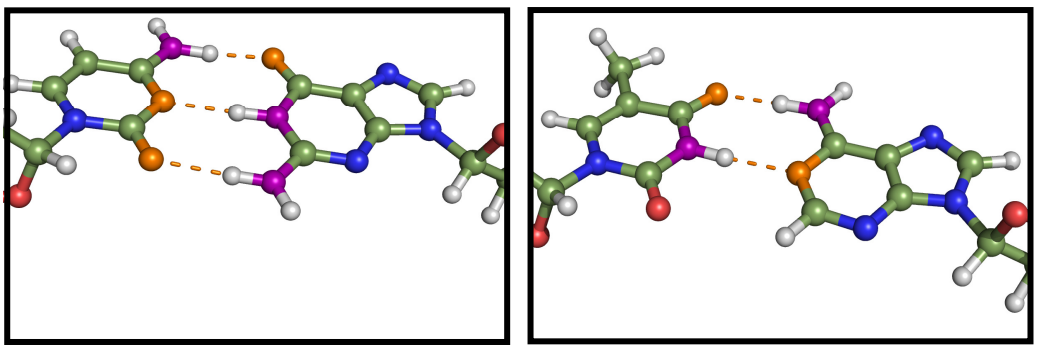


Figura 12. Apareamientos de Watson-Crick por enlaces de hidrógeno. Izquierda: par de bases C:G. Derecha: par de bases T:A. En naranja, grupos aceptores y en violeta, grupos dadores.

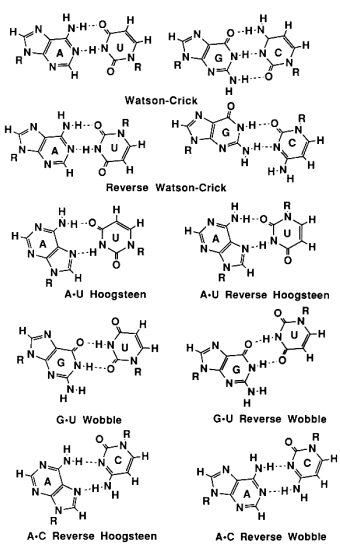


Figura 13. Pares de bases posibles entre bases púricas y pirimidínicas

Los apareamientos son siempre C:G y A:T, en los que intervienen tres y dos enlaces de hidrógeno, respectivamente. La "limitación" de que C sólo se pueda enfrentar a G (y viceversa) y de que A sólo se pueda enfrentar a T (y viceversa) establece al ADN como una estructura capaz de duplicarse con fidelidad, permitiendo la transmisión de la información genética. Además, las bases no saturan su capacidad de donar y aceptar enlaces de hidrógeno una vez apareadas, lo que dota al ADN de un patrón de interacción directamente relacionado con la secuencia. Esto permite explicar, por ejemplo, la alta especificidad en la interacción de algunos fármacos y proteínas con determinadas secuencias de ADN.

Durante el proceso de fusión del ADN a una determinada temperatura, conocida como temperatura de fusión (*melting*), estos enlaces de hidrógeno entre bases complementarias se ven alterados debido a distintos cambios conformacionales en la doble hélice que tendrán como consecuencia final la separación de ambas hebras (28). El *fraying* consiste en una pérdida de la estructura WC en los extremos del *dúplex* por deshilachamiento. El *peeling* se refiere a movimientos donde la distorsión se ve aumentada por enlaces de hidrógeno no canónicos con el esqueleto principal de una cadena. *Searching* es el proceso de búsqueda que da lugar al intercambio de enlaces de hidrógeno tipo WC como consecuencia de un cambio en las pautas de emparejamiento de una cadena con respecto a la otra (**Figura 14**).

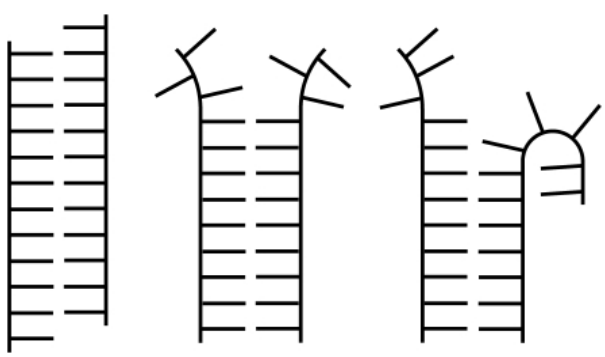


Figura 14. Esquema mostrando los estados conformacionales *searching* (izquierda), *fraying* (centro) y *peeling* (derecha) (28)

1.3.4.2. Interacciones de Apilamiento (*Stacking)*

Las interacciones de apilamiento entre bases son particularmente importantes en medio acuoso ya que la estabilidad de los enlaces de hidrógeno se verá alterada por competición con las moléculas de agua. Los anillos aromáticos de las bases se disponen paralelos unos sobre otros a una distancia de 3.4 Å, evitando contactos electrostáticos no deseados. La razón de la naturaleza del apilamiento es doble; por un lado, es guiado por interacciones hidrófobas, de modo que se minimiza el contacto de las bases con el disolvente polar y, por otro lado, es guiado por interacciones π–π entre las nube de electrones π de los enlaces dobles de los anillos aromáticos de una base con los de la base adyacente. Aunque esta interacciones sean no covalentes, y por lo tanto más débiles que las covalentes, la suma de todas ellas confiere un alto grado de estabilidad, pero también de plasticidad, a la molécula de ADN.

1.3.5 Conformación del ADN

1.3.5.1 Conformación del Esqueleto Azúcar-fosfato

La conformación del esqueleto azúcar-fosfato de una cadena de ADN con la numeración secuencial de átomos P-O5'-C5'-C4´-C3'-O3' etc. se define por los ángulos de torsión α, β, γ, δ, ε, y ζ (**Figura 15**).

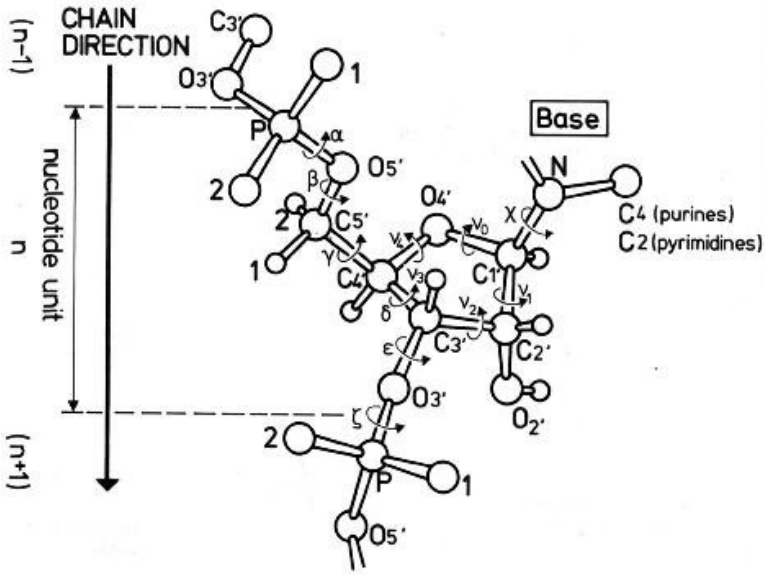


Figura 15. Definición de los ángulos diedros en una cadena nucleotídica y torsionales internos del azúcar. (Saenger,W., *Principles of Nucleic Acid Structure*, Springer Verlag, New York, 1984)

En términos generales, según la IUPAC, las conformaciones correspondientes a ángulos de torsión reciben distintos nombres en función de los valores que adoptan: de -30º a 30º *sinperiplanar* (*syn* o *cys*); de -30º a -90º o +30º a +90º *sinclinal* o *gauche* (g- o g+, en función del sentido de las agujas del reloj); de 90º a 150º y de -90º a -150º *anticlinal*; de +/- 150 a 180º *antiperiplanar* o *trans* (t o *anti*) (**Figura 16**).

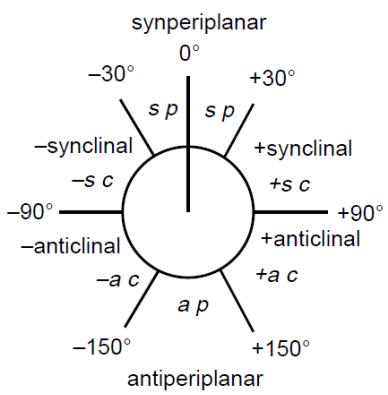


Figura 16. Conformaciones posibles de los ángulos diedros.

La rotación alrededor del enlace exocíclico C4'-C5' permite al O5' adoptar diferentes posiciones relativas a la desoxirribosa. Cabe destacar acoplamientos de distintos torsionales que modifican la conformación del ADN. Los ángulos α y γ están normalmente relacionados, formando un confórmero tipo *gauche-*/*gauche+* (*g-/g+*) en un ADN canónico tipo B. Sin embargo, estos valores cambian al producirse transiciones a formas no canónicas, por ejemplo al formar complejos ADN-proteínas, en cuyo caso pueden adoptar formas del tipo *g+/g-* o *g+/t+*. A su vez, los ángulos de torsión ε y ζ determinan la posición del fosfato respecto al azúcar del nucleótido anterior. Estos torsionales correlacionados pueden adoptar dos conformaciones: BI tipo t/g-, de menor energía, y BII tipo g-/t. En general, el ADN está en su forma BI el 85% del tiempo siendo las transiciones entre ambos estados muy frecuentes. La torsión δ determina en gran medida la conformación del plegamiento del anillo de la ribosa (*puckering*), por lo que se utiliza a menudo para discriminar entre formas A y B de ADN.

1.3.5.2 Conformación del Azúcar

La orientación relativa de la base con respecto al azúcar viene dada por χ (O4'-C1'-N9-C4 en las bases púricas y O4'-C1'-N1-C2 en las bases pirimidínicas) (Figura 3.7) y puede adoptar dos principales orientaciones: *syn* y *anti* (Figura 3.8). La conformación *syn* orienta la base y el azúcar hacia el mismo lado y *anti* indica que la base y el azúcar no se solapan. Ambas dependen mucho del tipo de base y de la conformación del ADN. La conformación más normal en un ADN tipo B es *anti*, siendo la guanina la base que presenta una mayor población de *syn*.

Los ángulos de torsión del azúcar se nombran de ν_0 a ν_4 (**Figura 15**). Debido a que están incluidos en un anillo, estos ángulos de torsión están más restringidos que los anteriores.

La conformación del plegamiento del anillo (*puckering*) más estable para la ribosa es aquella en la que uno o dos átomos quedan fuera del plano formado por el resto de los átomos. Cuando los átomos se desplazan hacia el mismo lado que el C5' se llama *endo* y si se orientan hacia el lado contrario, *exo* (**Figura 17**).

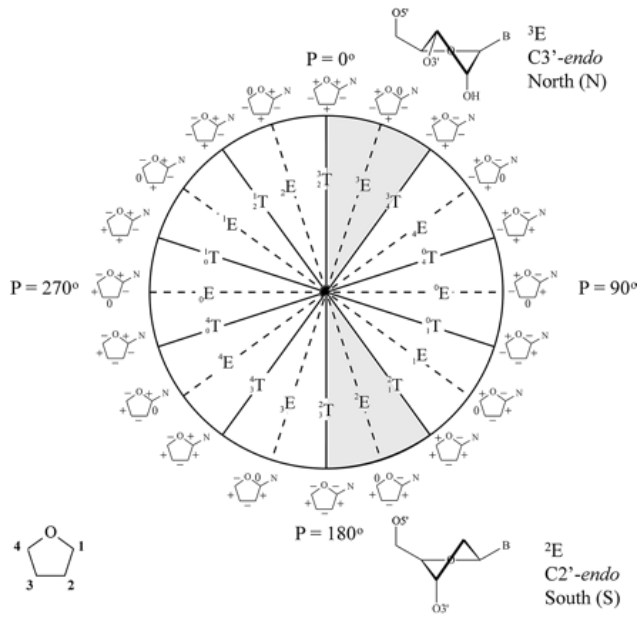


Figura 17. Ciclo de pseudorotación del azúcar y conformaciones preferidas de la ribosa (arriba) y desoxirribosa (abajo). Cada punto en el círculo representa un valor específico de la pseudorotación del ángulo P. Las regiones sombreadas indican las conformaciones que se encuentran en hélices de tipo A y B. En la periferia del ciclo, las ribosas se indican con los signos de los ángulos de torsión endocíclicos (v0 a v4) positivos, negativos y nulo(29).

1.3.5.3. Parámetros de las Bases

Estos parámetros definen la posición relativa de las bases. Cabe distinguir dos tipos, los que describen la posición relativa de una base con respecto a su base complementaria y los que definen la orientación de un par de bases respecto al siguiente par de bases en la cadena (**Figura 18**).

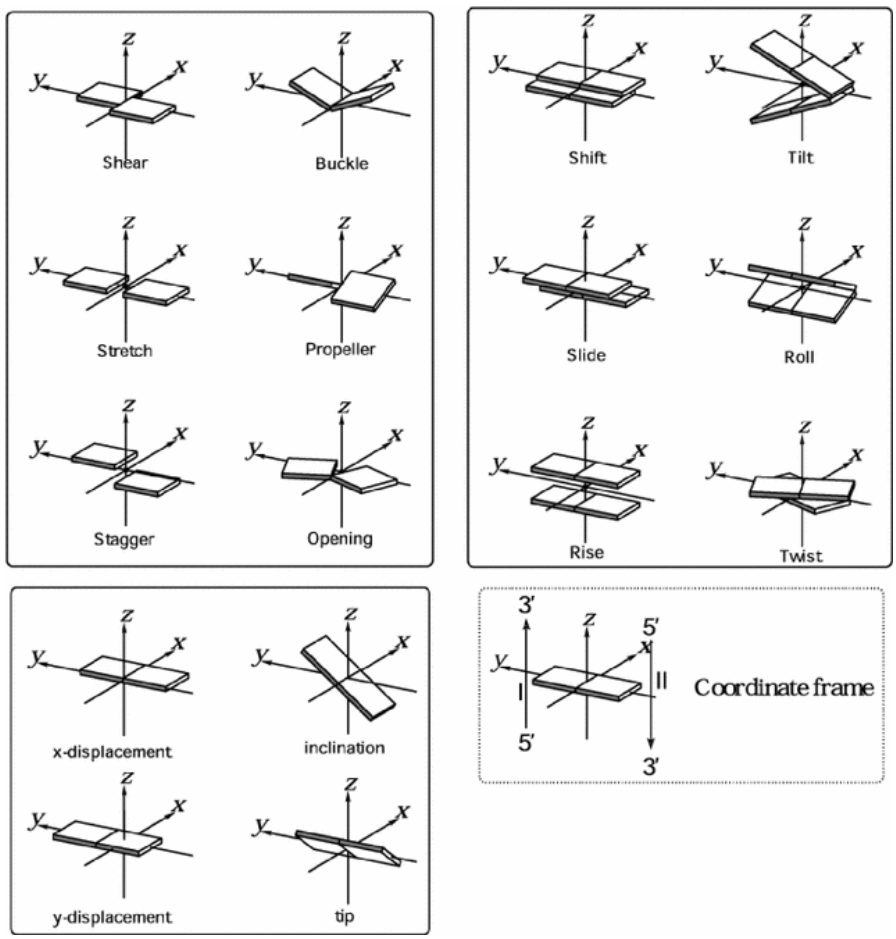


Figura 18. Parámetros de rotación y traslación de las bases del ADN. Los ejes de coordenadas se definen en el recuadro inferior derecho.

1.3.5.4. Familias Conformacionales

La doble hélice de ADN fue descrita por primera vez por Watson y Crick (3). El modelo fue construido con los datos obtenidos a partir de un espectro de difracción de rayos X resuelto por Rosalind Franklin y corresponde al llamado ADN tipo B. Las formas canónicas del ADN son A, B y Z. Las dos primeras son dextrógiras (se enrolla hacia la derecha) y la conformación Z es levógira (enrollamiento hacia la izquierda) (**Figura 19**).

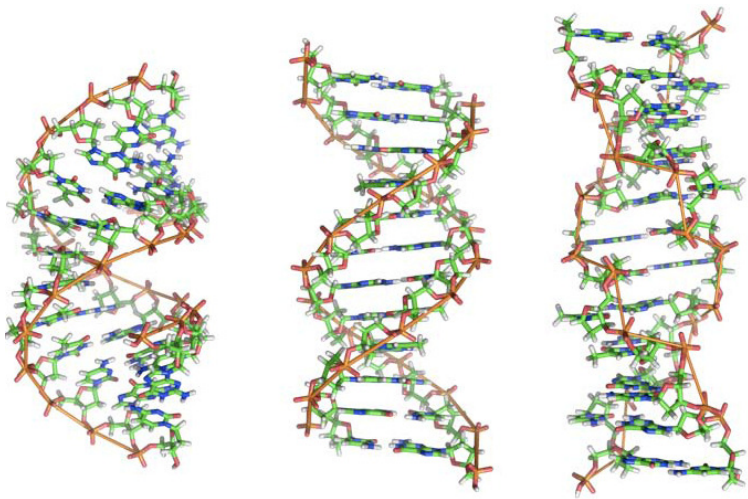


Figura 19. Familias conformacionales del ADN: A–ADN, B–ADN y Z–ADN

1.3.5.5. Estructura Terciaria

El núcleo celular contiene unos 2 metros de ADN compactado en un núcleo de 5 µm de radio, por lo que resulta evidente la necesidad de un sistema altamente eficiente de empaquetamiento. Por ello, el ADN adopta niveles de estructuración mayores que la doble hélice. Estos niveles de compactación no son fijos, ya que el ADN tiene que cambiar de nivel de estructuración dependiendo del momento del ciclo en el que se encuentre la célula. Primero, el ADN–B se va deformando de manera uniforme para producir una estructuración superior en forma de superhélice levógira. En organismos eucariotas esta superhélice se forma gracias al enrollamiento del ADN alrededor de un núcleo formado por unas proteínas llamadas histonas (H2A, H2B, H3 y H4) para dar lugar al nucleosoma (**Figura 20**).

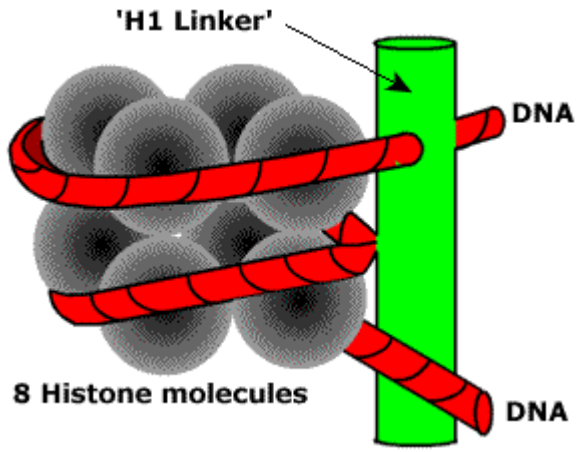


Figura 20. Nucleosoma

Estas proteínas son muy ricas en hélices α y poseen un alto contenido en aminoácidos cargados positivamente, lo que favorece la interacción electrostática con el ADN. La formación del nucleosoma permite un nivel de compresión 10 veces superior al inicial. Cada nucleosoma queda separado por una secuencia de ADN de hasta 80 pares de bases, formando la estructura que se conoce como "collar de perlas" o, más correctamente, una fibra de cromatina, siendo ésta la estructura propia del núcleo interfásico, que no ha entrado en división. Este collar de nucleosomas vuelve a enrollarse y en este nuevo empaquetamiento participan otras histonas (H1), que se unen a cada nucleosoma provocando un cambio en la dirección de salida del ADN, lo cual permite formar una estructura en zig–zag. Este tipo de estructuras se combinan para conseguir el siguiente nivel de compactación, que sigue sin estar bien caracterizado. El modelo más sencillo formulado es el de solenoide, donde se forman repeticiones con 5 nucleosomas por paso. Este nivel de compactación recibe el nombre de fibras de cromatina, las cuales tienen unos 30 nm de diámetro. Éste es el mayor grado de empaquetamiento en el núcleo durante la interfase. El nivel de compactación más alto ocurre durante la metafase, cuando se define el cromosoma. No se conoce tampoco con suficiente detalle qué proteínas, ARN u otras macromoléculas están involucradas en este último proceso**.** El resultado global es una longitud 50.000 veces inferior a la longitud inicial del ADN desenrollado **(Figura 21).**

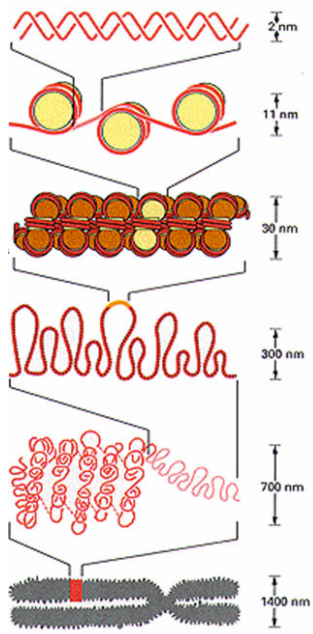


Figura 21. Niveles de compactación del ADN.

1.3.5.6. Interacción del ADN con fármacos

En cada una de las formas de ADN que hemos descrito, la superficie interna de los surcos está formada por los grupos funcionales característicos de cada base nitrogenada, los cuales ofrecen posibilidades de interacción mediante enlaces de hidrógeno con proteínas y ligandos **(Figura 22).**

Figura 22. Aceptores y donadores de enlace de hidrógeno entre dos pares de bases

La diversidad estructural entre los distintos fármacos que se pueden unir al ADN refleja la gran flexibilidad de esta macromolécula como receptor. Aunque el surco mayor tiene un potencial de reconocimiento superior al del surco menor, la mayoría de los fármacos son específicos de este último. Por el contrario, el surco mayor del ADN parece más apropiado para la unión de proteínas reguladoras de la expresión de genes donde numerosos enlaces de hidrógeno específicos con las bases nitrogenadas sientan las bases para el reconocimiento secuencias. No hay que olvidar tampoco la influencia que la estructura del ADN tiene en esta lectura, donde variaciones en parámetros del propio ADN como, por ejemplo, la anchura de los surcos, puede permitir o impedir la unión de ligandos y proteínas. La unión de fármacos al ADN puede ser covalente o no covalente, y puede tener lugar por intercalación o mediante unión a uno de los surcos.

La intercalación conlleva un efecto muy dramático en la conformación del ADN, ya que estos agentes, generalmente de geometría plana y heterociclos, se introducen entre dos pares de bases adyacentes. Con ello, la longitud total de la cadena se incrementa en unos 3,4 Å. Este incremento en la longitud se consigue disminuyendo el grado de enrollamiento (*twist*) de la cadena. La contribución a la estabilidad del complejo viene determinada fundamentalmente por las interacciones de apilamiento.

La unión al surco menor consiste en el posicionamiento de la molécula en su interior, por lo que se desplazan así, total o parcialmente, las moléculas de agua que pudieran encontrarse hidratando esa zona del ADN. Los efectos que sufre la doble hélice de ADN no son tan dramáticos como los que se producen por intercalación, pero también se originan cambios como, por ejemplo, el ensanchamiento del surco. Esta unión puede establecerse tanto en zonas ricas en AT (e.g. netropsina, Hoechst–33258, DAPI), como a secuencias CG (e.g. cromomicina A_3, mitomicina C). La estabilidad de los compuestos viene dada por interacciones de enlaces de hidrógeno entre los grupos funcionales del surco y los del fármaco.

1.4. Propiedades Estructurales de las Proteínas

1.4.1. Introducción

Las proteínas, en su conjunto, son las biomoléculas más importantes desde un punto de vista funcional. Sus propiedades son tan útiles que se emplean como agentes terapéuticos y catalizadores, así como fuente de materiales diversos. Muchas enfermedades tienen su origen en mutaciones que se producen en las proteínas, provocando la pérdida de su función al bloquear su actividad enzimática o al alterar su estructura tridimensional. En general, las enzimas y los receptores de membrana constituyen las dianas más comunes de los medicamentos, ya sea para restaurar o modificar la función celular o para destruir agentes infecciosos y células cancerígenas.

Entre los objetivos del estudio de las proteínas se incluyen el poder predecir la estructura y la función de una molécula concreta a partir de su secuencia de aminoácidos, así como el conocer los detalles de su interacción con pequeños ligandos u otras macromoléculas. Este conocimiento ya está permitiendo, en casos favorables, diseñar y sintetizar catalizadores y materiales de uso industrial, entender las bases moleculares de muchas enfermedades, así como el diseño de fármacos a medida para combatir ciertas enfermedades. En los últimos años se han conseguido avances significativos en este sentido: grupos experimentales han llegado a alterar la actividad y la estabilidad de algunas proteínas mientras que grupos teóricos han sido capaces de simular aspectos del plegamiento y mecanismo catalítico de algunas proteínas, así como de caracterizar cambios estructurales importantes para su función.

La función biológica de una proteína depende en gran medida de la conformación o conformaciones que es capaz de adoptar. En una proteína cada aminoácido desempeña una función específica, de modo que unos se encuentran en el interior de la molécula plegada y son esenciales para el mantenimiento de la estructura, mientras que otros participan en el mecanismo de una reacción enzimática o en la estabilización de ligandos unidos o en la estabilización de su unión

a otras proteínas. Un principio básico de la bioquímica es que toda función fisiológica tiene una base estructural: sin estructura no hay función. Como regla general, hay una estructura para cada función y una función para cada estructura. Así, cada proteína ha evolucionado para llevar a cabo una función específica. La amplia variedad de funciones biológicas requiere, por tanto, de una diversidad estructural igualmente amplia. En este sentido, las mutaciones aleatorias y la selección natural han sido los grandes impulsores de la evolución que han sufrido las proteínas en el transcurso del tiempo, independientemente de si la aparición espontánea de una nueva estructura abre la posibilidad de una nueva función o de si la necesidad de una nueva función obliga a las proteínas a cambiar hasta que se genera una estructura capaz de ejercer dicha función.

1.4.2. Principios Básicos de Estructura de Proteínas

Péptidos y proteínas son polímeros formados por secuencias de aminoácidos. En ellas vamos a encontrar 21 aminoácidos distintos responsables de la versatilidad estructural y funcional que las caracteriza **(Figura 23).**

Middle Base / 5' Base	U	C	A	*G	Middle Base / 3' Base
*U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	*Stop*	*Sec / *Stop*	*A
	Leu	Ser	*Stop*	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	▲Met *Initiator*	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Figura 23. Los 20 aminoácidos de las proteínas (21 al incluir la selenocisteína) (Extraído de *JNCI J Natl Cancer Inst.).*

Todos los aminoácidos están formados por un grupo amino, de carácter básico, un grupo carboxilo de carácter ácido y una cadena lateral, todos ellos unidos al mismo átomo de carbono (Cα). Las primeras estructuras de rayos X revelaron que las proteínas no formaban estructuras regulares o simétricas sino que eran mucho más complejas. Sin embargo, ciertos motivos estructurales solían aparecer frecuentemente, como la hélice α y la hebra β, que puede dar lugar a láminas paralelas y antiparalelas **(Figura 24)**.

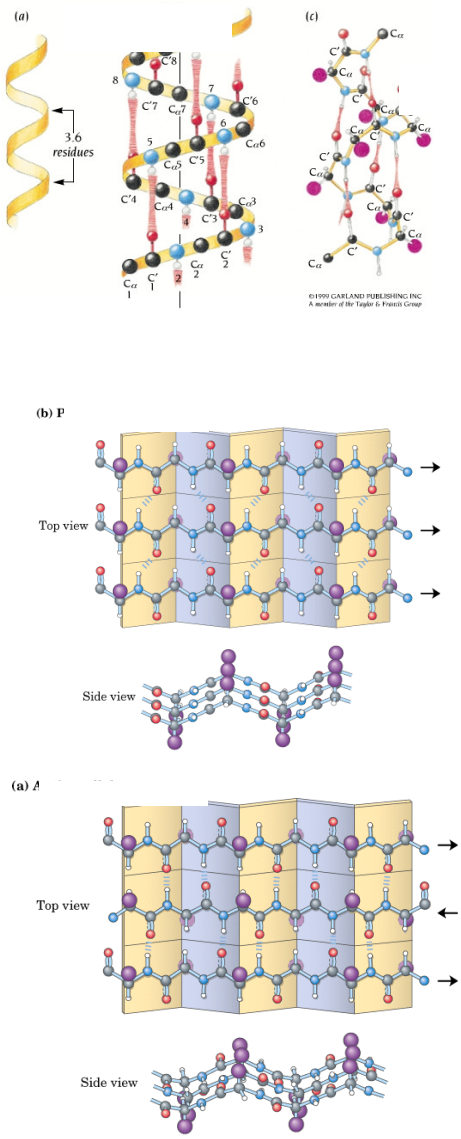


Figura 24. Geometrías de a) hélice alfa, b) lámina beta paralela, y c) lámina beta antiparalela (Lehninger, *Principles of Biochemistry, 3rd edition*)

Estos motivos constituyen la estructura secundaria de la proteína, siendo la estructura primaria el orden de los aminoácidos en la secuencia, la terciaria, el plegamiento de los elementos anteriores en una forma más o menos compacta, y la cuaternaria, la asociación de subunidades proteicas en un ensamblado molecular de mayor tamaño y complejidad. Los elementos de estructura secundaria se conectan entre sí por regiones conocidas como asas o giros (*loops*), que constituyen

conformaciones menos regulares. Cada tipo de estructura secundaria puede ser completamente descrito por los ángulos de torsión de cada residuo (**Figura 25).**

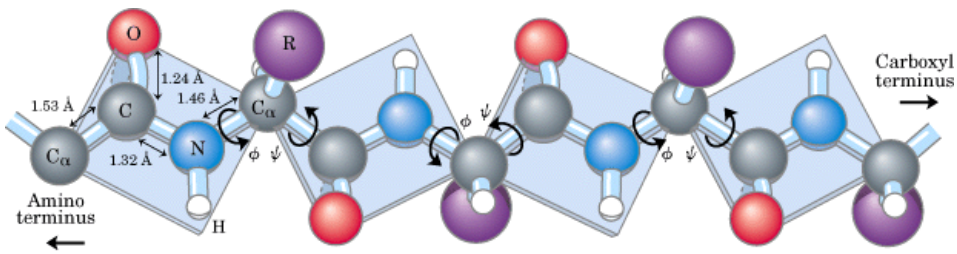


Figura 25. Geometría del enlace peptídico

Dado que el enlace peptídico (ω) es plano y raramente se desvía de 0º ó 180º, serán los ángulos ϕ y ψ los que determinarán la estructura secundaria, como puede verse en el mapa de Ramachandran (**Figura 26**).

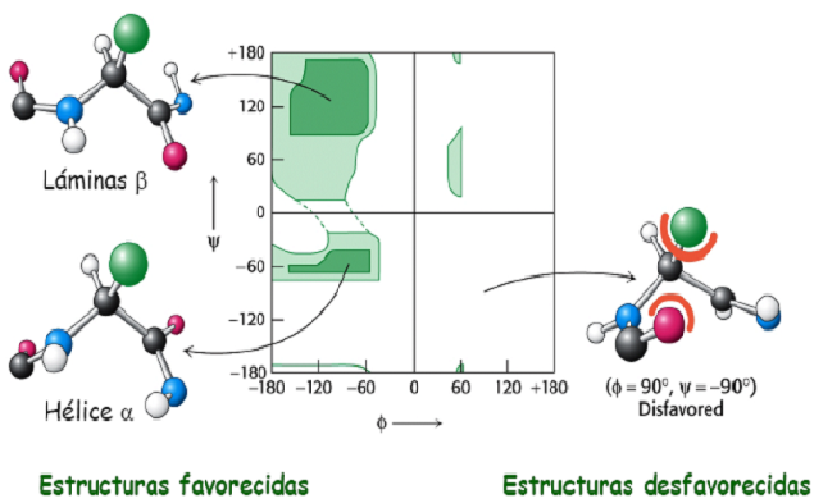


Figura 26. Mapa de Ramachandran

A medida que se fueron conociendo más estructuras de proteínas, se observó que muchas contenían más de una región diferente, a menudo con funciones separadas. A cada una de estas regiones se la conoce como dominio, que se define como una cadena polipeptídica que puede plegarse independientemente en una estructura tridimensional estable.

1.4.3 Fuerzas de Interacción

En los complejos proteicos que se localizan en las membranas celulares predominan las interacciones hidrófobas pero existen otras asociaciones estables en las que el mantenimiento de la estructura depende mayoritariamente de interacciones de otra naturaleza, como los enlaces de hidrógeno o las interacciones de van der Waals, como ocurre en las proteínas multiméricas. En general, todos estos tipos de enlaces son relativamente débiles de modo individual, pero pueden proporcionar una gran estabilidad a los complejos moleculares cuando se dan simultáneamente en número suficiente. El estudio de otro tipo de interacciones, las transitorias, que ocurren entre proteínas o entre una enzima y su sustrato, son mucho más difíciles de abordar que las interacciones que conducen a complejos estables, ya que implican no sólo la formación del complejo inestable propiamente dicho, sino que también deben regular los procesos internos de adaptación y acoplamiento que permitan que la reacción tenga lugar.

El proceso de reconocimiento molecular, sobre todo a corta distancia, está controlado en gran medida por las características de la superficie de las proteínas, que vienen determinadas por la diferente naturaleza de las cadenas laterales de los aminoácidos. Con frecuencia, los residuos de igual naturaleza se concentran en determinadas regiones, haciendo que las proteínas presenten una distribución irregular con áreas de marcado carácter hidrófobo y otras de naturaleza polar. Las características superficiales de las proteínas van a ser decisivas, por tanto, en la formación de complejos transitorios funcionalmente activos. De hecho, la distribución superficial de carga de las proteínas desempeña un papel tan importante como la carga global en el proceso de asociación y reconocimiento entre dos

proteínas. Así, las interacciones electrostáticas de larga distancia facilitan el acercamiento de las proteínas: cuando una proteína posee una carga global positiva y la otra una carga global negativa se produce el movimiento de atracción entre ambas. A corta distancia, sin embargo, son las cargas locales y los parches hidrófobos los que determinan la asociación y la consiguiente formación del complejo. Las interacciones electrostáticas a larga y corta distancia serían, simplificando mucho, del tipo monopolo–monopolo y dipolo–dipolo, respectivamente.

1.4.4. Movimientos de Proteínas

Hasta hace poco se decía que las proteínas con capacidad catalítica, denominadas enzimas y encargadas de acelerar las reacciones metabólicas, reconocían a los ligandos sobre los que debía actuar con enorme especificidad, al igual que una llave encaja en su cerradura de manera única. Esta manera de visualizar la relación entre las enzimas y sus sustratos tiene gran valor pedagógico a la hora de explicar cómo una proteína es capaz de reconocer e interaccionar de modo específico con una molécula concreta, pero da a entender que las proteínas son estructuras rígidas e indeformables. Hoy se sabe que las proteínas son moléculas de gran plasticidad, sobre todo en las zonas de reconocimiento con otras proteínas y en sus centros activos, mientras que las zonas internas, responsables de mantener su estructura y conformación tridimensional, presentan cierta rigidez. En efecto, las proteínas son moléculas en las que la flexibilidad estructural constituye una de las claves que explican su funcionalidad. Las proteínas pueden modificar su conformación tras unirse al sustrato o por reacciones enzimáticas específicas. La Naturaleza aprovecha los cambios conformacionales así inducidos para controlar la actividad de las proteínas reguladoras y convertir el trabajo mecánico en energía físico–química. Así ocurre en las proteínas alostéricas con varios sitios de unión al sustrato, en las que la unión de una primera molécula del sustrato provoca cambios estructurales que facilitan la entrada de las siguientes. Entre los ejemplos más representativos y mejor estudiados se encuentra la hemoglobina, proteína

tetramérica en la que la unión de la primera molécula de oxígeno a una de las subunidades favorece la entrada de las otras tres moléculas.

Los cambios estructurales asociados a la función de las proteínas pueden ser relativamente sutiles, como los que explican el alosterismo de la hemoglobina, pero en ocasiones conllevan movimientos drásticos y pronunciados de, al menos, ciertos dominios de la estructura (30). Hoy en día los cambios conformacionales de muchas proteínas se conocen lo suficientemente bien como para explicar su funcionamiento y modo de acción, pero la diversidad es tan amplia que resulta difícil analizarlos de manera homogénea y sacar conclusiones o "leyes generales" (31).

1.4.5. Predicción de la Estructura de Proteínas Mediante Modelado por Homología

La predicción de la estructura de proteínas mediante modelado por homología se aprovecha de la relación evolutiva entre proteínas con un ancestro común, lo que las hace compartir características estructurales muy similares. Esta relación fue cuantificada por primera vez por Chothia y Lesk en 1986 (32-34), y el concepto fue ampliado posteriormente por Sander y Schneider en 1991 (35). En general, para pares de proteínas de más de 50 aminoácidos hay un límite en torno al 30% de identidad de secuencia de aminoácidos a partir del cual ambas proteínas serían estructuralmente semejantes (o estructuralmente homólogas). En la práctica, este hecho permite una aproximación al plegamiento de la cadena polipeptídica de una proteína tomando directamente como molde la estructura de otra cuyo porcentaje de identidad en el alineamiento de aminoácidos sea mayor del 30%.

Construcción del modelo.

En esencia, el modelado por homología consiste en la extrapolación de la estructura para una nueva secuencia aminoacídica (modelo) a partir de una estructura tridimensional conocida de uno o más miembros (moldes) de la misma

familia de proteínas (homólogos). Los pasos, en lo que se refiere a la construcción del modelo, se resumen a continuación:

(i) Identificación del molde adecuado para la secuencia problema. El sistema habitualmente más utilizado consiste en buscar la mejor estructura (o estructuras) candidata(s) a molde, haciendo un alineamiento de la secuencia problema frente a una base de datos de secuencias de proteínas con estructura tridimensional conocida mediante algoritmos como BLAST (*Basic Local Alignment Search Tool*) (36). Una buena estructura candidata a molde debe mostrar, al menos, un 30% de homología de secuencia con la proteína problema.

(ii) Alineamiento entre la secuencia problema y las del molde o moldes. Se puede obtener un buen alineamiento utilizando programas que consideran simultánemante múltiples secuencias, como CLUSTALW (37) o T–COFFEE (38). Pueden descartarse los segmentos no alineados que corresponden a regiones no conservadas y que se modelarán de forma independiente.

(iii) Construcción de regiones conservadas. La construcción del modelo propiamente dicho comienza con la colocación espacial de los átomos comunes a ambas cadenas: secuencia problema y molde. Tales átomos corresponderán a los aminoácidos totalmente conservados y a la traza de Cα (sin las cadenas laterales) de residuos que, si bien no son idénticos, mantienen un alto grado de homología.

(iv) Construcción de regiones no conservadas. La falta de información estructural en estas zonas, al carecer de segmentos homólogos en el molde, obliga a que sean construidas aparte. Una de las metodologías empleadas consiste en predecir primero la estructura secundaria de esta secuencia de aminoácidos no conservada, con un servidor web, como JPred (39) y la posterior búsqueda de fragmentos con la estructura secundaria predicha (servidor web PROSITE, http://expasy.org/prosite/), para que puedan ser acomodados en el modelo. En el caso de que la estructura secundaria de estas regiones no conservadas fuera un asa o giro (*loop*), el problema se hace más complicado, dada la gran variedad de conformaciones que puede adoptar. Podemos recurrir en estos casos a predicciones conformacionales *ab initio*, como las que hace el servidor web ModLoop (40).

(v) Adición de las cadenas laterales en las posiciones no conservadas. Una vez que se posee la estructura completa del esqueleto peptídico, es necesario acomodar la posición de los átomos correspondientes a las cadenas laterales de los aminoácidos en las regiones recién construidas no conservadas, así como la de los aminoácidos no totalmente conservados entre la cadena molde y la secuencia problema. Para ello se dispone, para la cadena lateral de cada aminoácido, del conjunto de las conformaciones más probables (rotámeros), cada uno de los cuales se evalúa en función de su posibilidad de colisión con los aminoácidos de alrededor, y se elige el más favorable. Los servidores como What If (http://swift.cmbi.ru.nl/servers/html/index.html) y MaxSprout (http://www.ebi.ac.uk/Tools/maxsprout/index.html) hacen este proceso de manera automática, pero podemos hacerlo de manera más "artesanal" e interactiva con la colección de rotámeros que ofrece el programa de visualización molecular PyMOL (http://pymol.org/).

(vi) Refinamiento del modelo. Por último, con métodos de minimización de energía como los descritos previamente podremos mejorar la geometría de los términos enlazados (distancia, ángulo y torsión) y disminuir el número de colisiones presentes en el modelo.

(vii) Evaluación del modelo. Servidores como What If (41) y ProCheck (42) utilizan el mapa de Ramachandran para evaluar el modelo obtenido. En general, valores alejados de los óptimos durante el proceso de evaluación de un modelo son casi siempre debidos a falta de optimización en el alineamiento de partida, siendo éste el paso limitante para la obtención de modelos fiables de estructuras de proteínas.

Existen diversos programas y servidores que permiten la obtención de modelos de proteínas por homología de forma automática o casi automática. Destacan, por su aceptación casi general, Swiss–Model (43), MODELLER (44,45), Phyre (46) e I-TASSER (47).

1.4.6. Predicción del Modo de Unión de Pequeños Ligandos a sus Dianas: *Docking*

Predecir cómo se unirá un ligando a una proteína o a una secuencia de ácido nucleico de interés no es en ningún caso una tarea trivial ni libre de subjetividad. El problema se puede abordar de diferentes maneras según la información de que dispongamos: si existe una estructura cristalográfica de la macromolécula formando un complejo con uno o más ligandos, se puede utilizar la información de la posición relativa del ligando en el bolsillo de unión de la proteína o ácido nucleico para predecir el modo de unión de otros ligandos semejantes.

El acoplamiento (*docking*) también puede ser manual, porque puede que no tengamos datos a priori del modo de unión. En este caso, debemos contar con algo de información para que el modelo sea razonablemente fiable como, por ejemplo, el tipo de inhibición del ligando, o los residuos clave del receptor para su interacción. Además, en estos casos resulta muy útil la utilización de métodos como GRID (48), que predice los sitios de unión energéticamente favorables de diferentes grupos funcionales en macromoléculas de importancia biológica. Estos grupos funcionales se representan por sondas y, seleccionando unos niveles energéticos adecuados, se pueden visualizar gráficamente contornos tridimensionales que resaltan las zonas donde esas sondas o grupos funcionales dan lugar a una interacción favorable con la macromolécula.

Un tercer método menos sesgado consiste en explorar de forma automática y exhaustiva los posibles modos de unión de un ligando a su receptor mediante algoritmos que examinan el bolsillo de unión, encajan el ligando en él y evalúan las interacciones en cada posición estudiada. Para este último paso se requieren funciones de evaluación o tanteo (*scoring*) que cuantifiquen la interacciones que caracterizan el complejo, generalmente mediante su descomposición en contribuciones electrostáticas y de van der Waals, así como la energía de desolvatación de la superficie de contacto entre el ligando y el receptor.

$$\Delta G_{unión} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{enlace} + \Delta G_{desolv} + \Delta G_{tors} \qquad (1.18)$$

Los programas de *docking*, por lo tanto, realizan dos pasos consecutivos: exploración y evaluación. Como resultado de la exploración se generan un gran número de posibles modos de unión ligando–receptor ("poses"), que son evaluadas por la función de tanteo para identificar las soluciones más favorables, desde un punto de vista energético, al considerarse que tienen una mayor probabilidad de corresponder a la solución experimental. La estrategia utilizada por este tipo de programas se detalla a continuación:

(i) Caracterización del sitio de unión. Uno de los programas más utilizados para estudiar el centro activo de una enzima o el bolsillo de unión a ligandos de una proteína es CASTp (49), que emplea un algoritmo para localizar estas cavidades, y calcular tanto su área como su volumen.

(ii) Precálculo de las energías de interacción. Una vez delimitado el bolsillo del receptor donde vamos a proceder a realizar la búsqueda de poses para el ligando, definimos en su interior una malla tridimensional y precalculamos en cada nudo de la malla potenciales de afinidad, tanto electrostáticos como de van der Waals, para cada tipo de átomo del ligando. De esta manera, aceleramos el proceso de evaluación posterior porque, mediante interpolación, podremos conocer la interacción entre cada uno de los átomos del ligando y los átomos del receptor.

(iii) Generación de poses dentro del centro activo. A la hora de buscar las posibles conformaciones y orientaciones del ligando hay que considerar todos sus grados de libertad posibles tanto los rotacionales y traslacionales como su flexibilidad (ángulos torsionales activos). El receptor, por el contrario, normalmente se considera rígido aunque ya existen programas, como por ejemplo GLIDE (50-52) y AutoDock 4.0 (http://autodock.scripps.edu/), que incluyen en el cálculo del acoplamiento ligando–receptor una cierta flexibilidad al considerar diferentes conformaciones de las cadenas laterales de los aminoácidos que constituyen el centro activo.

(iv) Evaluación energética. El propósito de una función de evaluación o tanteo es intentar discriminar, en un tiempo de cálculo razonable, entre las soluciones correctas y el resto de las soluciones encontradas en el proceso de búsqueda. Con este objetivo se hace una estimación de la energía de unión ligando–receptor en cada complejo mediante campos de fuerzas, potenciales estadísticos u otras funciones.

1.5. Aplicaciones del Modelado Molecular para Comprender el Mecanismo de Acción de Fármacos Antitumorales.

En el desarrollo de esta tesis, se ha investigado el mecanismo de acción de la mitomicina C, un antitumoral clásico, y de otros fármacos antitumorales novedosos de origen marino desarrollados por la empresa española PharmaMar, que tiene como uno de sus principales objetivos explorar el universo marino en busca de nuevos tratamientos contra el cáncer.

Durante millones de años, innumerables formas de vida marina han evolucionado para producir una inmensa variedad de entidades químicas de una sofisticación exquisita. Entre éstas se encuentran moléculas con actividades biológicas potentes, desarrolladas supuestamente como una forma de defensa bioquímica para sobrevivir en un entorno extremadamente competitivo. La evolución ha definido y seleccionado diversas estrategias de supervivencia, defensa, ataque, adaptación y comunicación entre organismos marinos a través de un verdadero arsenal de moléculas (metabolitos), algunas de las cuales podrían llegar a emplearse en el tratamiento de enfermedades humanas (53).

A partir de las muestras de organismos marinos, PharmaMar lleva a cabo la identificación y aislamiento de moléculas de interés y la rápida elucidación de sus estructuras químicas. Las moléculas con actividad antitumoral innovadora y no descritas hasta el momento son sintetizadas en el laboratorio mediante un proceso químico para evitar recurrir a la fuente natural, normalmente escasa y a veces protegida, y para obtener cantidades suficientes para su desarrollo y posible futura comercialización.

Se descubren continuamente nuevas moléculas dotadas de exclusivos mecanismos de acción (MoA), lo que representa una gran oportunidad para avanzar en las ciencias básicas y, lo que es más importante, abrir nuevas vías potenciales para el tratamiento del cáncer **(Figura 27**). PharmaMar tiene colaboraciones con más de 200 centros de I+D en todo el mundo, incluyendo centros de investigación de prestigio y universidades.

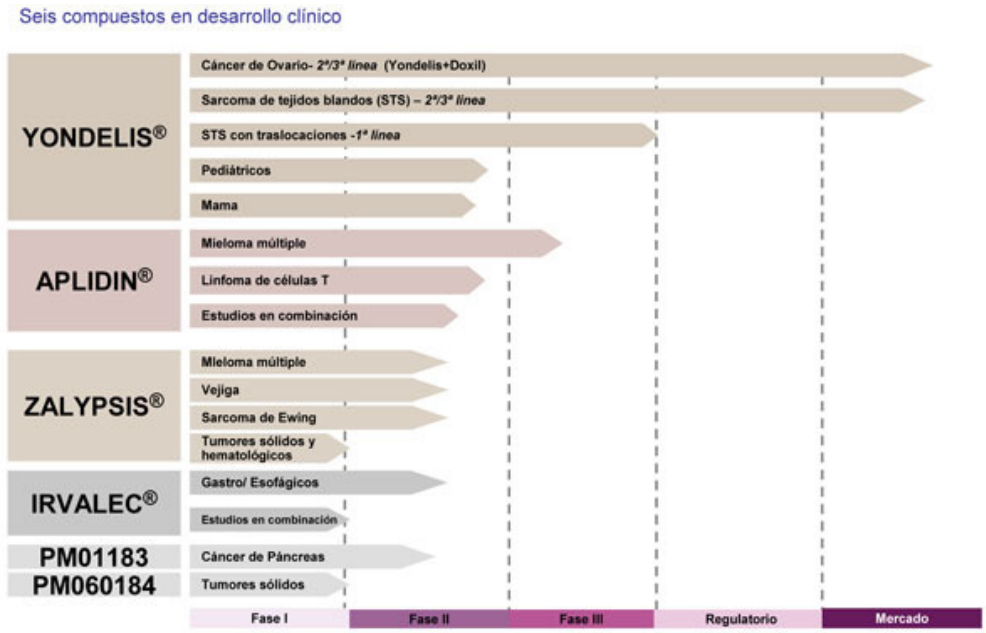


Figura 27. Productos de PharmaMar en desarrollo clínico

En las investigaciones llevadas a cabo en el laboratorio se estudiaron los mecanismos de acción de *Yondelis®, Zalypsis®* y PM01183 mediante modelado molecular. La colaboración con grupos de investigación experimentales nos permite conocer el mecanismo de acción de estas moléculas y justificar la especificidad de su unión al ADN con un nivel de detalle atomístico.

1. Yondelis®

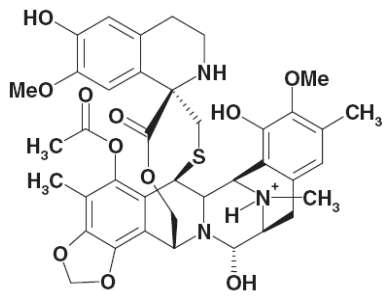


Figura 28. Estructura química de Yondelis®

Yondelis® (trabectedina) **(Figura 28)** es un agente antitumoral de origen marino descubierto en el tunicado colonial *Ecteinascidia turbinata* **(Figura 30).** Actualmente se produce por síntesis química a partir de la cianosafracina B. Yondelis® tiene un mecanismo de acción único: se une al surco menor del ADN e interfiere con la división celular y los procesos de transcripción genética y reparación del ADN (54). Se ha observado que dosis clínicamente relevantes de trabectedina inducen dobles roturas en el ADN (*double strand breaks,* DSBs) de células humanas, activando la reparación por recombinación homóloga de un modo similar al producido por la mitomicina C (MMC), un agente antitumoral clásico que produce entrecruzamientos intercatenarios (*interstrand cross-link,* ICL*)* en el ADN (55-57) **(Figura 29).**

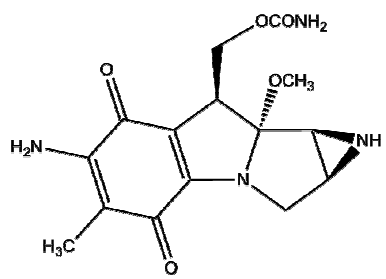


Figura 29. Estructura química de mitomicina C

En septiembre de 2007, Yondelis® recibió la autorización de comercialización por la Comisión Europea para el tratamiento del sarcoma de tejidos blandos avanzado o metastático. En noviembre de 2009 Yondelis® recibió su segunda aprobación de comercialización por parte de la Comisión Europea para el tratamiento del cáncer de ovario recurrente platino-sensible en combinación con DOXIL® / Caelyx®.

Figura 30. ***Ecteinascidia turbinata***

Yondelis® ha sido designado fármaco huérfano por la Comisión Europea (CE) y la Administración de Fármacos y Alimentos(FDA) de los EE.UU. para el sarcoma de tejidos blandos y el cáncer de ovario. PharmaMar comercializa Yondelis® en Europa y en Japón, y Ortho Biotech Products, L.P. lo hace en Estados Unidos y el resto del mundo. Actualmente se están llevando a cabo ensayos de fase II con Yondelis® para cáncer de mama, de próstata, de pulmón y para tumores pediátricos.

2. Zalypsis®

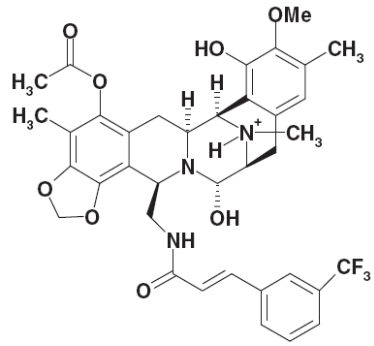


Figura 31. Estructura química de Zalypsis®

Zalypsis® es una entidad química novedosa que está relacionada con el compuesto natural marino jorumicina y con la familia de las renieramicinas que se derivan de moluscos y esponjas, respectivamente (58) **(Figura 31).**

Figura 31. *Jorunna funebris*, nudibranquio del que se aisló la jorumicina.

Está actualmente en fase II de desarrollo clínico en cáncer de endometrio y de cérvix y, de forma simultánea, se encuentra en evaluación farmacológica primaria en líneas celulares y modelos animales frente a distintos tipos de tumores.

3. Triptamicidina (PM01183)

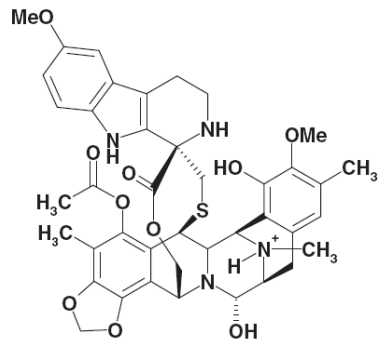


Figura 32. Estructura química de PM01183

PM01183 **(Figura 32)** es un nuevo compuesto sintético de origen marino que también se une covalentemente al surco menor del ADN. Estos aductos PM01183-ADN dan lugar a DSBs y a alteraciones del ciclo celular, induciendo la muerte de las células cancerosas (59). En estudios preclínicos, el compuesto ha mostrado una potente actividad citotóxica contra líneas celulares tumorales de diferentes orígenes. Actualmente se están llevando a cabo ensayos clínicos de fase I con PM01183 para tumores sólidos.

1.6. Vías de Reparación de Entrecruzamientos Intercatenarios en el ADN

Muchos fármacos anticancerosos empleados en la clínica han sido utilizados durante décadas y son altamente eficaces frente a células tumorales. Un daño importante en el ADN provoca una parada del ciclo celular y muerte celular. Además, los daños en el ADN que persisten en la fase S del ciclo pueden bloquear la progresión de la horquilla de replicación provocando la formación de DSBs en el ADN asociadas a la replicación (60). Las DSBs son consideradas como las lesiones más tóxicas para el ADN.

Los daños más importantes que se producen en el ADN y que interfieren en la progresión de las horquillas de replicación son modificaciones químicas (aductos)

de las bases del ADN. Éstas se crean como consecuencia de la unión covalente de fármacos reactivos directamente o después de su metabolización en el organismo. Estos agentes se clasifican en dos categorías: los monofuncionales (por ejemplo, Yondelis®, Zalypsis® y PM01183), que presentan una única región reactiva que modifica una única base, y los bifuncionales (por ejemplo, mitomicina C), que poseen dos sitios reactivos y pueden entrecruzar dos bases de ADN de la misma hebra (entrecruzamientos intracatenarios) o, más frecuentemente, de hebras opuestas (ICL) **(Figura 33)**. Estos últimos bloquean muy eficientemente las horquillas de replicación y transcripción.

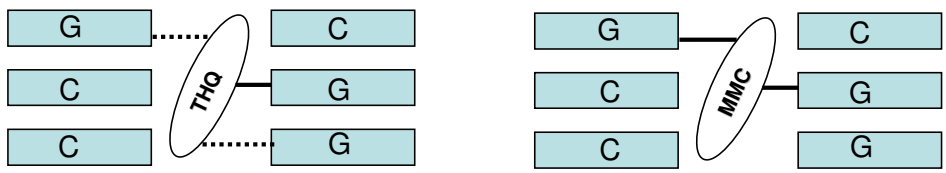


Figura 33. Esquema del modo de unión de agentes alquilantes monofuncionales (izquierda, derivados tetrahidroisoquinolínicos) y agentes que inducen la formación de ICLs (derecha, mitomicina C). Las líneas continuas corresponden a enlaces covalentes y las líneas discontinuas a enlaces de hidrógeno.

Es especialmente relevante el hecho de que fármacos como Yondelis® y sus derivados tetrahidroisoquinolínicos, además de unirse covalentemente al amino exocíclico de una guanina determinada, también son capaces de establecer interacciones no covalentes con el esqueleto azúcar-fosfato y otras bases tanto de la hebra modificada como de la complementaria **(Figura 33)**. En consecuencia, se ha sugerido que estos fármacos se comportan funcionalmente como entrecruzadores intercatenarios (54).

La reparación de ICLs implica múltiples vías de reparación de ADN entre las que destacan la reparación por escisión de nucleótidos (NER) y la recombinación homóloga (HR). Esta última implica la formación de DSBs (61) **(Figura 34)**. El hecho de que el tratamiento de células con Yondelis® (54) o Zalypsis® (62) induzca la formación de DSBs refuerza la idea de que estos fármacos se comportan funcionalmente como

entrecruzadores intercatenarios, a pesar de que se unan covalentemente a una sola hebra del ADN.

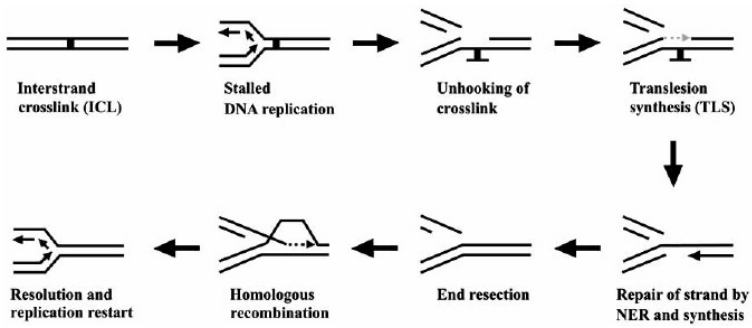


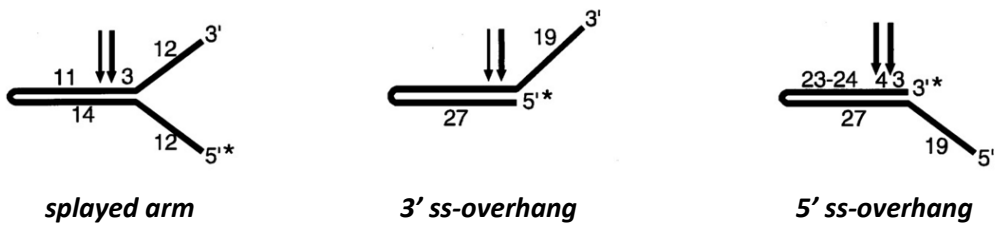
Figura 34. Esquema de la reparación de ICLs en el ADN.

Extraído de *Andreassen et al., (2009) Current cancer drug targets, 9, 101-117.*

Los estudios de los síndromes de inestabilidad cromosómica y de ciertos cánceres de origen genético han sido críticos para identificar los mecanismos implicados en las respuestas celulares a los daños en el ADN. Además ciertas líneas celulares de pacientes con este tipo de enfermedades tienen especial sensibilidad o resistencia a determinados agentes alquilantes. En particular, la anemia de Fanconi (FA) está asociada a hipersensibilidad a agentes que inducen ICLs. De hecho, células de FA que poseen deficiencias en algunas de las proteínas de la vía de HR son extremadamente sensibles a Yondelis® (54), al igual que a MMC, que es el fármaco que se utiliza para su diagnóstico (63). Por el contrario, Yondelis® es más citotóxico en células que sobreexpresan proteínas implicadas en NER, como por ejemplo XPG y ERCC1-XPF (64).

Con el objetivo de obtener más detalles acerca del mecanismo de acción de estos fármacos tetrahidroisoquinolínicos, se investigó el efecto que tienen en la reparación de ADN en colaboración con investigadores del *Institut de Génétique et de Biologie Moléculaire et Cellulaire* de Estrasburgo y de la compañía biotecnológica

PharmaMar. En particular, las investigaciones se centraron en el efecto que Yondelis® y Zalypsis® tienen en la actividad de XPF. Esta enzima es una endonucleasa capaz de hidrolizar estructuras irregulares de ADN durante la replicación, HR o NER (65). XPF en eucariotas funciona como un heterodímero con ERCC1 pero sólo XPF tiene un dominio nucleasa. Esta enzima hidroliza específicamente el ADN de doble cadena en la región 5' de su punto de bifurcación con una cadena sencilla (66) **(Figura 34)**.



splayed arm ***3' ss-overhang*** ***5' ss-overhang***

Figura 34. Elementos estructurales mínimos requeridos para la actividad de ERCC1-XPF. La posición exacta del punto de corte (flechas) puede variar entre 2 y 8 nucleótidos desde el punto de unión entre ds-y ss-ADN.
Extraído de *de Laat et al., (1998) The Journal of Biological Chemistry, 14,7835-7842.*

Esta acción de XPF/ERCC1 es esencial en NER y, junto a la de XPG, una endonucleasa que actúa por el lado contrario de la lesión, interviene para eliminar el fragmento dañado, que quedará después resintetizado por polimerasas.

Las investigaciones demostraron que la hidrólisis de una molécula de ADN por XPF se incrementaba notablemente cuando el oligo contenía un aducto con Yondelis® o Zalypsis®. Parte del trabajo de esta memoria consistió en explicar estos resultados experimentales mediante la generación y refinado de modelos moleculares que representan complejos ternarios entre ADN, fármaco y XPF/ERCC1 **(Figura 35).**

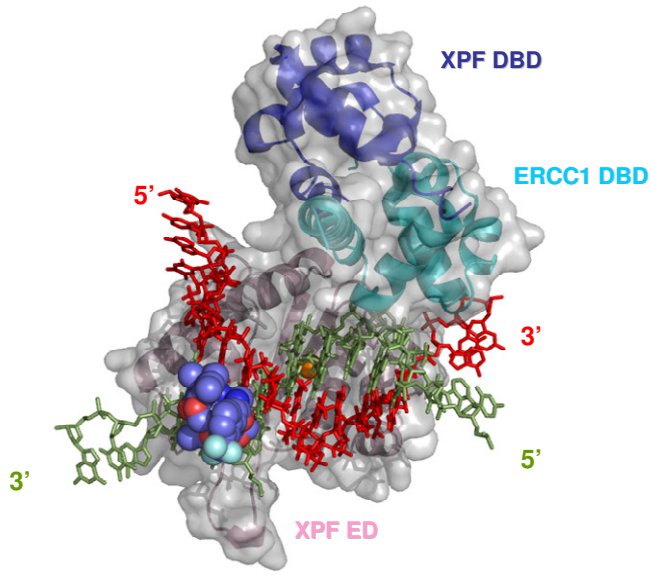


Figura35. Modelo molecular de la interacción de los dominios endonucleasa (ED) y de unión al ADN (DBD) de XPF/ERCC1 con el ADN y Zalypsis®

También se demostró que estos fármacos, además de estimular la actividad de XPF, eran capaces de bloquear la transcripción por la ARN polimerasa II (67) **(Figura 36)**. De nuevo, utilizamos métodos teóricos y gráficos moleculares para intentar proporcionar una explicación en términos atomísticos a estas observaciones macroscópicas.

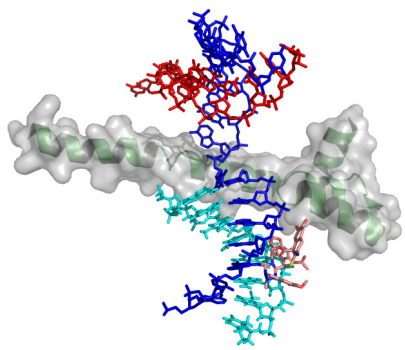


Figura 36. Modelo de la interacción entre ADN (azul), ARN (rojo), ARN polimerasa II (verde) y Yondelis (anaranjado).

1.7. Determinación del Mecanismo de Acción de Nucleasas de Interés Biológico por Modelado Molecular.

Los métodos actuales de modelado molecular son capaces de aportar detalles fundamentales para comprender el mecanismo catalítico de las reacciones enzimáticas (68). El objetivo principal de modelar una reacción enzimática es intentar resolver algunas de las cuestiones que son difíciles de abordar con métodos puramente experimentales. Los métodos "estándar" que emplean mecánica molecular (con campos de fuerzas tipo AMBER, CHARM o GROMOS) son importantes para estudiar la estructura y dinámica de la proteína pero no son aplicables para estudiar reacciones químicas. Por este motivo se combinan con los métodos mecánico-cuánticos (métodos QM/MM) que tienen en cuenta la estructura electrónica para poder investigar la formación y rotura de enlaces.

Utilizando estos métodos es posible llegar a conocer con exactitud el mecanismo de acción de numerosas enzimas. En particular, y para completar los estudios sobre nucleasas, nos centramos en el mecanismo de acción de las endonucleasas no específicas. Estas nucleasas son capaces de cortar el ADN o ARN de cadena sencilla o doble con poca o ninguna especificidad. Las enzimas más representativas de este grupo son la nucleasa de *Serratia marcescens*, la nucleasa A de *Anabaena* sp. y la nucleasa periplasmica de *Vibrio vulnificus* (*Vvn*). Desde un punto de vista estructural, todas estas enzimas pertenecen a la superfamilia de nucleasas del tipo ββα-metal que emplean únicamente un solo ión metálico divalente. Debido a que su mecanismo de acción no está totalmente claro hasta la fecha, como se puede deducir del gran número de sugerencias que aparecen en la literatura, nos propusimos investigar el mecanismo catalítico de esta familia de enzimas utilizando la nucleasa de *Vvn* como enzima representativa de la familia **(Figura 37).**

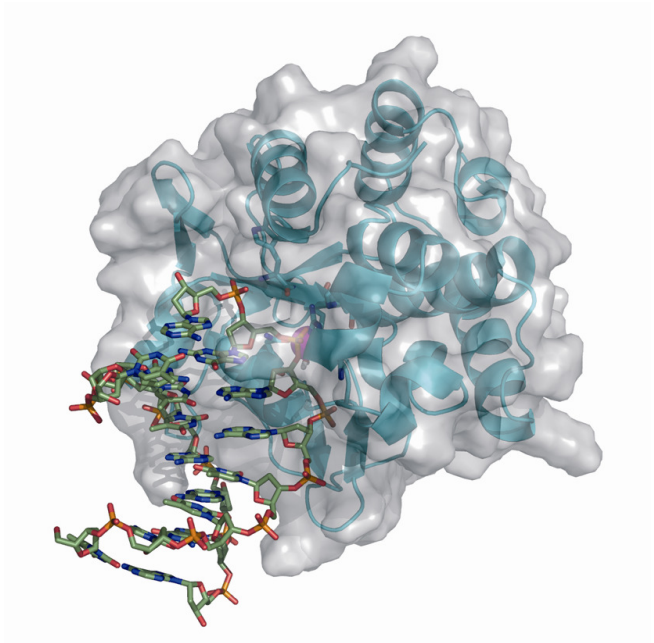


Figura 37. Modelo estructural de *Vvn* (azul) formando un complejo con ADN (verde)

2. OBJETIVOS

"Confía en el tiempo, que suele dar dulces salidas a muchas amargas dificultades."

Miguel de Cervantes

Los objetivos principales de esta tesis fueron, en primer lugar estudiar la interacción del ADN con diversos fármacos antitumorales y con diversas familias de proteínas aplicando técnicas de modelado molecular por ordenador en estrecha colaboración con distintos grupos experimentales de la Universidad de Alcalá e investigadores de *PharmaMar*. Además, con el objetivo de completar algunas de las cuestiones sin resolver en cuanto al mecanismo de acción de algunas nucleadas de interés biológico, se emplearon simulaciones de MD y métodos híbridos de QM/MM para proponer un mecanismo catalítico para la nucleasa de *Vibrio vulnificus*, enzima para la que se cuenta con una amplia información estructural.

Las investigaciones llevadas a cabo se dividen en los siguientes apartados:

1. Descripción de la interacción del ADN con el nuevo fármaco antitumoral *Zalypsis®* y análisis de su especificidad por técnicas de modelado molecular.

2. Descripción de la interacción del ADN con el nuevo fármaco antitumoral *PM01183* y análisis de su especificidad por técnicas de modelado molecular.

3. Estudio de la interacción entre el ADN y la enzima de reparación XPF y la ARN polimerasa II. Consecuencias de incorporación de un aducto covalente con los fármacos antitumorales *Yondelis®* y *Zalypsis ®*.

4. Estudio mediante simulaciones de MD de la fusión inducida por temperatura de un segmento de ADN de doble cadena en ausencia y presencia de fármacos covalentemente unidos.

5. Razonamiento para la esteroquímica opuesta de los monoaductos y entruzamientos intercatenarios principales formados por mitomicina C y su análogo descarbamoilado en pasos CpG de ADN y efecto de citosinas modificadas en su reactividad.

6. Estudio por simulaciones de MD del mecanismo de acción de la endonucleasa de *Vibrio vulnificus* e implicaciones en distintas familias de endonucleasas.

3. RESULTADOS Y DISCUSIÓN

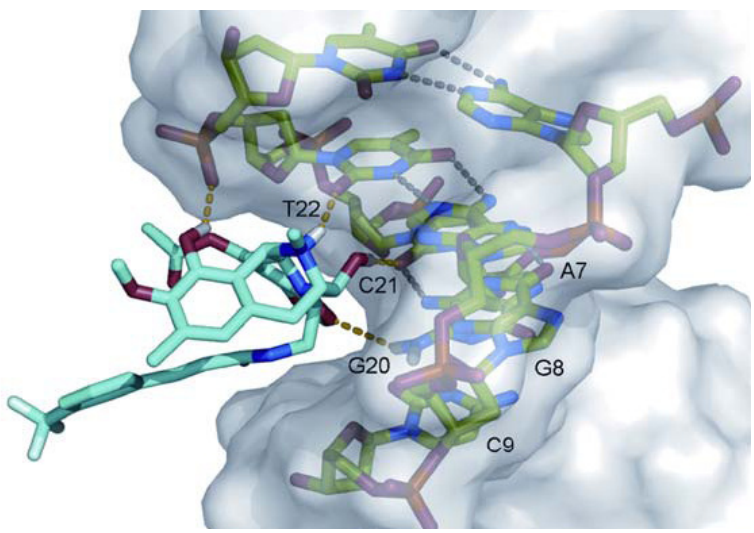
"*La patience est amère, mais ses fruits sont doux*."

Jean Jacques Rousseau

Artículo I

"Molecular pharmacology and antitumor activity of Zalypsis in several human cancer cell lines"

***Biochem Pharmacol.* 2009 Jul 15;78(2):162-70.**



Zalypsis es un nuevo alcaloide tetrahidrisoquinolínico sintético con un grupo carbinolamino reactivo. Este grupo funcional da lugar a la formación de un enlace covalente con el grupo amino de determinadas guaninas en la doble hélice del ADN, tanto en ausencia como en presencia de citosinas metiladas. El complejo resultante es, además, estabilizado por la formación de uno o más enlaces de hidrógeno con los nucleótidos adyacentes en la cadena opuesta, así como por interacciones de van der Waals en el surco menor. Experimentos de desnaturalización térmica basados en fluorescencia han demostrado que los tripletes de ADN más favorables para la formación de aductos covalentes son AGG, GGC, CAG, CGG y TGG, y estas preferencias se pueden racionalizar gracias a los resultados del modelado molecular. Los aductos de Zalypsis y ADN con el tiempo dan lugar a roturas de doble cadena, lo que provoca la acumulación de la fase S y la muerte celular apoptótica. La potente actividad citotóxica de Zalypsis se comprobó en un panel de células línea de 24. El valor medio de la IC (50) fue de 7nM y las líneas celulares de leucemia y tumor de estómago se encontraban entre las más sensibles. La administración de Zalypsis en cuatro modelos de xenoinjerto murino de cáncer humano demuestra una inhibición significativa del crecimiento del tumor que es más alta en la línea celular de cáncer gástrico Hs746t sin pérdida de peso de los animales tratados. En conjunto, estos resultados indican que la potente actividad antitumoral de Zalypsis apoya su desarrollo actual en la clínica como un agente contra el cáncer.

Biochemical Pharmacology 78 (2009) 162–170

Contents lists available at ScienceDirect

Biochemical Pharmacology

journal homepage: www.elsevier.com/locate/biochempharm

Molecular pharmacology and antitumor activity of *Zalypsis*® in several human cancer cell lines

Juan F.M. Leal [a], Verónica García-Hernández [b], Victoria Moneo [a], Alberto Domingo [b], Juan Antonio Bueren-Calabuig [c], Ana Negri [c], Federico Gago [c], María José Guillén-Navarro [a], Pablo Avilés [a], Carmen Cuevas [a], Luis Francisco García-Fernández [a], Carlos María Galmarini [a,*]

[a] *Cell Biology Department, Pharmamar SA, Avda. de los Reyes, 1, 28770 Colmenar Viejo, Madrid, Spain*
[b] *Department of Biochemistry and Molecular Biology, University of Alcalá, Spain*
[c] *Department of Pharmacology, University of Alcalá, Spain*

ARTICLE INFO

Article history:
Received 13 February 2009
Accepted 6 April 2009

Keywords:
Antineoplastic agents
Tetrahydroisoquinolines
DNA breaks
Apoptosis

ABSTRACT

Zalypsis® is a new synthetic alkaloid tetrahydroisoquinoline antibiotic that has a reactive carbinolamine group. This functionality can lead to the formation of a covalent bond with the amino group of selected guanines in the DNA double helix, both in the absence and in the presence of methylated cytosines. The resulting complex is additionally stabilized by the establishment of one or more hydrogen bonds with adjacent nucleotides in the opposite strand as well as by van der Waals interactions within the minor groove. Fluorescence-based thermal denaturation experiments demonstrated that the most favorable DNA triplets for covalent adduct formation are AGG, GGC, AGC, CGG and TGG, and these preferences could be rationalized on the basis of molecular modeling results. Zalypsis®–DNA adducts eventually give rise to double-strand breaks, triggering S-phase accumulation and apoptotic cell death. The potent cytotoxic activity of Zalypsis® was ascertained in a 24 cell line panel. The mean IC_{50} value was 7 nM and leukemia and stomach tumor cell lines were amongst the most sensitive. Zalypsis® administration in four murine xenograft models of human cancer demonstrates significant tumor growth inhibition that is highest in the Hs746t gastric cancer cell line with no weight loss of treated animals. Taken together, these results indicate that the potent antitumor activity of Zalypsis® supports its current development in the clinic as an anticancer agent.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Zalypsis® (Fig. 1A) is a new synthetic dimeric isoquinoline alkaloid that is currently in Phase I clinical development for the treatment of solid tumors and hematological malignancies. This compound is structurally related to jorumycin, a natural compound isolated from the skin and mucus of the Pacific nudibranch *Jorunna funebris* [1], and also to renieramycins isolated from sponges and tunicates [2,3], safracins and saframycins isolated from bacteria and marine sponges [4], and ecteinascidins isolated from marine tunicates [5]. These carbinolamine-containing compounds react, through an iminium intermediate that is generated by dehydration [4,6–8], with the exocyclic amino group of selected guanines in the minor groove of DNA. The resulting adduct is additionally stabilized through the establishment of van der Waals interactions and one or more hydrogen bonds with neighboring nucleotides in the opposite strand of the DNA double helix [8,9], thus creating the equivalent to a functional interstrand crosslink [10] that can lead to strong inhibition of the early phases of transcription [11,12].

In this study we characterized both molecular interactions with the DNA and the cytotoxic activity of Zalypsis® in an attempt to shed light into its mechanism of antitumor action. The DNA-binding characteristics were studied using a combination of electrophoretic mobility shift assays in polyacrylamide gels and a fluorescence-based method that employs tailor-made oligonucleotides. We also looked at the type of DNA damage subsequently generated in living cells as a consequence of Zalypsis®–DNA adduct formation and whether cell cycle arrest and cell death induced by this compound mechanistically depend on caspase activation. Then, *in vitro* cytotoxic activity was evaluated in a panel of 24 different cancer cell lines. Finally, the *in vivo* antitumor activity of Zalypsis® was investigated using several murine xenograft models of human prostate, stomach, bladder and pancreas carcinomas.

* Corresponding author. Tel.: +34 91 846 6158; fax: +34 91 846 6001.
E-mail address: cgalmarini@pharmamar.com (C.M. Galmarini).

0006-2952/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bcp.2009.04.003

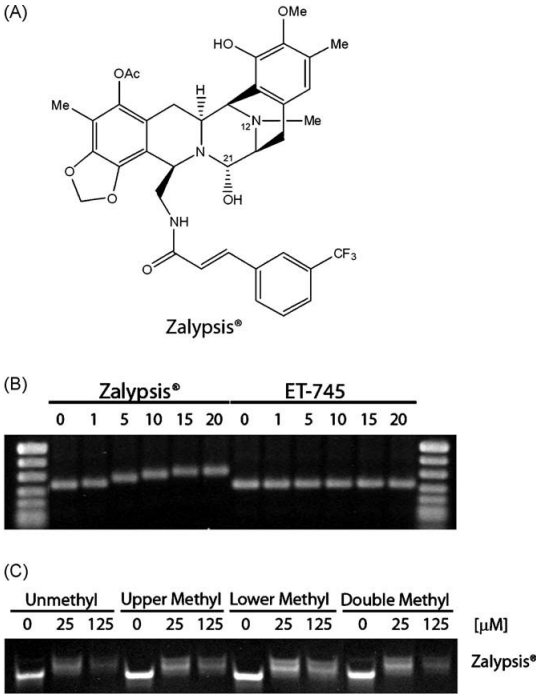


Fig. 1. Cytotoxic effect and DNA-binding properties of Zalypsis®. (A) Structure of Zalypsis® showing the positions of *C21* and *N12* atoms discussed on the main text. (B) Binding to naked DNA. Drugs were incubated with a naked 250 bp PCR product at 25 °C during 1 h and the electrophoresis run in 2% agarose-TAE. (C) Binding to methylated DNA. The methylated ds-oligo with a unique binding site for Zalypsis® was incubated with the drug at 25 °C during 1 h, then electrophoresis was run in a 20% acrylamide-TAE gel.

2. Materials and methods

2.1. Drugs

Zalypsis® is a Pharmamar SAU proprietary drug synthesized at our facilities. ET-745 is a synthetic isoquinoline derived from trabectedin (ET-743) that lacks the carbinolamine group (Supplementary Material, Figure 1) and does not alkylate DNA [13]. Z-Vad-fmk is a pan-caspase inhibitor purchased from Sigma (St. Louis, MO, USA).

2.2. DNA electrophoretic mobility shift assay

The binding assays were performed either with a 250 bp PCR product from the human adiponectin gene or with a double-stranded oligonucleotide (ds-oligo) synthesized by Sigma (St. Louis, MO, USA). Briefly, appropriate concentrations of the compound were added at 25 °C during 1 h to 250 bp DNA or ds-oligo. Then, DNA was subjected to electrophoresis in a 2% (w/v) agarose/TAE gel, stained with 1 μg/ml ethidium bromide and photographed. The sequences of the oligonucleotides used in the assay are the following: NonMeth-Fw 5′-TAAATATTTCCGGATATAATTAT-3′; NonMeth-Rev 5′-ATAATTATATCCGGAAATATTTA-3′; Meth-Fw 5′-TAAATATTTC**MeC**GGATATAATTAT-3′; Meth-Rev 5′-ATAATTATATC**MeC**GGAAATATTTA-3′.

2.3. DNA melting assay

Synthetic oligodeoxynucleotides with one strand 5′-end-labeled with the fluorophore 6-carboxyfluorescein (6-FAM, F) and the complementary strand 3′-end-labeled with the quencher tetramethylrhodamine (TAMRA, Q) were synthesized at Bonsai Technologies (Madrid, Spain) (Supplementary Table 1). For the experiments, we followed the methodology previously described in a 7500 Fast Real-Time PCR System (ABI Prism, Applied Biosystems, Foster City, CA, USA) [10,14]. The raw data obtained were analyzed to estimate the increases in melting temperatures (ΔT_m) brought about by drug binding as well as the ligand concentration that produces half the maximal change in melting temperature (C_{50}). The inverse of this value ($1/C_{50}$) was taken as a measure of the relative DNA-binding affinity. We additionally used the parameter ΔT_m(max) to reflect the relative stability of the DNA-ligand complexes [14]. The complete analysis was carried out using an in-house developed Visual Basic Application running on Microsoft Excel (Microsoft, Redmond, WA, USA).

2.4. Computational methods

We studied in atomic detail the feasibility of achieving the geometries required to activate dehydration of the carbinolamine prior to the nucleophilic attack that leads to covalent bond formation. To this end, the initial structures of the precovalent complexes between Zalypsis® and a series of oligonucleotides of general sequence 5′-d(ATAATAXYZATAATA)/5′-d(TATTATZ′Y′X′TATTAT), where XYZ/Z′Y′X′ stands for AGA/TCT, AGG/CCT and AGC/GCT, were built as previously reported for trabectedin [15]. The complexes were then refined using energy minimization techniques in the AMBER force field and their stability was studied by means of 5 ns of unrestrained molecular dynamics simulations in an explicit water box using the same conditions as described previously.

2.5. Cell culture and cytotoxicity

All the tumor cell lines used in this study were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA). For the cytotoxicity experiments, cells were seeded in 96-well trays. Serial dilutions of the compound dissolved in dimethyl sulfoxide (DMSO) were prepared and added to the cells in fresh medium, in triplicates. Exposure to the compounds was maintained during 72 h. Then 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Sigma, St. Louis, MO, USA) was added to the cells and formazan crystals dissolved in DMSO. Absorbance at 540 nM was measured with a POLARStar Omega Reader (BMG Labtech, Offenburg, Germany). Determination of IC_{50} values was performed by iterative non-linear curve fitting with the Prism 5.0 statistical software (GraphPad, La Jolla, CA, USA). The data presented are the average of three independent experiments performed in triplicate.

2.6. Fluorescent microscopy

Cells were treated with the appropriate concentration of Zalypsis® during 6 h, washed out and cultured for 18 additional hours. Cells were then fixed (4% paraformaldehyde), permeabilized (0.5% Triton X-100) and incubated with the primary anti-γ-H2AX monoclonal antibody (Upstate, Temecula, CA) for 1 h at 37 °C. Then cells were washed and incubated with the secondary anti-mouse AlexaFluor 594 (Invitrogen, Carlsbad, CA, USA) for 30 min at 37 °C. Finally the slides were incubated with Hoesch 33342 (Sigma, St. Louis, MO, USA) and mounted with Mowiol mounting medium. Pictures were taken with a Leica DM IRM fluorescence microscope equipped with a 100× oil immersion objective and a DFC 340 FX digital camera (Leica, Wetzlar, Germany).

2.7. Comet assay

For the determination of double-strand breaks (DSBs) a single cell gel electrophoresis assay was used (Trevigen's CometAssay™), following the manufacturer's instructions after treatment of cells for 12 h with the appropriate concentration of Zalypsis®. Pictures were taken with a Leica DM IRM fluorescence microscope equipped with a DFC 340 FX digital camera (Leica, Wetzlar, Germany). Quantitation of the DNA in the tails of the comets was performed with Adobe Photoshop CS3 (Adobe Systems Incorporated, San Jose, CA, USA). For each condition 30 cells were analyzed, and the experiments were repeated several times.

2.8. Cell cycle analysis

For the cell cycle experiments, cells were treated with the appropriate amount of the compound for 24 h, and then stained with 0.4 μg/ml propidium iodide. Samples were analyzed with a FACScalibur flow cytometer (Beckton and Dickinson, Franklin Lakes, NJ, USA) and the FlowJo7 cytometry analysis software.

2.9. Apoptosis

For the chromatin condensation assay, cells were treated with the appropriate amount of the compound for 24 h and stained with DAPI. Early apoptotic cells show chromatin condensation that was assessed with a DM IRM microscope (Leica, Wetzlar, Germany). For the M30-Apoptosense solid-phase sandwich enzyme immunoassay (Peviva, Bromma, Sweden), the kit's instructions were followed. Briefly, cells were exposed to the appropriate concentration of the compound for 24 h. Cells lysates were obtained and transferred to an assay well, M30 HRP conjugate added and incubated for 4 h. Once the antigen-M30 was bound to the M5-coated surface of the wells, they were extensively washed and incubated with TMB substrate in the dark for 20 min. Finally, the reaction was stopped with the addition of a stop solution and plates were read at 450 nM within 30 min with a Victor3 platform (PerkinElmer).

2.10. Western blot assays

For immunoblotting, cells were treated with the appropriate concentration of the compound for 24 h and lysed with RIPA lysis buffer. Protein content was determined by the modified Bradford method. Samples were separated in 7.5% SDS-PAGE, transferred onto an Immobilon-P membrane. Then, membranes were incubated with the appropriate primary antibody 1 h, washed and incubated with the secondary antibody. Finally protein was visualized using the ECL System (GE Healthcare, Fairfield, CT, USA). We used anti-PARP rabbit polyclonal, anti-p53FL rabbit polyclonal and anti-p21/WAF1 rabbit polyclonal antibodies from Santa Cruz (Santa Cruz, CA, USA), and an anti-α-tubulin monoclonal antibody from Sigma (St. Louis, MO, USA). Secondary antibodies were HRP-conjugated goat anti-rabbit secondary antibody (R&D Minneapolis, MN, USA) and HRP-conjugated goat anti-mouse secondary antibody (Santa Cruz, CA, USA).

2.11. Antitumor activity in xenograft murine models

Four- to six-week-old athymic *nu*/*nu* mice (Harlan Sprague Dawley, Madison, WI, USA) were s.c. xenografted into their right flank with ca. $(0.5–1) \times 10^7$ cells in 0.2 ml of a mixture (50:50; v:v) of Matrigel basement membrane matrix (Beckton Dickinson, Frankin Lakes, NJ, USA) and serum-free medium. When tumors reached ca. 150 mm^3, mice were randomly assigned into treatment or control groups. Zalypsis® was intravenously administered either in 3 consecutive weekly doses (0.9 mg/kg/day) or in 2 cycles of 5 consecutive daily doses (0.3 mg/kg/day) in pancreas or in breast and prostate xenograft studies, respectively. Control animals received an equal volume of vehicle. Caliper measurements of the tumor diameters were done twice weekly and tumor volumes calculated according to the following formula: $(ab)^2/2$, where a and b were the longest and shortest diameters, respectively. Animals were humanely euthanized, according to Institutional Animal Care and Use Committee of PharmaMar, Inc. (Cambridge, MA, USA) guidelines, when their tumors reached 3,000 mm^3 or if significant toxicity (e.g. severe body weight reduction) was observed. Differences in tumor volumes between treated and control groups were evaluated using the unpaired t-test. Statistical significance was defined as $p < 0.05$. Statistical analyses were performed by LabCat® v8.0 SP1 (Innovative Programming Associates, Inc. NJ, USA). All animal studies were conducted under approval from an IACUC in an AAALAC accredited animal facility.

3. Results

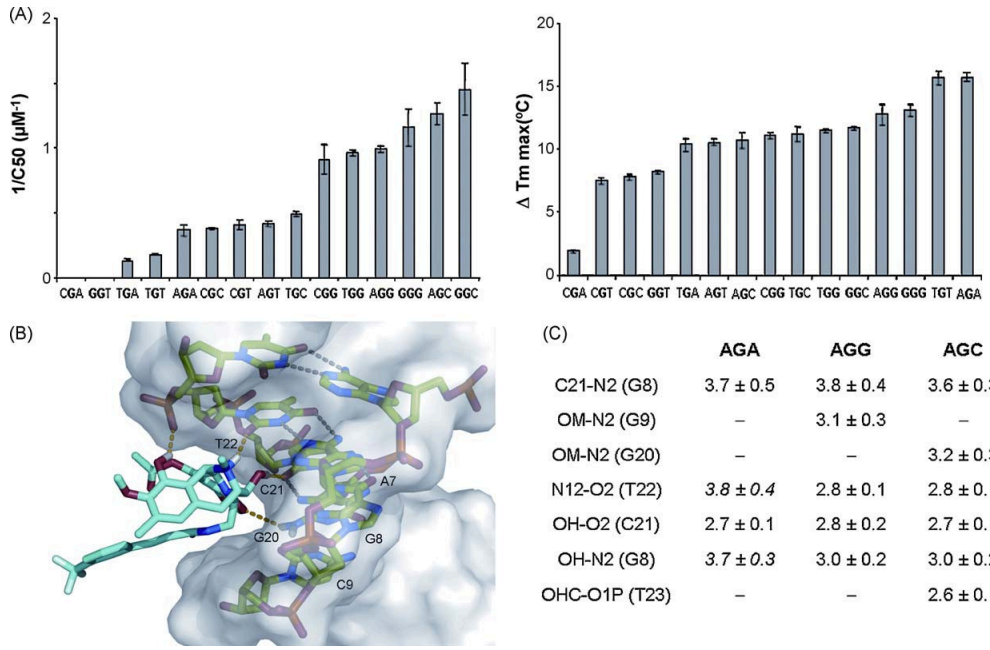
3.1. Zalypsis® binds covalently to DNA

The mechanism of action of Zalypsis® is expected to rely on the alkylation of selected guanines in the DNA double helix. Thus, we investigated its binding to DNA by means of band shift assays that allow the qualitative assessment of the molecular weight increase of Zalypsis®–DNA adducts. Fig. 1B shows a typical band shift experiment using increasing amounts of Zalypsis® in order to saturate the 250 bp DNA probe with the compound. ET-745, which lacks the crucial hydroxyl group required for activation and covalent DNA modification, was used as a negative control (Supplementary Figure 1). As expected, Zalypsis® bound to naked DNA and delayed its electrophoretic migration in the gel. Although 1 μM of Zalypsis® was enough to produce a slight delay in band migration, DNA saturation required ~15 μM of the compound.

We further assessed whether Zalypsis® could alkylate methylated DNA, using a ds-oligo containing a methylated cytosine in the sequence *me*CGG. Fig. 1C shows the result of a typical experiment. Zalypsis® bound to methylated and unmethylated CGG sequences with similar efficiencies, suggesting that methylation of the cytosine in a CpG sequence context in the major groove does not interfere with guanine modification in the DNA minor groove.

3.2. Zalypsis® binds to guanines in selected DNA triplets

To investigate the ability of Zalypsis® to stabilize the double helix, we designed several fluorophore-labeled oligonucleotides encompassing most of the possible DNA triplet combinations involving a central guanine (Supplementary Table 1). Fig. 2A summarizes the results of the DNA melting experiments. The average $1/C_{50}$ and ΔT_m(max) values for Zalypsis® were ~0.6 μM^{-1} and ~10 °C, respectively. There were differences in the affinity of Zalypsis® for the different triplets assayed, with the highest affinity being observed for GGC, AGC, AGG and TGG triplets, and the lowest, almost negligible, being displayed by the CGA triplet. Nonetheless, some of the selected triplets (e.g. AGC or AGG) do not provide a good binding site if embedded in an alternative sequence context (e.g. AGCA vs. AGCG and AGGA vs. AGGG). On the other hand, the highest stabilities were observed for adducts formed with the oligos containing the triplets AGA, TGT, AGG, GGC and TGG. Thus, according the affinity indicator $1/C_{50}$, the most favorable triplets for Zalypsis® bonding to DNA appear to be AGG, GGC, AGC, CGG and TGG whereas the thermal stability of the DNA-adduct involving an AGA or TGT site is slightly above average



	AGA	AGG	AGC
C21-N2 (G8)	3.7 ± 0.5	3.8 ± 0.4	3.6 ± 0.3
OM-N2 (G9)	–	3.1 ± 0.3	–
OM-N2 (G20)	–	–	3.2 ± 0.3
N12-O2 (T22)	*3.8 ± 0.4*	2.8 ± 0.1	2.8 ± 0.1
OH-O2 (C21)	2.7 ± 0.1	2.8 ± 0.2	2.7 ± 0.1
OH-N2 (G8)	*3.7 ± 0.3*	3.0 ± 0.2	3.0 ± 0.2
OHC-O1P (T23)	–	–	2.6 ± 0.1

Fig. 2. Characteristics and modeling of DNA–Zalypsis® complexes. (A) Relative binding affinities for different DNA triplet sequences. Zalypsis® was incubated with the labeled ds-oligos at 25 °C during 1 h, then the melting assay was started in a 7500 Fast Real-Time PCR System by increasing the temperature up to 95 °C in small steps of 1 °C min^{-1}. Analysis of the $1/C_{50}$ parameter was done with an in-house developed Visual Basic Application (VBA) running on Microsoft Excel. An additional parameter, ΔT_m(max), which is related to the stability of the drug-DNA complex, was also analyzed using the VBA. The values correspond to the mean and standard deviation from at least two independent experiments for each sequence. (B) View of the central part of the precovalent Zalypsis®:AGC model complex in which a semitransparent solvent-accessible surface envelops the DNA atoms. Drug and DNA carbon atoms have been colored cyan and green, respectively. Only hydrogens attached to *N12* and *O21* of Zalypsis® and the amino groups of G8 and G20 are shown. Crucial hydrogen bonds are displayed as dotted yellow lines for the drug-DNA complex and dashed black lines for the DNA base pairs. (C) Attacking distances and relevant intermolecular distances (Å) between hydrogen-bonding donor and acceptor atoms in the three precovalent complexes studied (the italics for the AGA complex highlight longer distances than those needed to establish a hydrogen bond).

and those involving CGC, CGT, GGT and CGA sites are well below average.

3.3. Molecular modeling data

The molecular models of the equilibrated Zalypsis®–DNA precovalent complexes in aqueous solution provided a rationale for the distinct binding of the drug to three representative central triplets that provide good (AGC), intermediate (AGG) or poor (AGA) binding sites for this drug, the only difference among them being the nature of the base pair on the 3′-side of the guanine that will undergo alkylation. As shown in Table 1, the geometry required for nucleophilic attack of the exocyclic amino group of the central guanine (G8) at the carbinolamine of the drug (*C21*) was rapidly attained in all cases, as assessed by an interatomic distance of less than 4 Å between the two reactive atoms. Likewise, the carbinol OH was seen to act as a hydrogen bond acceptor from N2(G8) and as a hydrogen bond donor to O2 in C21, that is, the cytosine pairing with G8 (Fig. 2B). These findings are consistent with the fact that Zalypsis® is able to react with these triplet sites, and the geometry that we find in these precovalent complexes lends strong support to the proposal that the latter hydrogen bond weakens the *N12–H12* covalent bond, thus facilitating transfer of the *H12* proton to the carbinol OH, whereas the first hydrogen bond debilitates the *C21–OH* bond, thus favoring the exit of the water molecule. On the other hand, the protonated *N12* of Zalypsis® remained hydrogen-bonded to the O2 acceptor atom of T22 in the AGG and AGC triplets but not in the AGA triplet (Fig. 2C). This finding, together with the fact that AGA cannot provide a hydrogen-bonding partner to the methylenedioxy oxygen facing the minor groove may account for the fact that AGA is a poorer binding site even though the DNA in the resulting complex is stabilized to a greater extent than those containing the other sequence contexts studied (Fig. 2A). When this hydrogen-bonding partner is found, as is the case for the AGC and AGG sequences which possess the amino group of either G20

Table 1
Cell cycle perturbations[a] induced by Zalypsis®.

	Cell cycle phase	Control	15 nM	150 nM
A549	Sub-G_1	1.7	13	5.7
	G_0/G_1	60	24	49
	S-phase	25	49	30
	G_2/M	8.6	7.6	7.1
HT29	Sub-G_1	0.3	15	21
	G_0/G_1	64	17	38
	S-phase	38	63	36
	G_2/M	2	9.2	3.8
MDA-MB-231	Sub-G_1	2.3	2.2	7
	G_0/G_1	37	12	23.3
	S-phase	30	47	33
	G_2/M	13	22	18
IGROV-1	Sub-G_1	2.8	56	68
	G_0/G_1	39	11	6.5
	S-phase	33	21	18
	G_2/M	16	8.5	1.7

[a] Cells were treated with the appropriate amount of the compound for 24 h, and then stained with 0.4 μg/ml propidium iodide. Samples were analyzed with a FACScalibur flow cytometer (Beckton and Dickinson, Frankin Lakes, NJ, USA) and the FlowJo7 cytometry analysis software. Values represent % of cells in each phase of the cell cycle.

or G9, respectively, the additional hydrogen bond formed (Fig. 2B and C) facilitates the full insertion of the drug into the minor groove and leads to enhanced reactivity. As regards the small stabilization brought about by Zalypsis® on the CGA-containing oligonucleotide, this can be rationalized by proposing that it arises only from the precovalent complex, which cannot evolve into a covalent adduct because of an incorrect geometry for attack (14).

3.4. The initial Zalypsis®–DNA adducts eventually give rise to double-strand breaks in living cells

We analyzed the formation of γ-H2AX foci as a surrogate indicator of DSB formation [16,17]. Fig. 3A shows typical results of MCF7 and A549 cells treated with Zalypsis® for 6 h at the indicated concentrations. At 5 nM, only a few MCF7 or A549 cells showed γ-H2AX staining and, in the positive cells, few foci were visible (data not shown). In contrast, at 25 nM, most of the cells were stained for γ-H2AX and each of the positive cells carried a large number of foci, indicating that several DSBs per cell were induced at this drug concentration (Fig. 3A). At 100 nM, nearly every single cell stained positive for γ-H2AX, with a large number of foci per nucleus, again an indication of extensive DNA damage and induction of numerous DSBs (data not shown).

Zalypsis®-dependent formation of DSBs was additionally evaluated using the comet assay, which provides a direct measurement of damaged DNA. Both A549 and MCF7 cell lines were treated with Zalypsis® at 10 and 100 nM for 24 h and the percentage of the total DNA that was in the tail of the comets was measured. A typical comet assay result (Fig. 3B) shows a clear concentration-dependent increase in DSBs in both cell lines following treatment with Zalypsis®. The use of Mann–Whitney *U*-test demonstrated that differences between each treated-measure against the baseline were statistically significant.

3.5. Zalypsis®-treated cells accumulate in S-phase

Table 1 summarizes the results of the assessment of possible cell cycle perturbations after 24 h of exposure to Zalypsis®. At both low (15 nM) and high (150 nM) drug concentrations there was a clear increase of the S-phase population in A549 (lung), HT29 (colon) and MDA-MB-231 (breast) cancer cells. Of note, the S-phase accumulation was higher at 15 nM, probably due to a higher cell survival. In IGROV-1 cells we did not observe any S-phase accumulation, most likely due to the extensive cell death induced by the compound at both concentrations after 24 h treatment (data not shown).

3.6. Zalypsis® induces apoptosis in tumor cells

Since the antitumor mechanism of action of DNA-damaging agents is known to involve induction of apoptosis [18], we investigated whether Zalypsis® treatment was able to induce a canonical apoptosis response in tumor cells. Firstly, we determined whether the drug was able to generate late apoptosis by detecting

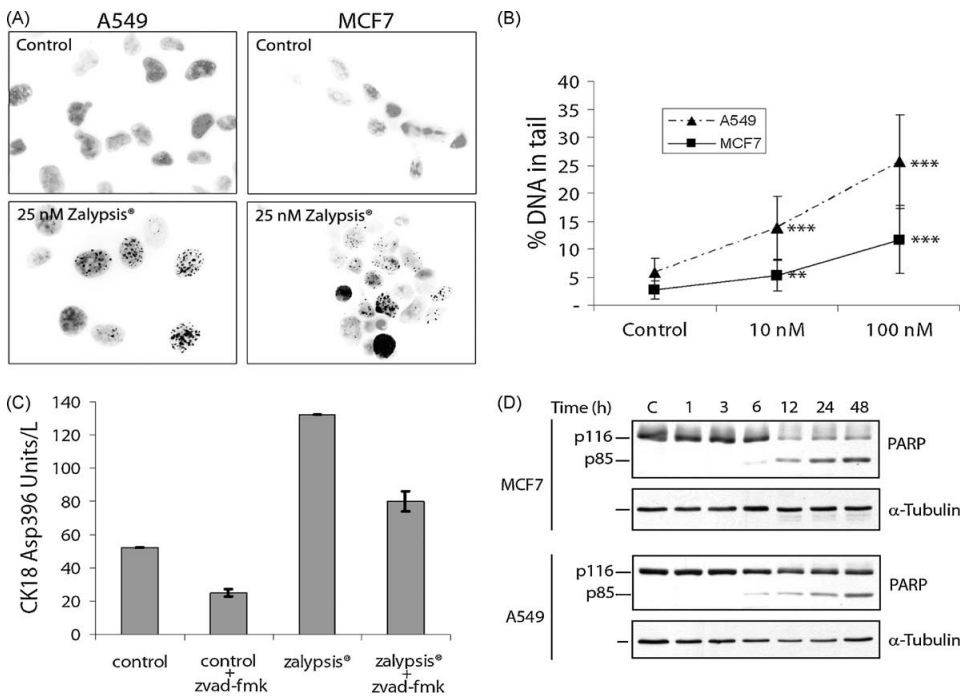


Fig. 3. Zalypsis® treatment induces DNA double-strand breaks and apoptosis in living cells. (A) A549 or MCF7 cells were treated with the drug at the indicated concentrations during 6 h, followed by additional 18 h of incubation without the drug. After fixation, cells were immunostained for γ-H2AX and nuclei were visualized with Hoestch 33342. (B) Zalypsis®-treated cells were simultaneously analyzed through comet assay to determine the amount of DSBs induced by the drug. After 12 h of treatment with the indicated concentrations of Zalypsis®, cells were washed with PBS, detached mechanically and included in low melting point agarose. After a treatment of lysis, cells were subjected to electrophoresis and the comets were stained with SYBR green. Quantitation of the amount of DNA in the comets was performed with Photoshop CS3. $^{**}p < 0.001$; $^{***}p < 0.0001$. (C) Apoptosis induced by Zalypsis® is caspase-dependent. A549 cells were treated with the drugs during 24 h, washed and processed following the protocol of the M30-Apoptosense solid-phase sandwich enzyme immunoassay. (D) Caspase activation and PARP-1 processing. MCF7 and A549 cells were treated with Zalypsis® 100 nM during 24 h and then washed and lysed with RIPA buffer supplemented with a protease inhibitor cocktail. Western blot was then performed using a primary anti-PARP-1 polyclonal antibody.

the flow-cytometric sub-G_1 cell population (Table 1). A549 and HT-29 cell treatment resulted in an increase in the sub-G_1 population. On the other hand, MDA-MB-231 cells died with a slower kinetics while IGROV-1 cells died faster (Table 1). These data indicate that the percentage of cells entering the sub-G_1 peak after 24 h of treatment with Zalypsis® was cell type- and concentration-dependent. After treatment of MCF7 and A549 cells with 100 nM Zalypsis® for 24 h, an apoptotic nuclear morphology with strong chromatin condensation was clearly visible in fluorescent microscopy in both cell types (data not shown).

We further analyzed the onset of apoptosis using a commercially available solid-phase sandwich enzyme immunoassay that detects the caspase cleavage of cytokeratin-18 as a sensitive and specific marker of early canonical apoptosis [19]. Treatment with 100 nM Zalypsis® induced a marked increase of the apoptosis-related CK-18 Asp396 epitope in A549 cells. Moreover, simultaneous treatment with the pan-caspase inhibitor Z-VAD-fmk strongly reversed this effect (Fig. 3C), which clearly indicates that the pro-apoptotic effect of Zalypsis® was dependent on caspase activity.

Finally, we studied the caspase-dependent cleavage of the nuclear protein PARP as a marker of canonical apoptosis. The p85 band, generated through proteolytic cleavage of p116 PARP by activated caspases, was clearly visible as early as 6 h after the treatment of MCF7 and A549 cells with 100 nM Zalypsis® (Fig. 3D). After 48 h of treatment a considerable proportion of the PARP protein had already been cleaved by caspases, indicating that most of the cells were undergoing canonical apoptosis.

3.7. Zalypsis® shows a potent cytotoxic activity

The *in vitro* cytotoxicity of Zalypsis® was determined using a panel of 24 tumor cell lines that represent 11 relevant types of human cancer (Table 2). Most of the cell lines were very sensitive to Zalypsis® treatment, with IC_{50} values in the low nanomolar range. The panel average IC_{50} was ~7 nM, with MOLT-4 (acute lymphoblastic leukemia) cells showing the lowest IC_{50} and CAKI-1 (renal carcinoma) and HCT116 (colon carcinoma) cells presenting the highest IC_{50}s in the mid-nanomolar range.

Table 2
Cytotoxicity of Zalypsis® in a panel of 24 human cancer cell lines[a].

	Cell line	IC_{50} (nM)	Tissue average (nM)
Prostate	PC3	2.4 ± 1.1	2.1 ± 0.4
	22RV1	1.7 ± 0.8	
Pancreas	PANC-1	4.1 ± 0.8	2.7 ± 1.9
	MiaPaCa-2	1.3 ± 0.4	
Ovary	IGROV-1	3.5 ± 0.1	2.8 ± 0.9
	A2780	2.1 ± 1.3	
Lung	NCI-H460	3 ± 1.2	2.5 ± 0.4
	NCI-H23	2.2 ± 0.3	
	A549	2.4 ± 0.8	
Liver	SK-HEP-1	2.5 ± 0.8	2.5 ± 0.1
	HEPG2	2.6 ± 1.5	
Leukemia	MOLT4	0.5 ± 0.2	1 ± 0.7
	K562	1.6 ± 1.5	
Kidney	RXF393	2.1 ± 0.5	21.9 ± 27
	CAKI-1	41 ± 5	
Stomach	HS746T	1.5 ± 0.1	1.6 ± 0.07
	HGC-27	1.6 ± 0.5	
Colon	LoVo	2.4 ± 0.8	23 ± 34
	HT29	3.8 ± 0.2	
	HCT-116	62 ± 30	
Bladder	SW780	10 ± 0	10
Breast	MDA-MB-231	4.8 ± 0.8	4.1 ± 0.9
	MCF-7	4.5 ± 0.1	
	BT-474	3 ± 1.2	

[a] Values represent mean ± SD of three different experiments.

3.8. Zalypsis® exhibits in vivo antitumor activity in xenograft murine models

We performed xenograft studies to test whether the cytotoxicity of Zalypsis® translates into *in vivo* antitumor activity. 22RV1 (prostate), SW780 (bladder), Hs746t (gastric), and MiaPaCa-2 (pancreas) cells were xenografted into the right flank of athymic nu/nu mice. Once the tumors reached ca. 150 mm^3, the mice were randomized into groups of 10 and Zalypsis® was intravenously administered either in 3 consecutive weekly doses (0.9 mg/kg/day) (prostate and bladder) or in 2 cycles of 5 consecutive daily doses (0.3 mg/kg/day) (gastric and pancreas). Control animals received an equal volume of vehicle. At the drug doses used in the experiment, no significant toxicity or body weight loss was observed in the treated animals. Tumors in mice treated with Zalypsis® as a single agent strongly reduced their proliferation over the assay period, which finished when paired vehicle-treated animals had to be sacrificed (Fig. 4). Tumor growth reduction ranged from 69% in the Hs746t gastric cancer cell line to 42% in the MiaPaCa-2 pancreas cancer cell line. Tumor growth reductions in the 22RV1 (human prostate cancer), and SW780 (human bladder cancer) xenografts treated with Zalypsis® were 63% and 55%, respectively. The differences in tumor volumes between the vehicle- and the Zalypsis®-treated cohorts were statistically significant ($p < 0.05$) from Day 16 (22RV1, prostate), Day 24 (SW780, bladder), Day 19 (Hs746t, gastric) and Day 22 (MiaPaCa-2, pancreas) until the end of the experiments (Fig. 4).

4. Discussion

In the present work we introduce Zalypsis®, a new synthetic tetrahydroisoquinoline alkaloid that induces DNA damage and apoptosis in a variety of cancer cell lines in the low nanomolar range and shows a potent antitumor activity in several mouse xenograft models.

We first investigated whether Zalypsis® was able to bind either naked or methylated DNA through band shift assays, a methodology that has been previously used for similar analysis with other alkylating agents [13,15,20]. Zalypsis® was able to bind to both a naked PCR product of 250 bp with multiple guanines spread over its whole length and an oligonucleotide with a single binding site. This binding was dependent on the presence of the reactive carbinolamine moiety in the drug since ET-745, a related compound without this reactive group, did not produce any band shifts at any of the concentrations tested. Furthermore, Zalypsis® was able to bind to guanines that are preceded by a methylated cytosine (CpG natural methylation sites), demonstrating that methylation in the major groove does not affect binding of the drug to the minor groove.

Structurally, Zalypsis® belongs to a family of compounds that bind to the exocyclic amino group of guanines in the DNA double helix through an iminium intermediate generated *in situ* by dehydration of the carbinolamine moiety present in the A-ring [4,6–8]. The resulting adduct is additionally stabilized through van der Waals interactions and the establishment of one or more hydrogen bonds with surrounding nucleotides [8,9,21]. In fact, hydrogen bonding rules seem to determine the binding sequence specificity of the tetrahydroisoquinoline family of drugs [9,15]. Considering the diversity of regulatory mechanisms that rely on specific DNA sequences, it can be speculated that sequence selectivity may drive therapeutic specificity for the different DNA-binding compounds. Thus, while guanines are the target of many agents that alkylate DNA, anthramycin selects AGA and AGG triplets [22]; saframycins prefer GGG or GGC [23]; trabectedin favors TGG, CGG, AGC and GGC (while CGA is completely refractory, unlike the *N12*-demethylated analogue ET-729 [15]),

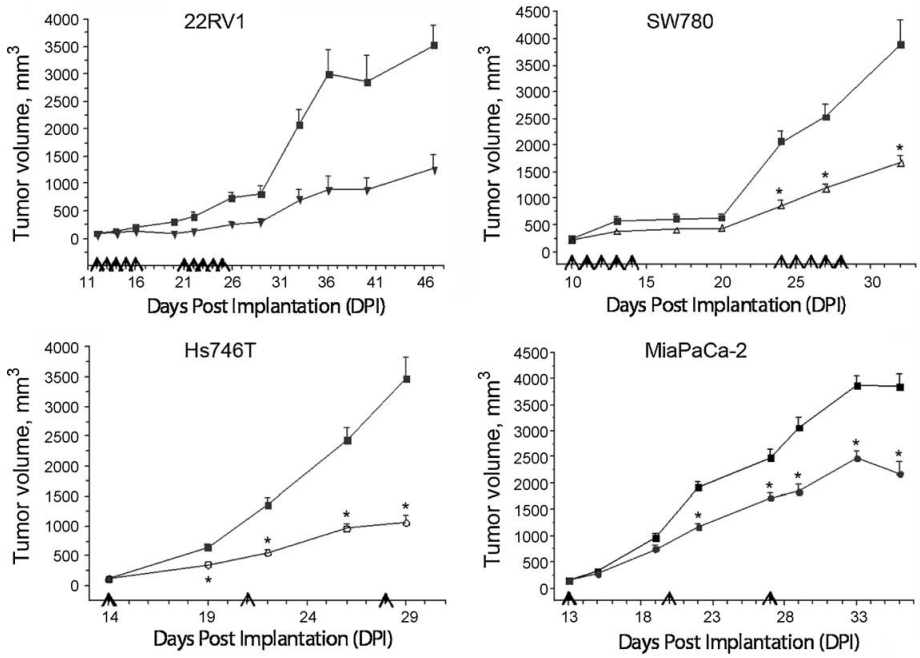


Fig. 4. *In vivo* antitumor activity of Zalypsis®. Treatment with Zalypsis® results in statistically significant ($^{*}p < 0.05$) tumor growth inhibition in 22RV1 prostate (▼), SW780 bladder (△), Hs746t gastric (○), and MiaPaCa-2 pancreas (●) xenograft models compared to vehicle (■) treated animals. Treatments started at a tumor volume size of ca. 150 mm^3 and were intravenously administered either in 2 cycles of 5 consecutive daily doses at 0.3 mg/kg (prostate and bladder) or 3 consecutive weekly doses at 0.9 mg/kg (gastric and pancreas). Each point represents mean ± SEM value ($n = 10$). Arrows indicate dosing days.

and Zalypsis® is shown here to have a preference for GGC, AGC, AGG, TGG and CGG. Whether these subtle differences in sequence selectivity may eventually generate differences in activity against diverse tumor types, is presently not known. On the other hand, the average DNA stabilization brought about on the double helix by drug binding (ΔT_m (max)) was different for Zalypsis® (~10 °C) compared to Yondelis® (~20 °C) (Supplementary Figure 1) [10] although it is still as high as that caused by a tight-binding bisintercalating agent such as echinomycin or thiocoraline [14]. The differences in reactivity and stabilization found for the rather comprehensive collection of alternative DNA triplets containing a central guanine could be rationalized on the basis of molecular models that showed the importance of specific hydrogen-bonding interactions involving drug and DNA atoms in the minor groove for attaining a suitable geometry for nucleophilic attack and covalent adduct formation.

We then verified that Zalypsis® induces DSBs in human cancer cells other than those from multiple myeloma for which γ-H2AX foci formation following exposure to this compound has been recently reported [24]. Indeed, our data show that drug treatment at nanomolar concentrations for 6 h gives rise to a high proportion of positive cells with abundant γ-H2AX foci per cell, as well as comets with a considerable percentage of DNA in the tail. Taken together, these findings most likely indicate that the Zalypsis®–DNA adducts are eventually transformed into DSBs in replicating cells. This is not surprising in light of similar findings recently reported for Yondelis® [10] but there still might be some differences in the way both compounds induce these DSBs. In fact, it has already been demonstrated that in Yondelis®-treated cells, stalling of the transcription-coupled NER (TC-NER) machinery at the adduct position is largely responsible for the toxicity of the compound [25,26]. In this action, the C-ring present in Yondelis®, which protrudes from the DNA minor groove, has been proposed to be directly involved in a hydrogen-bonding interaction with a key basic residue of the NER endonuclease XPG [27]. Since the ring in Zalypsis® that sticks out of the DNA minor groove is very different, this type of interaction may not take place and the repair complex can be stalled at a different step or even proceed further allowing in this case the effective repair of the DNA adducts. Preliminary data support this view as fission yeast mutants defective in the XPG orthologue Rad13 have been shown to be partially resistant to Yondelis® but not to Zalypsis® [28]. We can then envisage the possibility that Zalypsis® behaves as other alkylating/crosslinking agents giving rise to DNA adducts that escape repair by the NER machinery and eventually lead to replication fork stalls in the S-phase, followed by translesion synthesis (TLS) and generation of DSBs (reviewed in [29]). Moreover, the recent finding that Zalypsis® inhibits transcription with a higher potency than Yondelis® [11,12] may indicate that this latter drug principally targets TC-NER-dependent DNA repair whereas Zalypsis® blocks transcription through a stalled initiation or early elongation complex.

In this regard, the data presented here show that Zalypsis® induces an S-phase arrest at both low and high nanomolar concentrations. During S-phase, cells are continuously checking the integrity of their DNA to ensure the accuracy of the copying process. If any alteration is found, there are two safety mechanisms (DNA damage and DNA replication checkpoints) that stop S-phase progression and coordinate the repair of damaged DNA [29–31]. Thus, our data suggest that DNA damage induced by this drug arrests cells in S-phase and activates the DNA replication checkpoints. This scenario is different from that obtained in multiple myeloma cells where Zalypsis® treatment was described to induce an increased in the G_0/G_1 phase [24]. Whether these

effects are cell line- or disease-dependent is presently not known. In our experiments, the S-phase accumulation was higher at 15 nM than at 150 nM. Probably, this is due to a higher cell survival rate at the lower concentration. In fact, although some of the cells treated with high concentrations of Zalypsis® would still accumulate in S-phase, most of them would stop in the cell cycle phase in which they originally were at the onset of drug treatment, due to the high amount of DNA damage that they endure, and they would die with a slower kinetics. Our data also show that Zalypsis®-induced cell death presents the characteristics of apoptosis, with chromatin condensation, dependence on caspase activity and cleavage of the caspase targets PARP-1 and cytokeratin-18. This is in accordance with data obtained in multiple myeloma cell lines on which Zalypsis® treatment induced a strong apoptotic response that was partially dependent on caspase activation [24].

When assayed in a panel of 24 cancer cell lines, Zalypsis® demonstrated a potent cytotoxic activity, with an average IC_{50} of 7 nM in most cell lines, with the notable exception of HCT116 and CAKI-1, both of which are substantially less sensitive to the drug. Myeloma cell lines have also been recently shown to be highly sensitive to Zalypsis®, with IC_{50} values ranging from picomolar to low nanomolar concentrations [24]. Other alkaloids of the tetrahydroisoquinoline family, including jorumycin, renieramycins and saframycins, have been previously shown to possess strong cytotoxic activity, with IC_{50}s also in the low nanomolar range [3,32,33]. According to our data, Zalypsis® exhibits higher cytotoxic activity against leukemia and stomach tumor cell lines, with breast cancer cell lines being less sensitive than the average. This tissue/cell-type sensitivity pattern is clearly different from that observed for Yondelis® as this ecteinascidin shows selectivity against sarcoma, breast, ovary and lung tumor cell lines, with head-and-neck and colon tumor cell lines being slightly less sensitive [34,35].

Finally, the *in vitro* cytotoxicity of Zalypsis® translated well into antitumor activity *in vivo* in four xenograft models of human cancer. The models analyzed corresponded to cell lines ranging from slightly more sensitive (Hs746T, gastric cancer) to substantially less sensitive (SW780, bladder cancer) than the average in the *in vitro* cytotoxicity assays. Zalypsis® demonstrated a clear antitumor activity that was highest for the Hs746t gastric cancer model, with a tumor growth reduction of 69%. The pancreatic cancer model MiaPaCa was the least sensitive to Zalypsis® but, even in this case, the compound achieved a 42% tumor growth reduction. When we assayed the maximum tolerated dose of the drug in two different administration regimes no significant differences among them and no toxic effects to the animals were observed.

In summary, we have demonstrated highly potent *in vitro* and *in vivo* anticancer activities of Zalypsis® and gained some insight into its molecular mechanism of action. This agent exerts its anticancer effects through the covalent modification of guanines in the DNA minor groove that eventually give rise to DNA double-strand breaks, S-phase arrest and apoptosis in cancer cells. These encouraging results strongly support its development as a novel anticancer agent with a view to a wide spectrum of clinical settings.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bcp.2009.04.003.

References

[1] Fontana A, Cavaliere P, Wahidullah S, Naik CG, Cimino G. A new antitumor isoquinoline alkaloid from the marine Nudibranch *Jorunna funebris*. Tetrahedron 2000;56:7305–8.

[2] Frinke JM, Faulkner DJ. Antimicrobial metabolites of the sponge *Reniera sp*.. J Am Chem Soc 1982;104:265–9.

[3] Oku N, Matsunaga S, van Soest RW, Fusetani N. Renieramycin J: a highly cytotoxic tetrahydroisoquinoline alkaloid, from a marine sponge *Neopetrosia sp*.. J Nat Prod 2003;66:1136–9.

[4] Kishi K, Yazawa K, Takahashi K, Mikami Y, Arai T. Structure–activity relationships of saframycins. J Antibiot 1984;37:847–52.

[5] Rinehart K, Holt TG, Fregeau NL, Stroh JG, Keifer PA, Sun F, et al. Ecteinascidins 729, 743, 745, 759A, 759B and 770: potent antitumor agents from the Caribbean tunicate *Ecteinascidia turbinate*. J Org Chem 1990;55:4512–5.

[6] Guan Y, Sakai R, Rinehart K, Wang AH. Molecular and crystal structures of Ecteinascidins: potent antitumor compounds from the Caribbean tunicate *Ecteinascidia turbinata*. J Biomol Struct Dyn 1993;10:793–818.

[7] Pommier Y, Kohlhagen G, Bailly C, Waring M, Mazumder A, Kohn KW. DNA sequence- and structure-selective alkylation of guanine N2 in the DNA minor groove by ecteinascidin 743, a potent antitumor compound from the Caribbean tunicate *Ecteinascidia turbinata*. Biochemistry 1996;35:13303–9.

[8] Moore BM, Seaman FC, Hurley LH. NMR-based model of an Ecteinascidin 743–DNA adduct. J Am Chem Soc 1997;119:5475–6.

[9] Seaman FC, Hurley LH. Molecular basis for the DNA sequence selectivity of ecteinascidin 736 and 743: evidence for a dominant role of direct readout via hydrogen bonding. J Am Chem Soc 1998;120:13028–41.

[10] Casado JA, Rio P, Marco E, Garcia-Hernandez V, Domingo A, Perez L, et al. Relevance of the Fanconi anemia pathway in the response of human cells to trabectedin. Mol Cancer Ther 2008;7:1309–18.

[11] Magro P, Alvarez E, Kolb E, Mandola M, Scotto KW. Characterization of tumor cell response to the novel anti-tumor agent Zalypsis®. In: 98th AACR annual meeting; 2007. p. 1529.

[12] Feuerhahn S. Mechanistic insights into the effect of oxidative lesions and anticancer drugs on RNA pol II transcription and transcription coupled repair. In: Unité 596 Inserm "Institut de génétique et de biologie moléculaire et cellulaire", Université Louis Pasteur Strasbourg I, Strasbourg; 2008. p. 148.

[13] David-Cordonnier MH, Gajate C, Olmea O, Laine W, de la Iglesia-Vicente J, Perez C, et al. DNA and non-DNA targets in the mechanism of action of the antitumor drug trabectedin. Chem Biol 2005;12:1201–10.

[14] Negri A, Marco E, Garcia-Hernandez V, Domingo A, Llamas-Saiz AL, Porto-Sanda S, et al. Antitumor activity, X-ray crystal structure, and DNA binding properties of thiocoraline A, a natural bisintercalating thiodepsipeptide. J Med Chem 2007;50:3322–33.

[15] Marco E, David-Cordonnier MH, Bailly C, Cuevas C, Gago F. Further insight into the DNA recognition mechanism of trabectedin from the differential affinity of its demethylated analogue ecteinascidin ET729 for the triplet DNA binding site CGA. J Med Chem 2006;49:6925–9.

[16] Redon C, Pilch D, Rogakou E, Sedelnikova O, Newrock K, Bonner W. Histone H2A variants H2AX and H2AZ. Curr Opin Genet Dev 2002;12:162–9.

[17] Guirouilh-Barbat J, Redon C, Pommier Y. Transcription-coupled DNA double-strand breaks are mediated via the nucleotide excision repair and the Mre11-Rad50-Nbs1 complex. Mol Biol Cell 2008;19:3969–81.

[18] Norbury CJ, Zhivotovsky B. DNA damage-induced apoptosis. Oncogene 2004;23:2797–808.

[19] Kramer G, Erdal H, Mertens HJ, Nap M, Mauermann J, Steiner G, et al. Differentiation between cell death modes using measurements of different soluble forms of extracellular cytokeratin 18. Cancer Res 2004;64:1751–6.

[20] David-Cordonnier MH, Laine W, Lansiaux A, Rosu F, Colson P, de Pauw E, et al. Covalent binding of antitumor benzoacronycines to double-stranded DNA induces helix opening and the formation of single-stranded DNA: unique consequences of a novel DNA-bonding mechanism. Mol Cancer Ther 2005;4:71–80.

[21] Garcia-Nieto R, Manzanares I, Cuevas C, Gago F. Bending of DNA upon Binding of ecteinascidin 743 and phthalascidin 650 studied by unrestrained molecular dynamics simulations. J Am Chem Soc 2000;122:7172–82.

[22] Pierce JR, Nazimiec M, Tang MS. Comparison of sequence preference of tomaymycin- and anthramycin–DNA bonding by exonuclease III and lambda exonuclease digestion and UvrABC nuclease incision analysis. Biochemistry 1993;32:7069–78.

[23] Rao KE, Lown JW. DNA sequence selectivities in the covalent bonding of antibiotic saframycins Mx1, Mx3, A, and S deduced from MPE.Fe(II) footprinting and exonuclease III stop assays. Biochemistry 1992;31:12076–82.

[24] Ocio EM, Maiso P, Chen X, Garayoa M, Alvarez-Fernandez S, San-Segundo L, et al. Zalypsis: a novel marine-derived compound with potent antimyeloma activity that reveals high sensitivity of malignant plasma cells to DNA double strand breaks. Blood 2008;113:3781–91.

[25] Erba E, Bergamaschi D, Bassano L, Damia G, Ronzoni S, Faircloth GT, et al. Ecteinascidin-743 (ET-743), a natural marine compound, with a unique mechanism of action. Eur J Cancer 2001;37:97–105.

[26] Takebayashi Y, Pourquier P, Zimonjic DB, Nakayama K, Emmert S, Ueda T, et al. Antiproliferative activity of ecteinascidin 743 is dependent upon transcription-coupled nucleotide-excision repair. Nat Med 2001;7(8):961–6.

[27] Herrero AB, Martin-Castellanos C, Marco E, Gago F, Moreno S. Cross-talk between nucleotide excision and homologous recombination DNA repair pathways in the mechanism of action of antitumor trabectedin. Cancer Res 2006;66:8155–62.

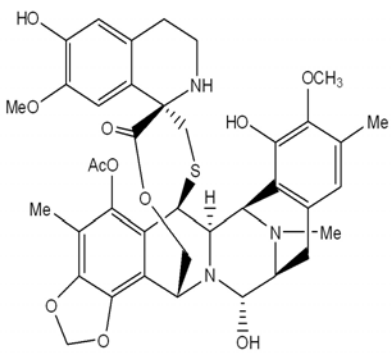
[28] Herrero A, Marco E, García-Hernández V, Martín-Castellanos C, Domingo A, Alvarez E, et al. Resilience of the cytotoxic effects of Zalypsis® (pm00104) to the lack of a functional nucleotide excision repair system. In: American association for cancer research annual meeting; 2007. p. A3470.

[29] Andreassen PR, Ho GP, D'Andrea AD. DNA damage responses and their many interactions with the replication fork. Carcinogenesis 2006;27:883–92.

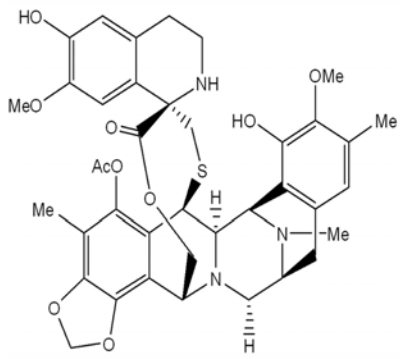
[30] Zhou BB, Elledge SJ. The DNA damage response: putting checkpoints in perspective. Nature 2000;408:433–9.

[31] Hakem R. DNA-damage repair: the good, the bad, and the ugly. Embo J 2008;27:589–605.
[32] Arai T, Takahashi K, Nakahara S, Kubo A. The structure of a novel antitumor antibiotic, saframycin A. Experientia 1980;36:1025–7.
[33] Lane JW, Estevez A, Mortara K, Callan O, Spencer JR, Williams RM. Antitumor activity of tetrahydroisoquinoline analogues 3-epi-jorumycin and 3-epi-renieramycin G. Bioorg Med Chem Lett 2006;16:3180–3.
[34] Poindessous V, Koeppel F, Raymond E, Comisso M, Waters SJ, Larsen AK. Marked activity of irofulven toward human carcinoma cells: comparison with cisplatin and ecteinascidin. Clin Cancer Res 2003;9:2817–25.
[35] Koeppel F, Poindessous V, Lazar V, Raymond E, Sarasin A, Larsen AK. Irofulven cytotoxicity depends on transcription-coupled nucleotide excision repair and is correlated with XPG expression in solid tumor cells. Clin Cancer Res 2004;10:5604–13.

Supplementary Figure 1 – Leal et al



ET-743



ET-745

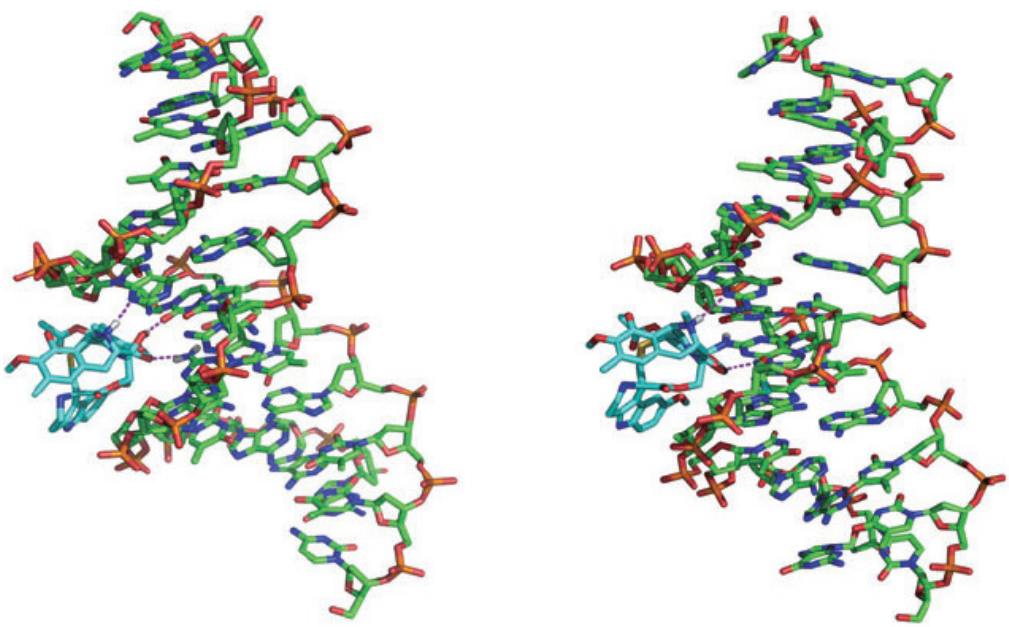
Supplementary Table 1. Sequence of the ds-Oligonucleotides used in the DNA melting assay

Triplet	Oligonucleotide Sequences
A**G**C	5'-F-AATATAAAGCAAATTAA-3'
T**G**C	3'-Q-TTATATTTCGTTTAATT-5'
T**G**G	5'-F-AATATAAACCAAATTAA-3'
G**G**T	3'-Q-TTATATTTGGTTTAATT-5'
A**G**A	5'-F-AATATAAAGAAATTAA-3' 3'-Q-TTATATTTCTTTAATT-5'
C**G**A	5'-F-AATATAAACGAAATTAA-3'
C**G**T	3'-Q-TTATATTTGCTTTAATT-5'
A**G**G	5'-F-AATATAAAGGCAAATTAA-3'
G**G**C T**G**C	3'-Q-TTATATTTCCGTTTAATT-5'
C**G**G	5'-F-AATATAAACGGAAATTAA-3'
G**G**A C**G**T	3'-Q-TTATATTTGCCTTTAATT-5'
C**G**T	5'-F-AATATAAAACGTTAAATTAA-3' 3'-Q-TTATATTTTGCAATTTAATT-5'
A**G**C	5'-F-AATATAAAGCGTTAAATTAA-3'
C**G**T C**G**C	3'-Q-TTATATTTCGCAATTTAATT-5'
A**G**G	5'-F-AATATAAAGGAAATTAA-3'
G**G**A	3'-Q-TTATATTTCCTTTAATT-5'
A**G**T	5'-F-AATATAAAGTAAATAA-3' 3'-Q-TTATATTTCATTTATT-5'
T**G**A	5'-F-AATATAATGAAAATTAA-3' 3'-Q-TTATATTACTTTTAATT-5'
T**G**C	5'-F-AATATAATGCAATTAA-3' 3'-Q-TTATATTACGTTAATT-5'
T**G**T	5'-F-AATATAAACAAATTAA-3' 3'-Q-TTATATTTGTTTAATT-5'
A**G**G	5'-F-AATATAAGGTAAATTAA-3'
G**G**T	3'-Q-TTATATTCCATTTAATT-5'
A**G**G	5'-F-AATATAAGGGAAATTAA-3'
G**G**G G**G**A	3'-Q-TTATATTCCCTTTAATT-5'
(AT)	5'-F-CAATTAAATATAAC-3' 3'-Q-GTTAATTTATATTG-5'

Artículo II

"PM01183, a new DNA minor groove covalent binder with potent in vitro and in vivo anti-tumour activity"

***Br J Pharmacol*. 2010 Nov;161(5):1099-110**



ANTECEDENTES Y OBJETIVOS: PM01183 es un nuevo alcaloide tetrahidroisoquinolínico sintético que está actualmente en fase I de desarrollo clínico para el tratamiento de tumores sólidos. En este estudio se han caracterizado las interacciones del PM01183 con determinadas secuencias y su citotoxicidad in vitro e in vivo.

ENFOQUE EXPERIMENTAL: las características de unión al ADN de PM01183 se estudiaron mediante ensayos de movilidad por electroforesis, experimentos cinéticos de fusión basados en fluorescencia y métodos de modelado computacional. Su mecanismo de acción se investigó mediante citometría de flujo, análisis de Western blot y microscopía de fluorescencia. In vitro, la actividad anti-tumoral se determinó por ensayo con bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolio y para medir su actividad in vivo se utilizaron varios modelos de cáncer humano.

PRINCIPALES RESULTADOS: los ensayos de movilidad electroforética demostraron que PM01183 se une al ADN. Los experimentos de desnaturalización térmica basados en fluorescencia demostraron que los tripletes de ADN más favorables para la formación de aductos covalentes son CAG, CGG, AGG y TGG. Estas preferencias de unión pudieron racionalizarse con el uso de modelado molecular.

CONCLUSIONES Y CONSECUENCIAS: Se demostró que PM01183 se une a determinadas secuencias de ADN y promueve la apoptosis mediante la inducción de roturas de doble cadena en concentraciones nanomolares. La potente actividad antitumoral de PM01183 en modelos murinos varios de cáncer humano apoya su desarrollo como un nuevo agente antineoplásico

BJP BRITISH JOURNAL OF PHARMACOLOGY
BPS BRITISH PHARMACOLOGICAL SOCIETY

BJP British Journal of Pharmacology

DOI:10.1111/j.1476-5381.2010.00945.x
www.brjpharmacol.org

RESEARCH PAPER

PM01183, a new DNA minor groove covalent binder with potent *in vitro* and *in vivo* anti-tumour activity

JFM Leal[1], M Martínez-Díez[1], V García-Hernández[2], V Moneo[1], A Domingo[2], JA Bueren-Calabuig[3], A Negri[3], F Gago[3], MJ Guillén-Navarro[1], P Avilés[1], C Cuevas[1], LF García-Fernández[1] and CM Galmarini[1]

[1]*Cell Biology Department, PharmaMar SA, Spain,* [2]*Department of Biochemistry and Molecular Biology, University of Alcalá, Spain, and* [3]*Department of Pharmacology, University of Alcalá, Spain*

Correspondence
Carlos M. Galmarini, PharmaMar Cell Biology Department, Avda. de los Reyes, 1, 28770 Colmenar Viejo, Madrid, Spain. E-mail: cgalmarini@pharmamar.com

Keywords
antineoplastic agents; tetrahydroisoquinolines; DNA breaks; apoptosis

Received
15 January 2010
Revised
12 April 2010
Accepted
8 June 2010

BACKGROUND AND PURPOSE
PM01183 is a new synthetic tetrahydroisoquinoline alkaloid that is currently in phase I clinical development for the treatment of solid tumours. In this study we have characterized the interactions of PM01183 with selected DNA molecules of defined sequence and its *in vitro* and *in vivo* cytotoxicity.

EXPERIMENTAL APPROACH
DNA binding characteristics of PM01183 were studied using electrophoretic mobility shift assays, fluorescence-based melting kinetic experiments and computational modelling methods. Its mechanism of action was investigated using flow cytometry, Western blot analysis and fluorescent microscopy. *In vitro* anti-tumour activity was determined by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay and the *in vivo* activity utilized several human cancer models.

KEY RESULTS
Electrophoretic mobility shift assays demonstrated that PM01183 bound to DNA. Fluorescence-based thermal denaturation experiments showed that the most favourable DNA triplets providing a central guanine for covalent adduct formation are AGC, CGG, AGG and TGG. These binding preferences could be rationalized using molecular modelling. PM01183–DNA adducts in living cells give rise to double-strand breaks, triggering S-phase accumulation and apoptosis. The potent cytotoxic activity of PM01183 was ascertained in a 23-cell line panel with a mean GI_{50} value of 2.7 nM. In four murine xenograft models of human cancer, PM01183 inhibited tumour growth significantly with no weight loss of treated animals.

CONCLUSIONS AND IMPLICATIONS
PM01183 is shown to bind to selected DNA sequences and promoted apoptosis by inducing double-strand breaks at nanomolar concentrations. The potent anti-tumour activity of PM01183 in several murine models of human cancer supports its development as a novel anti-neoplastic agent.

Abbreviations
CK-18, cytokeratin-18; DAPI, 4′,6-diamidino-2-phenylindole; DMSO, dimethyl sulphoxide; DSBs, double-strand breaks; ICL, interstrand crosslink; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Pol II, polymerase II; TC-NER, transcription-coupled-nucleotide excision repair

Introduction

DNA is a well-characterized intracellular target in cancer chemotherapy (Reddy *et al.*, 2001) and consists of two right-handed polynucleotide chains running in opposite directions to form a double helix. This DNA duplex is stabilized by both intra-strand stacking interactions and hydrogen bonds

© 2010 PharmaMar SA
British Journal of Pharmacology © 2010 The British Pharmacological Society

between paired bases in the two strands: adenine forms two hydrogen bonds with thymine while guanine forms three hydrogen bonds with cytosine. The base pairs are rotated by 36° with respect to each adjacent pair. This structure gives rise to two well-defined clefts known as the major and minor grooves. The major groove is deep and wide while the minor groove is narrow and shallow. As a result, DNA-binding proteins and drugs usually make contacts to the sides of the bases exposed in both the major (Pabo and Sauer, 1984) and the minor grooves (Rohs *et al.*, 2009). The DNA major groove represents a site of attack for cisplatin and many alkylating agents whereas other anti-tumour drugs such as mitomycin C, chromomycin A_3 or the ecteinascidins bind to the minor groove.

Currently, there is a renewed interest in the development of molecules that target the minor groove of DNA (Susbielle *et al.*, 2005). One of the best examples is trabectedin (Yondelis®), a member of the ecteinascidin family originally derived from the marine tunicate *Ecteinascidia turbinata* that is currently used for the treatment of patients with advanced or metastatic soft tissue sarcoma and relapsed platinum-sensitive ovarian cancer. Trabectedin reacts with the exocyclic amino group of certain guanines in the minor groove of DNA (Kishi *et al.*, 1984; Guan *et al.*, 1993; Pommier *et al.*, 1996; Moore *et al.*, 1997) forming a covalent bond. The resulting adduct is additionally stabilized through the establishment of van der Waals interactions and one or more hydrogen bonds with neighbouring nucleotides in the opposite strand of the DNA double helix (Moore *et al.*, 1997; Seaman and Hurley, 1998), thus creating the equivalent to a functional interstrand crosslink (Casado *et al.*, 2008).

PM01183 (Figure 1A) is a new synthetic alkaloid structurally related to ecteinascidins (Rinehart *et al.*, 1990; Manzanares *et al.*, 2001) that is currently in phase I clinical development for the treatment of solid tumours. In common with trabectedin, PM01183 contains a pentacyclic skeleton composed of two fused tetrahydroisoquinoline rings (subunits A and B) that is mostly responsible for DNA recognition and binding but the additional module (ring C) in PM01183 is a tetrahydro β-carboline rather than the additional tetrahydroisoquinoline present in trabectedin. This structural difference may confer pharmacokinetic benefits as well as intrinsic activity. In PM01183 the C-ring, which protrudes from the DNA minor groove, has been proposed to be involved in direct interactions with specific factors of DNA repair pathways (Herrero *et al.*, 2006; Casado *et al.*, 2008). In this study we have characterized: (i) the molecular interactions between PM11083 and several DNA oligo- and polynucleotides; and (ii) the *in vitro* and *in vivo* cytotoxic activity of the compound.

The DNA binding characteristics were studied using a combination of electrophoretic mobility shift assays, fluorescence-based melting kinetic experiments and computational modelling methods. We also looked at the type of DNA damage subsequently generated in living cells as a consequence of PM01183-DNA adduct formation as well as cell cycle perturbations and type of cell death induced by this compound. Finally, the *in vitro* and *in vivo* anti-tumour activity of PM01183 was investigated using several models of human cancers.

Methods

DNA electrophoretic mobility shift assay

The binding assay was performed with a 250 bp PCR product from the human adiponectin gene (Sigma, St Louis, MO, USA). After incubation with appropriate concentrations of the compound at 25°C during 1 h, the DNA was subjected to electrophoresis in a 2% (w/v) agarose/TAE gel, stained with 1 $\mu g \cdot mL^{-1}$ ethidium bromide and photographed.

DNA melting assay

Synthetic oligodeoxynucleotides with one strand 5′-end-labelled with the fluorophore 6-carboxyfluorescein and the complementary strand 3′-end-labelled with the quencher tetramethylrhodamine were synthesized at Bonsai Technologies (Madrid, Spain) (Table S1). For the experiments, we followed the methodology previously described (Negri *et al.*, 2007; Casado *et al.*, 2008) using a 7500 Fast Real-Time PCR System (ABI Prism, Applied Biosystems, Foster City, CA, USA). Analyses of the raw data obtained for each oligodeoxynucleotide using an in-house developed Visual Basic Application running on Microsoft Excel (Microsoft, Redmond, WA, USA) yielded estimates of both the increase in melting temperature (ΔT_m) brought about by drug binding, relative to the free DNA molecule, and the ligand concentration that produces half the maximal change in melting temperature (C_{50}). The inverse of this latter value ($1/C_{50}$) was taken as a measure of the relative DNA binding affinity, and the parameter $\Delta T_m(max)$ was used to reflect the relative stability of the DNA–ligand complexes (Negri *et al.*, 2007).

Computational methods

Molecular modelling and simulation methods were used to study in atomic detail: (i) the binding orientation required to activate dehydration of the

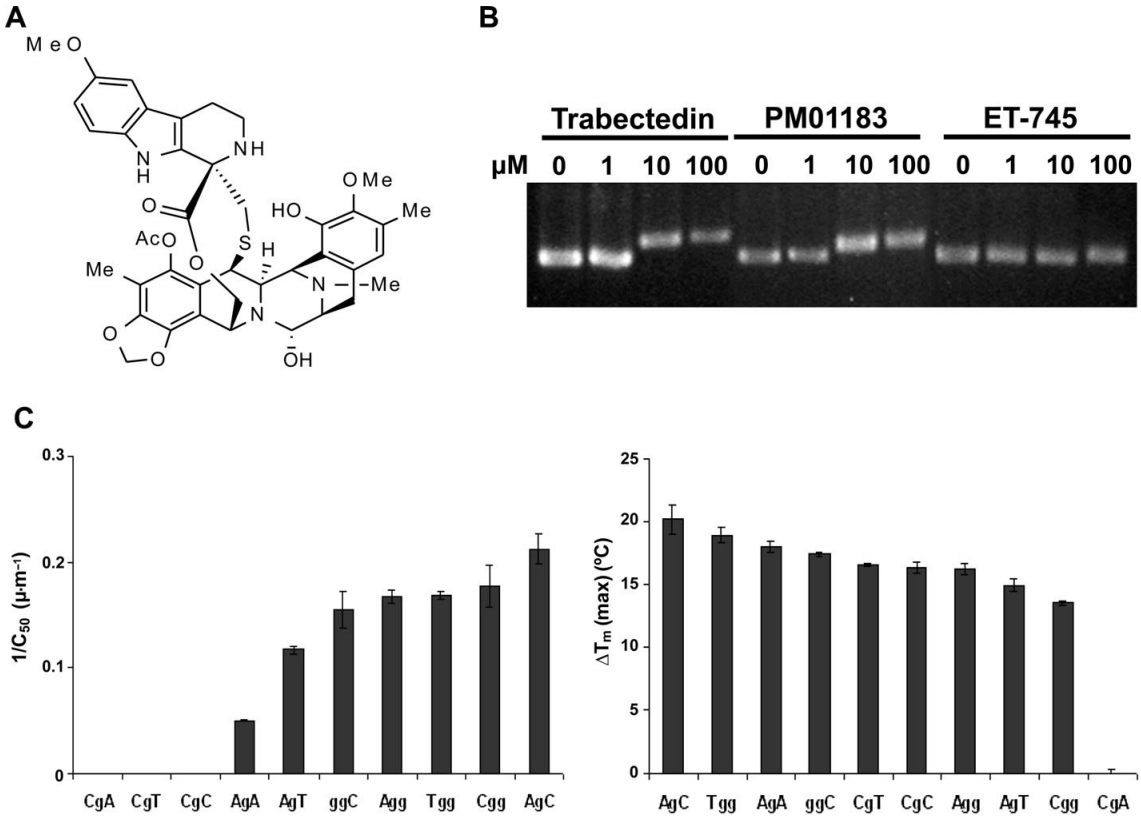


Figure 1

Chemical structure and DNA binding properties of PM01183. (A) Structure of PM01183; (B) binding to naked DNA. Drugs were incubated with a 250 bp PCR product at 25°C during 1 h and the electrophoresis was run in 2% agarose-TAE; (C) relative binding affinities for different DNA triplet sequences. PM01183 was incubated with the labelled double-strand oligodeoxynucleotides at 25°C during 1 h; then the melting assay was started in a 7500 Fast Real-Time PCR System by increasing the temperature up to 95°C in small steps of 1°C·min^{-1}. Analysis of the $1/C_{50}$ parameter (drug affinity for specific DNA sequences) and ΔT_m(max) (stability of the drug–DNA complex) were analysed with an in-house developed Visual Basic Application (VBA) running on Microsoft Excel. The values correspond to the mean and standard deviation from at least two independent experiments for each sequence.

carbinolamine and promote the nucleophilic attack by the amino group to a central guanine in a favoured DNA triplet; and (ii) the consequences of covalent adduct formation in the minor groove. To this end, a pre-covalent complex between PM01183 and an oligonucleotide of sequence 5′-d(CAATACG GATAAG)/5′-d(CTTATCCGTATTG), containing a central CGG as a target site, was built as previously reported for trabectedin (Marco *et al.*, 2006). The covalent complex with the underlined guanine was similarly built and both complexes were then refined in the presence of explicit solvent and counterions using energy minimization techniques in the AMBER force field (Case *et al.*, 2005). The stabilities and geometrical features of the complexes were studied by means of 20 ns of unrestrained molecular dynamics simulations using the same conditions as described previously.

Cell culture and cytotoxicity

All the tumour cell lines used in this study were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA). For the cytotoxicity experiments, cells were seeded in 96-well trays. Serial dilutions of PM01183 dissolved in dimethyl sulphoxide (DMSO) were prepared and added to the cells in fresh medium, in triplicate. Exposure to the drug was maintained during 72 h and cellular viability was estimated from conversion of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium

bromide (MTT) (Sigma, St Louis, MO, USA) to its coloured reaction product, MTT formazan, which was dissolved in DMSO so as to measure absorbance at 540 nm with a POLARStar Omega Reader (BMG Labtech, Offenburg, Germany). Determination of the concentration giving 50% growth inhibition (GI_{50}) values was performed by iterative non-linear curve fitting using the Prism 5.0 statistical software (GraphPad, La Jolla, CA, USA). The data presented are the average of three independent experiments performed in triplicate.

Fluorescent microscopy

Cells were treated with the appropriate concentration of PM01183 for 6 h, washed and then cultured for 18 additional hours. At the end of this period, they were fixed (4% paraformaldehyde), permeabilized (0.5% Triton X-100) and incubated with the primary anti-γ-H2AX monoclonal antibody (Upstate, Temecula, CA, USA) for 1 h at 37°C. Thereafter the cells were washed and incubated with the AlexaFluor 594 secondary goat anti-mouse IgG (Invitrogen, Carlsbad, CA, USA) for 30 min at 37°C. Finally, the slides were incubated with Hoechst 33 342 (Sigma, St Louis, MO, USA) and mounted with Mowiol mounting medium. Pictures were taken with a Leica DM IRM fluorescence microscope equipped with a 100× oil immersion objective and a DFC 340 FX digital camera (Leica, Wetzlar, Germany).

Cell cycle analysis

For the cell cycle experiments and sub-G_1 peak determination, cells were treated with the appropriate amount of PM01183 for 12 h and 24 h, and then stained with 0.4 $\mu g \cdot mL^{-1}$ propidium iodide. Samples were analysed with a FACScalibur flow cytometer (Beckton and Dickinson, Franklin Lakes, NJ, USA) and the FlowJo7 cytometry analysis software.

Apoptosis detection

For the chromatin condensation assay, cells were treated with the appropriate amount of the compound for 24 h and stained with 4′,6-diamidino-2-phenylindole (DAPI). Chromatin condensation in early apoptotic cells was assessed with a DM IRM microscope (Leica, Wetzlar, Germany). For the M30-Apoptosense® solid-phase sandwich enzyme immunoassay (Peviva, Bromma, Sweden), the kit instructions were followed. Briefly, cells were exposed to the appropriate concentration of the drug for 24 h. Cell lysates were obtained, transferred to an assay well and incubated for 4 h following addition of a mouse monoclonal (M30) antibody-horseradish peroxidase (HRP) conjugate. Once the antigen-M30 was bound to the M5-coated surface of the wells, they were extensively washed and incubated with the 3,3′,5,5′-tetramethyl-benzidine substrate in the dark for 20 min. Finally, the reaction was ended by addition of a stop solution and the plates were read at 450 nM within 30 min on a Victor3 platform (Perkin Elmer).

Western blot assays

For immunoblotting, cells were treated with the appropriate compound concentration for 24 h and lysed with RIPA lysis buffer. Protein content was determined by the modified Bradford method. Samples were separated in 7.5% SDS-PAGE, transferred onto an Immobilon-P membrane. Then, membranes were incubated with the appropriate primary antibody for 1 h, washed and incubated with the secondary antibody. Finally, protein was visualized using the ECL System (GE Healthcare, Fairfield, CT, USA). We used an anti-poly ADP ribose polymerase (PARP) rabbit polyclonal antibody from Santa Cruz (Santa Cruz, CA, USA), and an anti-α-tubulin monoclonal antibody from Sigma (St Louis, MO, USA). Secondary antibodies were HRP-conjugated goat anti-rabbit secondary antibody (R&D Minneapolis, MN, USA) and HRP-conjugated goat anti-mouse secondary antibody (Santa Cruz, CA, USA).

Anti-tumour activity in xenograft murine models

All animal care and experimental procedures were approved by the Institutional Animal Care and Use Committee of PharmaMar, Inc. (Cambridge, MA, USA), in an AAALAC accredited animal facility. Four to 6-week-old athymic *nu/nu* mice (Harlan Sprague Dawley, Madison, WI, USA) were s.c. xenografted with NCI-H460 (lung), A2780 (ovary), HT29 (colon) and HGC-27 (gastric) cells into their right flank with *c.* 3×10^6 cells in 0.2 mL of a mixture (50:50; v:v) of Matrigel basement membrane matrix (Beckton Dickinson, Frankin Lakes, NJ, USA) and serum-free medium. When tumours reached *c.* 150 mm^3, mice were randomly assigned to treatment or control groups. PM01183 was intravenously administered in three consecutive weekly doses (0.18 $mg \cdot kg^{-1} \cdot day^{-1}$) whereas the control animals received an equal volume of vehicle with the same schedule. Caliper measurements of the tumour diameters were made twice weekly and tumour volumes were calculated according to the following formula: $(a \cdot b)^2/2$, where a and b were the longest and shortest diameters respectively. Animals were humanely killed when their tumours reached 3000 mm^3 or if significant toxicity (e.g. severe body weight reduction) was observed. Differences in tumour volumes between treated and control groups were evaluated using the Mann–

Whitney *U*-test. Statistical analyses were performed by LabCat® v8.0 SP1 (Innovative Programming Associates, Inc., NJ, USA).

Materials

PM01183 is a PharmaMar SAU proprietary drug produced by chemical synthesis. ET-745 is a synthetic ecteinascidin that lacks the carbinolamine (hemiaminal) group (Manzanares *et al.*, 2001) and does not bind covalently to DNA (David-Cordonnier *et al.*, 2005). DAPI and Z-Vad-fmk (a pan-caspase inhibitor) were purchased from Sigma (St Louis, MO, USA) (Figure S1).

Results

DNA binding properties of PM01183

The mechanism of action of PM01183 was expected to rely on the covalent reaction with selected guanines in the DNA double helix. Thus, we first investigated its binding to DNA through band shift assays that allow the qualitative assessment of the molecular weight increase brought about in this macromolecule by the covalent bonding of PM01183 to the amino group of a central guanine in well-defined DNA triplets. As positive and negative controls we used, respectively, trabectedin and ET-745, a derivative unable to form a covalent adduct with DNA because it lacks the reactive hydroxyl group making up the hemiaminal in PM01183 (Figure S1, Manzanares *et al.*, 2001). Figure 1B shows the result of a typical band shift experiment that used increasing concentrations of PM01183 to saturate the 250 bp DNA probe with the compound and give rise to an increase in molecular weight. As expected, the delay in electrophoretic migration reflected that PM01183 does indeed bind to naked DNA. In these experimental conditions, DNA saturation required ~10 μM of the compound.

The ability of PM01183 to stabilize the double helix was assessed by means of several tailor-made fluorophore-labelled oligonucleotides encompassing most of the possible DNA triplet combinations involving a central guanine (Table S1). Figure 1C summarizes the results of the DNA melting experiments. The mean ± SD $1/C_{50}$ and ΔT_m(max) values for PM01183 were 0.15 ± 0.5 μM^{-1} and 15 ± 5°C respectively. According to the affinity indicator $1/C_{50}$, the most favourable triplets for PM01183 bonding to DNA appeared to be AGC, CGG, AGG and TGG triplets whereas the highest thermal stabilities were observed for DNA adducts involving an AGC, TGG, AGA and GGC site. The lowest affinities and stabilities were observed in oligonucleotides containing CGC, CGT and CGA sites.

Molecular modelling data

The general structure of the CGG-containing pre-alkylation binding complex shows PM01183 fitting snugly within the minor groove with subunit A protruding perpendicularly to the helical axis of the DNA molecule in front of the target guanine (G7), subunit B stacking over the sugar ring of the complementary cytosine (C20), and one side of subunit C making extensive contacts with the sugar ring of the nucleotide on the 3′ side of the central triplet. The N12 proton in subunit A of PM01183 plays a clear role in sequence recognition and also mediates the *in situ* dehydration of the hemiaminal to yield the reactive iminium intermediate, as reported previously for trabectedin (Garcia-Nieto *et al.*, 2000). Thus, in our simulation of this pre-covalent complex, the protonated N12 remains hydrogen bonded to the N3 acceptor atom of G21 whereas the OH acts both as a hydrogen bond donor to O2 of C20 and as a hydrogen bond acceptor from N2 of G7 (Figure 2A, left). Conversely, the methylendioxy oxygen from subunit B that faces the minor groove is also involved in a hydrogen bond with N2 of G8. Interestingly, the double helical structure is only minimally perturbed as a result of PM01183 binding, except for widening of the minor groove and a slight bend towards the major groove.

In the covalent complex containing the PM01183-G7 adduct (Figure 2A, right), the minor groove remains widened and the major groove appears compressed in the central region due to significant increases in roll at C6/G7 and G7/G8 steps. As a result, the conformation of the DNA is slightly bent. This finding together with the facts that the minor groove is occupied and some functional groups are exposed to the solvent are likely to have an impact on the ability of this binary complex to associate or interfere with one or more DNA-binding proteins involved in transcription, replication and/or repair within the cell, not necessarily the same that are affected by trabectedin-DNA adducts.

PM01183 treatment induces double-strand DNA breaks in cancer cells and S-phase accumulation

We further assessed whether PM01183-DNA adducts could eventually give rise to double-strand breaks (DSBs) in living cells. For this purpose, we analysed the formation of γ-H2AX foci as a surrogate indicator of DSB formation. Figure 2B shows typical results found in A549 cells treated with PM01183 for 6 h at the indicated concentrations. At 5 nM, few cells showed γ-H2AX staining. In contrast, at 25 nM,

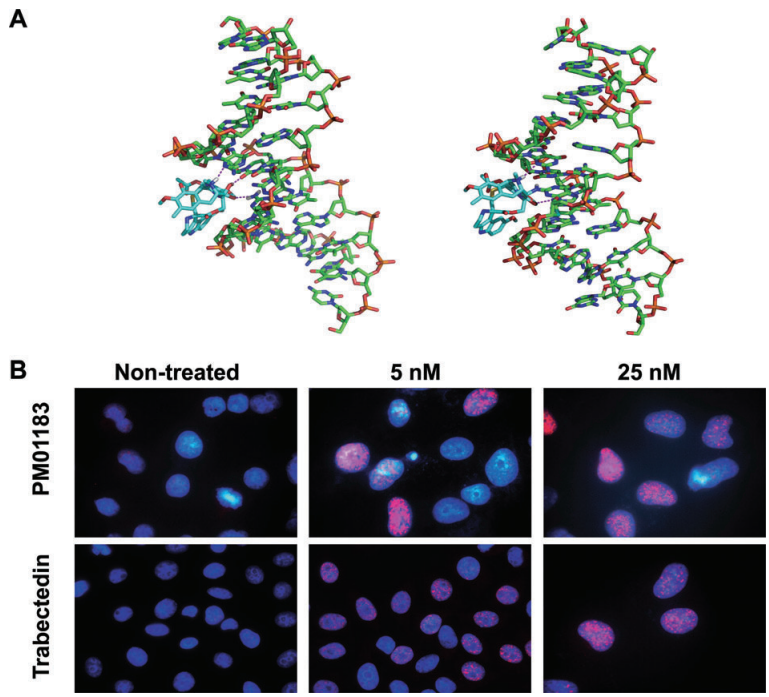


Figure 2

Modelling of DNA–PM01183 complexes and induction of DNA double-strand breaks following exposure to PM01183. (A) Refined and equilibrated molecular models of the pre-covalent (left) and covalent (right) complexes between PM01183 (C atoms in cyan) and the CGG-containing oligonucleotide (C atoms in green). The dotted lines represent hydrogen bonds. (B) A549 cells were treated with either PM01183 or trabectedin at the indicated concentrations during 6 h, followed by additional 18 h of incubation without the drug. After fixation, cells were immunostained for γ-H2AX and nuclei were visualized with Hoechst 33342.

most of the cells were stained for γ-H2AX and each of the positive cells carried a large number of foci, indicating that several DSBs per cell were induced at this drug concentration. As shown in Figure 2B, equimolar concentrations of PM01183 and trabectedin caused similar levels of DSBs.

Table 1 and Table S2 summarize the results of our assessment of possible cell cycle perturbations after 12 and 24 h of exposure to PM01183. At both time intervals, low (15 nM) and high (150 nM) drug concentrations induced an increase of the S-phase population in A549 (lung), HCT116 (colorectal), HT29 (colorectal) and MDA-MB-231 (breast) cancer cells that was more pronounced after exposure for 24 h. Of note, the S-phase accumulation was higher at 15 nM, probably due to a higher cell survival. In IGROV-1 cells, we did not observe any S-phase accumulation, most likely due to the extensive cell death induced by the compound at both concentrations after 24 h treatment (see below).

PM01183 induces cell death by apoptosis

We investigated whether exposure to PM01183 was able to induce a canonical apoptosis response in tumour cells. First, we determined whether the drug was able to generate late apoptosis by detecting the flow cytometric sub-G_1 cell population (Figure 3A, Table 1). Treatment of A549, HCT116 and HT-29 cells resulted in an increase in the sub-G_1 population at the highest concentration (150 nM). On the other hand, MDA-MB-231 cells died with a slower kinetics while IGROV-1 cells more rapidly. These data indicated that the percentage of cells entering the sub-G_1 peak after 24 h of treatment with PM01183 was cell type- and concentration-dependent. Fluorescence microscopy confirmed that exposure of the A549 cell line to 150 nM PM01183 for 24 h induced the appearance of cells with a visible apoptotic nuclear morphology including strong chromatin condensation (Figure 3B).

Table 1
Cell cycle perturbations induced by PM01183

	Cell cycle phase	Control	15 nM	150 nM
A549	Sub-G_1	3	11	28
	G_0/G_1	62	24	37
	S-phase	16	40	23
	G_2/M	19	25	12
HT29	Sub-G_1	7	8	25
	G_0/G_1	55	17	31
	S-phase	26	68	29
	G_2/M	12	7	15
HCT116	Sub-G_1	5	7	37
	G_0/G_1	59	21	36
	S-phase	17	45	14
	G_2/M	19	27	13
MDA-MB-231	Sub-G_1	3	3	3
	G_0/G_1	53	30	37
	S-phase	25	50	33
	G_2/M	19	17	27
IGROV-1	Sub-G_1	4	30	76
	G_0/G_1	53	33	9
	S-phase	20	15	9
	G_2/M	23	22	6

Cells were treated with the appropriate amount of the compound for 24 h, and then stained with 0.4 $\mu g \cdot mL^{-1}$ propidium iodide. Samples were analysed with a FACScalibur flow cytometer (Beckton and Dickinson, Frankin Lakes, NJ, USA) and the FlowJo7 cytometry analysis software. Values are expressed as % of cells.

We further studied the caspase-dependent cleavage of the nuclear protein PARP as a marker of canonical apoptosis. The p85 band, generated through proteolytic cleavage of p116 PARP by activated caspases, was clearly visible as early as 6 h after the treatment of A549 cells with PM01183 150 nM (Figure 3C). After 24 h of treatment, a considerable proportion of the PARP protein had already been cleaved by caspases, indicating that most of the cells were undergoing canonical apoptosis.

Finally, we analysed the onset of apoptosis using a commercially available solid-phase sandwich enzyme immunoassay that detects the cleavage of cytokeratin-18 (CK-18) by caspases as a sensitive and specific marker of early canonical apoptosis. Treatment with 150 nM PM01183 induced a marked increase of the apoptosis-related CK-18 Asp396 epitope in A549 cells. Moreover, simultaneous treatment with the pan-caspase inhibitor, Z-Vad-fmk, strongly reversed this effect (Figure 3D), which clearly indicates that the pro-apoptotic effect of PM01183 was dependent on caspase activity.

PM01183 shows a potent cytotoxic activity

The *in vitro* cytotoxicity of PM01183 was determined using a panel of 23 tumour cell lines that represent 11 relevant types of human cancer (Table 2). Most of the cell lines were very sensitive to PM01183 treatment, with GI_{50} values in the low nanomolar range (<10 nM). The panel average GI_{50} was 2.7 nM, with 22RV1, MiaPaCa-2, HS746T and MOLT4 cell lines showing the lowest GI_{50}, and IGROV-1, NCI-H23, RXF393 and HCT116 cell lines presenting the highest GI_{50} values.

We then performed xenograft studies to test whether the cytotoxicity of PM01183 translates into *in vivo* anti-tumour activity. NCI-H460 (lung), A2780 (ovary), HT29 (colon) and HGC-27 (gastric) cells were xenografted into the right flank of athymic *nu/nu* mice. Once the tumours reached *c.* 150 mm^3, the mice were randomized into groups of 10–15 mice each and either vehicle or PM01183 (0.18 $mg \cdot kg^{-1} \cdot day^{-1}$) was intravenously administered in three consecutive weekly doses. At the drug doses used in the experiment, no significant toxicity or body weight loss was observed in the treated animals (data not shown). As shown in Figure 4, PM01183 presented anti-tumour activity with a statistically significant inhibition of tumour growth in the four tested models.

Discussion

In the present work, we describe PM01183, a new synthetic tetrahydroisoquinoline alkaloid that induced DNA damage and apoptosis in a variety of cancer cell lines in the low nanomolar range and showed a potent anti-tumour activity in several human cancer cell lines and mouse xenograft models.

We first demonstrated that PM01183 was able to bind to a naked, 250 bp long PCR product containing multiple guanines spread over its whole length. This binding was dependent on the presence of the reactive hemiaminal moiety in the drug, as ET-745, a related compound devoid of the hydroxyl group, did not produce any band shifts at any of the concentrations tested. By using a highly sensitive and miniaturized fluorescence-based DNA melting assay, we observed that the preferred DNA triplets for PM01183 binding were AGC, CGG, AGG and TGG (the underlined guanine denoting the site of covalent adduct formation). These preferences are different from those of other DNA minor groove binders that alkylate DNA. For example, anthramycin selects AGA and AGG triplets (Pierce *et al.*, 1993), saframycins prefer GGG or GGC (Rao and Lown, 1992) while CC-1065 and derivatives (adozelesin,

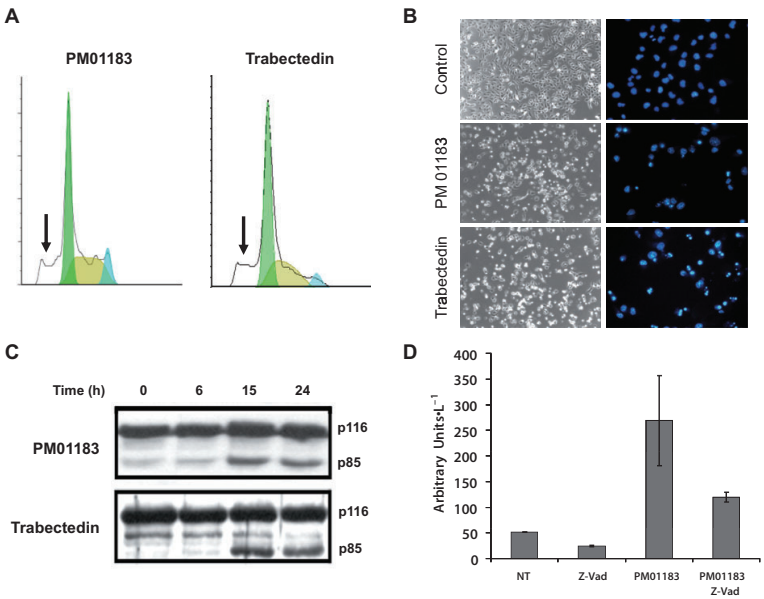


Figure 3

PM01183 induces apoptosis in living cells. (A) Flow cytometric DNA histograms showing sub-G1 peak induced by PM01183 and trabectedin (both drugs at 150 nM after 24 h exposure) in A549 cells. (B) Representative microphotographs of chromatin condensation of A549 cells without (Control) and after 24 h exposure to 150 nM of PM01183 or trabectedin. (C) Western blots of PARP in A549 cells after exposure to 150 nM of PM01183 and trabectedin at different time intervals. (D) Detection of cleaved cytokeratin-18 using the M30-Apoptosense kit in A549 cells after 24-hour exposure to PM01183 (150 nM) or not (NT).

bizelesin), as well as duocarmycin, prefer the minor groove of AT-rich regions (Baraldi *et al.*, 1999). On the other hand, preferential triplets for PM01183 binding were similar to those previously reported for trabectedin (TGG, CGG, AGC and GGC) (Marco *et al.*, 2006) and Zalypsis® (GGC and AGC) (Leal *et al.*, 2009). Nonetheless, although these three compounds share similar binding preferences, they differ in their kinetics of DNA adduct formation and in their degree of structural stabilization of the DNA double helix. PM01183 shows a higher affinity but gives rise to a lower duplex stabilization compared with trabectedin ($1/C_{50}$: 0.5 μM, data not shown; ΔT_m(max) ~20°C) (Casado *et al.*, 2008), which seems to indicate that PM01183 adducts are produced faster but they are slightly less effective in preventing strand separation, a process that must take place during transcription and replication. Whether these differences in affinity and DNA stabilization translate into differences in anti-tumour activities is not currently known in detail, although the present results appear to point in this direction.

A rationale for the reported binding preferences was provided by molecular modelling results which showed that a specific pattern of hydrogen bonds was necessary for both sequence recognition and PM01183 activation. The covalent complex with a representative oligonucleotide containing a central CGG triplet showed relatively minor structural distortions in the DNA molecule, the most significant being a widening of the minor groove in the region where the drug is bound and a distinct bending towards the major groove. Importantly, the guanine-bonded drug also establishes non-bonded (hydrophobic, van der Waals and electrostatic, including hydrogen bonding) interactions with surrounding nucleotides from both strands. A likely consequence, supported by the DNA melting results reported above, is that one or more PM01183-guanine adducts can hamper or prevent strand separation, thereby blocking replication and transcription forks. As a result, PM01183's behaviour could be akin to that of agents giving rise to true interstrand crosslinks (ICLs), such as mitomycin (Casado *et al.*, 2008). In this respect, it is of interest that ICL resolution is known to occur through the coordinated action of multiple DNA repair pathways, including homologous recombination, a process that gives rise to DSBs because both DNA strands are covalently modified and need to be incised.

Table 2
Cytotoxicity of PM01183 in a panel of 24 human cancer cell lines

	Cell line	PM01183 GI_{50} (nM)	Trabectedin GI_{50} (nM)
Prostate	PC3	1.2 ± 0.6	5.1 ± 2.4
	22RV1	0.3	1.1 ± 0.5
Pancreas	PANC-1	2.8 ± 1.4	8.4 ± 2.4
	MiaPaCa-2	1 ± 0.2	9.9 ± 2.1
Ovary	IGROV-1	9.7 ± 10	4.8 ± 3.1
	A2780	1.5 ± 0.4	3.1 ± 2.3
Lung	NCI-H460	1.5 ± 0.8	1.8 ± 0.8
	NCI-H23	5.3 ± 0.2	2.1 ± 1.3
	A549	1.3 ± 0.1	10.4 ± 7.6
Liver	SK-HEP-1	2.6 ± 1	4.7 ± 0.9
	HEPG2	3 ± 0.3	7.2 ± 1.3
Leukaemia	MOLT4	1.1 ± 0.04	1.6 ± 0.7
	K562	1.5 ± 1.8	3 ± 2.7
Kidney	RXF393	8.6 ± 0.5	3.3 ± 1
	CAKI-1	1.7 ± 1.1	5.9 ± 2
Stomach	HS746T	1 ± 3.1	9.9 ± 1.3
	HGC-27	2.8 ± 1.2	4.8 ± 1.2
Colon	LoVo	2 ± 1	1.5 ± 0.6
	HT29	2.4 ± 1	9.8 ± 4.3
	HCT-116	6.4 ± 3.7	4.9 ± 2.7
Breast	MDA-MB-231	3.4 ± 1.9	3.9 ± 4.2
	MCF-7	1.7 ± 0.6	2.6 ± 0.1
	BT-474	1.3 ± 0.4	4.6 ± 2

Values represent the mean ± SD of three different experiments.
GI_{50}, 50% growth inhibition.

When we tried to ascertain whether PM01183 treatment could induce DSB in human cancer cells, we confirmed that drug treatment at nanomolar concentrations did indeed give rise to a high proportion of DSB-positive cells with abundant γ-H2AX foci per cell. Moreover, the data presented here show that PM01183 induced an S-phase arrest at both low and high nanomolar concentrations. Altogether these findings indicate that some of the PM01183-DNA adducts are eventually transformed into DSB and induce an S-phase arrest by activation of the DNA replication checkpoints. Our data also showed that PM01183-induced cell death exhibited the characteristics of apoptosis, with chromatin condensation and cleavage of the caspase targets PARP-1 and CK-18.

These results are not surprising in light of similar findings recently reported for other DNA covalent minor groove binders. For example, trabectedin-DNA adducts trap the transcription-coupled-nucleotide excision repair (TC-NER) machinery, as it attempts to repair the lesion, at sites where those adducts have arrested RNA polymerase II (Pol II) (Erba *et al.*, 2001; Takebayashi *et al.*, 2001). In this action, the C-ring present in trabectedin, which protrudes from the DNA minor groove, has been proposed to interact with a key basic residue of the NER endonuclease XPG (Herrero *et al.*, 2006). Preliminary data support this view as fission yeast mutants defective in the orthologous Rad13 protein or tumour cells lacking XPG have been shown to be partially resistant to trabectedin (Takebayashi *et al.*, 2001; Herrero *et al.*, 2007). The formation of large ternary cytotoxic complexes (DNA–trabectedin–XPG) would result in DNA DSB during the S-phase and the repair of this type of DNA damage is known to require proteins from the homologous recombination machinery (Soares *et al.*, 2007; Tavecchio *et al.*, 2008). In contrast, the cytotoxic activity of Zalypsis®, which also induces DSB, is apparently not affected by NER deficiency as the formation of γ-H2AX foci was equally observed in models of NER-proficient and NER-deficient cells (Guirouilh-Barbat *et al.*, 2009). In this regard, it was recently found that PM01183 was three times less potent in NER-deficient cells compared with NER-proficient cells indicating that this compound could be also targeting TC-NER-dependent DNA repair (Aviles *et al.*, 2009). Trabectedin also induces the rapid and massive degradation of transcribing Pol II in various cancer cell lines (Aune *et al.*, 2008). Persistent Pol II degradation and lack of transcription would preclude the synthesis of essential cellular proteins. Similar alterations were observed after treatment of cancer cells with irofulven (Escargueil *et al.*, 2008). It is currently not clear whether PM01183 can also act on Pol II in a similar way.

When assayed in a panel of 23 cancer cell lines, PM01183 demonstrated a potent cytotoxic activity, with an average GI_{50} of 2.7 nM. All cell lines from the panel presented GI_{50} values lower than 10 nM. Other DNA minor groove binders of the tetrahydroisoquinoline family, including trabectedin, Zalypsis®, jorumycin, renieramycins and saframycins, have been previously shown to possess strong cytotoxic activity, with GI_{50}s also in the low nanomolar range (Arai *et al.*, 1980; Oku *et al.*, 2003; Lane *et al.*, 2006; Leal *et al.*, 2009). The *in vitro* cytotoxicity of PM01183 translated well into *in vivo* antitumour activity in four xenograft models of human cancer. We assayed the maximum tolerated dose of the drug with no toxic effects to the animals. PM01183 showed clear antineoplastic activity against xenograft models of different tissue types.

In summary, we have demonstrated highly potent *in vitro* and *in vivo* anti-cancer activities of PM01183. This agent exerts its anti-cancer effects

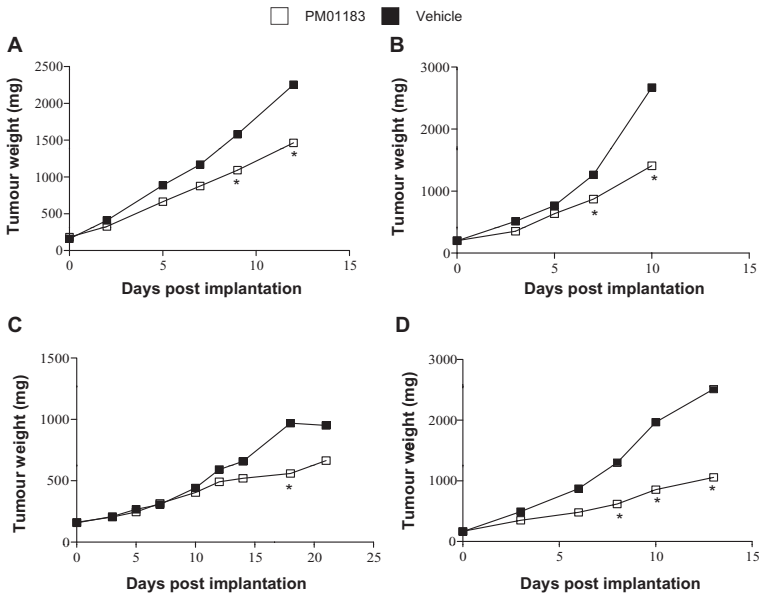


Figure 4

In vivo anti-tumour activity of PM01183. Treatment with PM01183 results in tumour growth inhibition in NCI-H460 lung (A), A2780 ovarian (B), HT29 colon (C) and HGC-27 gastric (D) xenograft models compared with vehicle-treated animals. Treatments started at a tumour volume size of *c.* 150 mm^3 and were intravenously administered in three cycles of three consecutive weekly doses at 0.18 $mg \cdot kg^{-1}$. Each point represents the median value ($n = 10$); $^*P < 0.05$.

through the covalent modification of guanines in the DNA minor groove that eventually give rise to DNA DSB, S-phase arrest and apoptosis in cancer cells. These encouraging results support its development as a novel anti-cancer agent in a wide spectrum of clinical settings.

Acknowledgements

The authors thank Beatriz de Castro, Guillermo Tarazona Ramos, Maribel Cercenado Mansilla, Patricia Martinez Rivas and Sofia Cascajares for their excellent technical assistance. This work was funded by PharmaMar SA.

Conflicts of interest

Alberto Domingo and Federico Gago have received a Research Grant from PharmaMar JFM Leal, M Martinez-Diez, V Moneo, MJ Guillen Navarro, P Aviles, C Cuevas, LF Garcia Fernandez and CM Galmarini are employees and shareholders of PharmaMar SA.

References

Arai T, Takahashi K, Nakahara S, Kubo A (1980). The structure of a novel antitumor antibiotic, saframycin A. Experientia 36: 1025–1027.

Aune GJ, Takagi K, Sordet O, Guirouilh-Barbat J, Antony S, Bohr VA *et al.* (2008). Von Hippel-Lindau-coupled and transcription-coupled nucleotide excision repair-dependent degradation of RNA polymerase II in response to trabectedin. Clin Cancer Res 14: 6449–6455.

Aviles PM, Galmarini CM, Cuevas C, Guillen MJ, Frapolli R, Uboldi S *et al.* (2009). Mechanism of action and antitumor activity of PM01183. In: Proceedings of the 100th Annual Meeting of the American Association for Cancer Research, AACR: Denver, CO, Abstract 2679.

Baraldi PG, Cacciari B, Guiotto A, Romagnoli R, Zaid AN, Spalluto G (1999). DNA minor-groove binders: results and design of new antitumor agents. Farmaco 54: 15–25.

Casado JA, Rio P, Marco E, Garcia-Hernandez V, Domingo A, Perez L *et al.* (2008). Relevance of the Fanconi anemia pathway in the response of human cells to trabectedin. Mol Cancer Ther 7: 1309–1318.

Case DA, Cheatham III, TE, Darden T, Gohlke H, Luo R, Merz Jr., KM *et al*. (2005). The Amber biomolecular simulation programs. J Comput Chem 26: 1668–1688.

David-Cordonnier MH, Gajate C, Olmea O, Laine W, Iglesia-Vicente J, Perez C *et al*. (2005). DNA and non-DNA targets in the mechanism of action of the antitumor drug trabectedin. Chem Biol 12: 1201–1210.

Erba E, Bergamaschi D, Bassano L, Damia G, Ronzoni S, Faircloth GT *et al*. (2001). Ecteinascidin-743 (ET-743), a natural marine compound, with a unique mechanism of action. Eur J Cancer 37: 97–105.

Escargueil AE, Poindessous V, Soares DG, Sarasin A, Cook PR, Larsen AK (2008). Influence of irofulven, a transcription-coupled repair-specific antitumor agent, on RNA polymerase activity, stability and dynamics in living mammalian cells. J Cell Sci 121: 1275–1283.

Garcia-Nieto R, Manzanares I, Cuevas C, Gago F (2000). Bending of DNA upon Binding of Ecteinascidin 743 and Phthalascidin 650 Studied by Unrestrained Molecular Dynamics Simulations. J Am Chem Soc 122: 7172–7182.

Guan Y, Sakai R, Rinehart K, Wang AH (1993). Molecular and crystal structures of ecteinascidins: potent antitumor compounds from the Caribbean thnicate ecteinascidia turbinata. J Biomol Struct Dyn 10: 793–818.

Guirouilh-Barbat J, Antony S, Pommier Y (2009). Zalypsis (PM00104) is a potent inducer of gamma-H2AX foci and reveals the importance of the C ring of trabectedin for transcription-coupled repair inhibition. Mol Cancer Ther 8: 2007–2014.

Herrero A, Marco E, García-Hernández V, Martín-Castellanos C, Domingo A, Alvarez E *et al*. (2007). Resilience of the cytotoxic effects of Zalypsis® (pm00104) to the lack of a functional nucleotide excision repair system. In: American Association for Cancer Research Annual Meeting. Proceedings: Los Angeles, CA, Abstract 1537.

Herrero AB, Martin-Castellanos C, Marco E, Gago F, Moreno S (2006). Cross-talk between nucleotide excision and homologous recombination DNA repair pathways in the mechanism of action of antitumor trabectedin. Cancer Res 66: 8155–8162.

Kishi K, Yazawa K, Takahashi K, Mikami Y, Arai T (1984). Structure-activity relationships of saframycins. J Antibiot (Tokyo) 37: 847–852.

Lane JW, Estevez A, Mortara K, Callan O, Spencer JR, Williams RM (2006). Antitumor activity of tetrahydroisoquinoline analogues 3-epi-jorumycin and 3-epi-renieramycin G. Bioorg Med Chem Lett 16: 3180–3183.

Leal JF, Garcia-Hernandez V, Moneo V, Domingo A, Bueren-Calabuig JA, Negri A *et al*. (2009). Molecular pharmacology and antitumor activity of Zalypsis in several human cancer cell lines. Biochem Pharmacol 78: 162–170.

Manzanares I, Cuevas C, Garcia-Nieto R, Marco E, Gago F (2001). Advances in the chemistry and pharmacology of ecteinascidins, a promising new class of anti-cancer agents. Curr Med Chem Anticancer Agents 1: 257–276.

Marco E, David-Cordonnier MH, Bailly C, Cuevas C, Gago F (2006). Further insight into the DNA recognition mechanism of trabectedin from the differential affinity of its demethylated analogue ecteinascidin ET729 for the triplet DNA binding site CGA. J Med Chem 49: 6925–6929.

Moore BM, Seaman FC, Hurley LH (1997). NMR-based model of an Ecteinascidin 743-DNA adduct. J Am Chem Soc 119: 5475–5476.

Negri A, Marco E, Garcia-Hernandez V, Domingo A, Llamas-Saiz AL, Porto-Sanda S *et al*. (2007). Antitumor activity, X-ray crystal structure, and DNA binding properties of thiocoraline A, a natural bisintercalating thiodepsipeptide. J Med Chem 50: 3322–3333.

Oku N, Matsunaga S, van Soest RW, Fusetani N (2003). Renieramycin J, a highly cytotoxic tetrahydroisoquinoline alkaloid, from a marine sponge Neopetrosia sp. J Nat Prod 66: 1136–1139.

Pabo CO, Sauer RT (1984). Protein-DNA recognition. Annu Rev Biochem 53: 293–321.

Pierce JR, Nazimiec M, Tang MS (1993). Comparison of sequence preference of tomaymycin- and anthramycin-DNA bonding by exonuclease III and lambda exonuclease digestion and UvrABC nuclease incision analysis. Biochemistry 32: 7069–7078.

Pommier Y, Kohlhagen G, Bailly C, Waring M, Mazumder A, Kohn KW (1996). DNA sequence- and structure-selective alkylation of guanine N2 in the DNA minor groove by ecteinascidin 743, a potent antitumor compound from the Caribbean tunicate Ecteinascidia turbinata. Biochemistry 35: 13303–13309.

Rao KE, Lown JW (1992). DNA sequence selectivities in the covalent bonding of antibiotic saframycins Mx1, Mx3, A, and S deduced from MPE.Fe(II) footprinting and exonuclease III stop assays. Biochemistry 31: 12076–12082.

Reddy BS, Sharma SK, Lown JW (2001). Recent developments in sequence selective minor groove DNA effectors. Curr Med Chem 8: 475–508.

Rinehart K, Holt TG, Fregeau NL, Stroh JG, Keifer PA, Sun F *et al*. (1990). Ecteinascidins 729, 743, 745, 759A, 759B and 770: potent antitumor agents from the Caribean tunicate Ecteinascidia turbinata. J Org Chem 55: 4512–4515.

Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B (2009). The role of DNA shape in protein-DNA recognition. Nature 461: 1248–1253.

Seaman FC, Hurley LH (1998). Molecular basis for the DNA sequence selectivity of ecteinascidin 736 and 743: evidence for a dominant role of direct readout via hydrogen bonding. J Am Chem Soc 120: 13028–13041.

Soares DG, Escargueil AE, Poindessous V, Sarasin A, de Gramont A, Bonatto D *et al*. (2007). Replication and homologous recombination repair regulate DNA double-strand break formation by the antitumor alkylator ecteinascidin 743. Proc Natl Acad Sci USA 104: 13062–13067.

Susbielle G, Blattes R, Brevet V, Monod C, Kas E (2005). Target practice: aiming at satellite repeats with DNA minor groove binders. Curr Med Chem Anticancer Agents 5: 409–420.

Takebayashi Y, Pourquier P, Zimonjic DB, Nakayama K, Emmert S, Ueda T *et al*. (2001). Antiproliferative activity of ecteinascidin 743 is dependent upon transcription-coupled nucleotide-excision repair. Nat Med 7: 961–966.

Tavecchio M, Simone M, Erba E, Chiolo I, Liberi G, Foiani M *et al*. (2008). Role of homologous recombination in trabectedin-induced DNA damage. Eur J Cancer 44: 609–618.

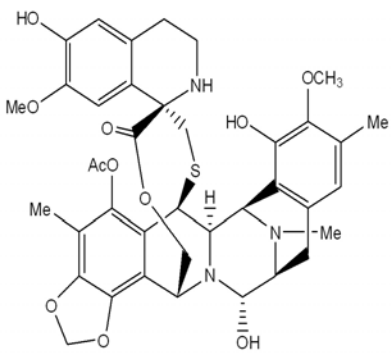
Supporting information

Additional Supporting Information may be found in the online version of this article:

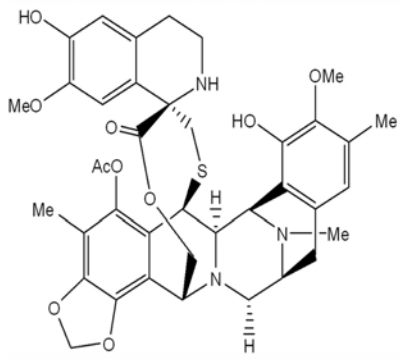
Figure S1 Chemical structures of trabectedin (ET-743) and ET-745.
Table S1 Sequences of the ds-oligonucleotides used in the DNA melting assays.
Table S2 Cell cycle perturbations after 12 h exposure of tumour cells to PM01183.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplementary Figure 1 – Leal et al



ET-743



ET-745

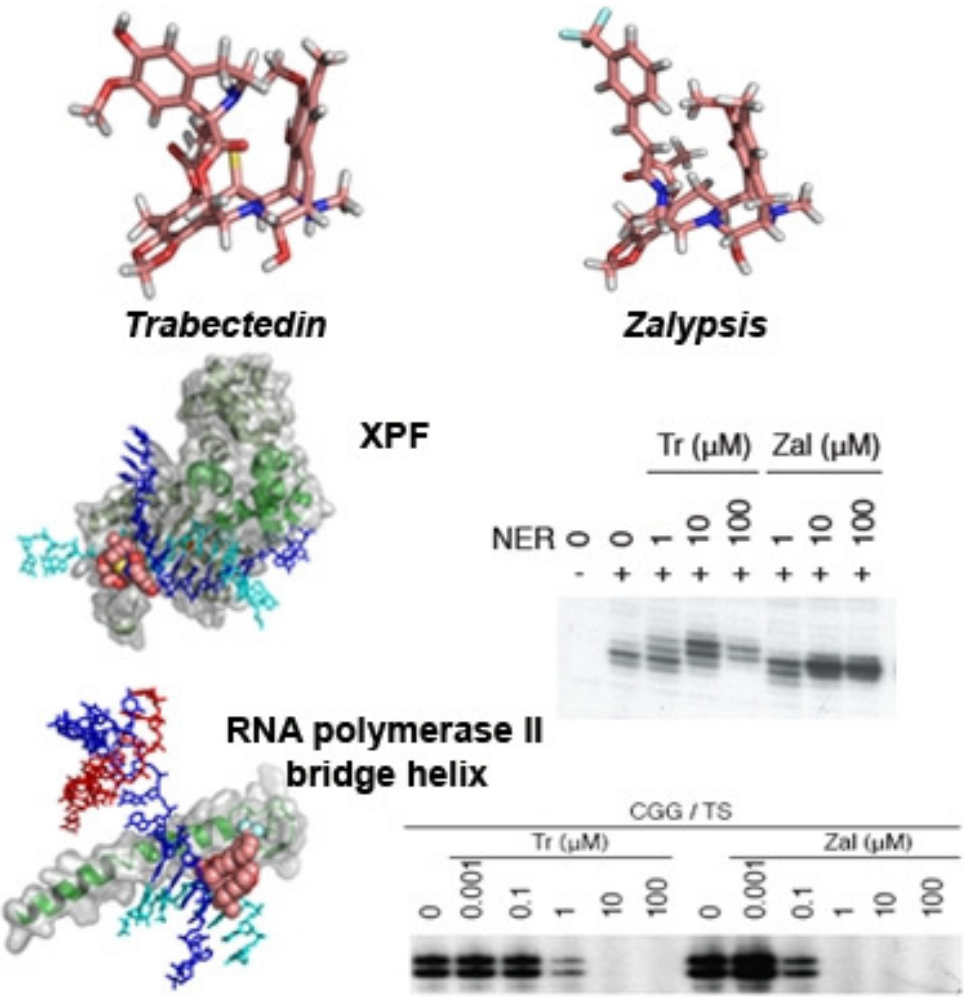
Supplementary Table 1. Sequence of the ds-Oligonucleotides used in the DNA melting assay

Triplet	Oligonucleotide Sequences
A**G**C	5'-F-AATATAAAGCAAATTAA-3'
T**G**C	3'-Q-TTATATTTCGTTTAATT-5'
T**G**G	5'-F-AATATAAACCAAATTAA-3'
G**G**T	3'-Q-TTATATTTGGTTTAATT-5'
A**G**A	5'-F-AATATAAAGAAATTAA-3' 3'-Q-TTATATTTCTTTAATT-5'
C**G**A	5'-F-AATATAAACGAAATTAA-3'
C**G**T	3'-Q-TTATATTTGCTTTAATT-5'
A**G**G	5'-F-AATATAAAGGCAAATTAA-3'
G**G**C T**G**C	3'-Q-TTATATTTCCGTTTAATT-5'
C**G**G	5'-F-AATATAAACGGAAATTAA-3'
G**G**A C**G**T	3'-Q-TTATATTTGCCTTTAATT-5'
C**G**T	5'-F-AATATAAAACGTTAAATTAA-3' 3'-Q-TTATATTTTGCAATTTAATT-5'
A**G**C	5'-F-AATATAAAGCGTTAAATTAA-3'
C**G**T C**G**C	3'-Q-TTATATTTCGCAATTTAATT-5'
A**G**G	5'-F-AATATAAAGGAAATTAA-3'
G**G**A	3'-Q-TTATATTTCCTTTAATT-5'
A**G**T	5'-F-AATATAAAGTAAATAA-3' 3'-Q-TTATATTTCATTTATT-5'
T**G**A	5'-F-AATATAATGAAAATTAA-3' 3'-Q-TTATATTACTTTTAATT-5'
T**G**C	5'-F-AATATAATGCAATTAA-3' 3'-Q-TTATATTACGTTAATT-5'
T**G**T	5'-F-AATATAAACAAATTAA-3' 3'-Q-TTATATTTGTTTAATT-5'
A**G**G	5'-F-AATATAAGGTAAATTAA-3'
G**G**T	3'-Q-TTATATTCCATTTAATT-5'
A**G**G	5'-F-AATATAAGGGAAATTAA-3'
G**G**G G**G**A	3'-Q-TTATATTCCCTTTAATT-5'
(AT)	5'-F-CAATTAAATATAAC-3' 3'-Q-GTTAATTTATATTG-5'

Artículo III

"*XPF-dependent DNA breaks and RNA polymerase II arrest induced by antitumor DNA interstrand crosslinking-mimetic tetrahydroisoquinoline alkaloids*"

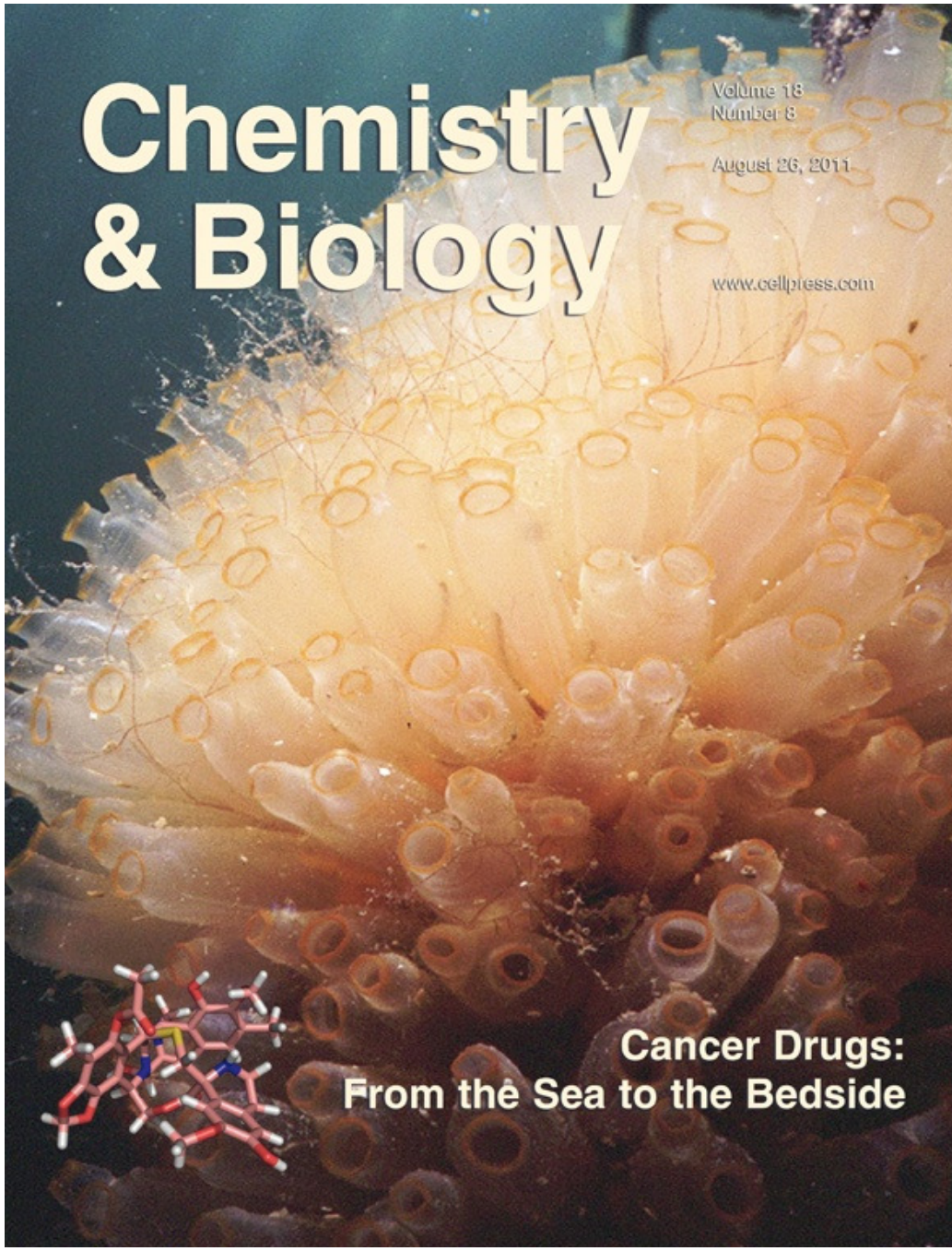
***Chem Biol*. 2011 Aug 26;18(8):988-999**



Yondelis®, un compuesto de origen marino y otro derivado tetrahidroisoquinolínico, *Zalypsis®*, se emplean en el tratamiento del sarcoma de tejidos blandos, cáncer de ovario y mielomas. Sin embargo, los mecanismos moleculares de acción siguen sin estar del todo claros. Aquí demostramos que actúan como entrecruzadores intercatenarios, a pesar de que se unen covalentemente a una sola cadena de ADN y por lo tanto no interfieren con la reparación por escisión de nucleótidos. De acuerdo con los ensayos in vitro y *comet*, estos aductos inducen una determinada estructura de ADN que favorece la escisión por la nucleasa ERCC1/XPF. Estos fármacos también inhiben la síntesis de ARN, ya sea al evitar el reclutamiento de factores de transcripción específicos y/o por la detención de la ARN polimerasa II, independientemente de si los compuestos se unen a la cadena transcrita o a su complementaria. El conocimiento profundo acerca de la mecánica de estos fármacos puede ser útil para entender la transcripción y la reparación de los entrecruzamientos intercatenarios, así como para mejorar la terapia del cáncer.

Aspectos destacados

- Yondelis y Zalypsis funcionalmente imitan a los entrecruzadotes intercatenarios.
- EL ADN dañado por Yondelis y Zalypsis es hidrolizado por la endonucleasa XPF/ERCC1.
- Los aductos producidos en el ADN por estos fármacos detienen la elongación de la transcripción por la ARN polimerasa II.


Chemistry
& Biology
Volume 18
Number 8
August 26, 2011
www.cellpress.com
Cancer Drugs:
From the Sea to the Bedside

Chemistry & Biology
Article

XPF-Dependent DNA Breaks and RNA Polymerase II Arrest Induced by Antitumor DNA Interstrand Crosslinking-Mimetic Alkaloids

Sascha Feuerhahn,[1,5] Christophe Giraudon,[1,5] Marta Martínez-Díez,[2] Juan A. Bueren-Calabuig,[3] Carlos M. Galmarini,[2] Federico Gago,[3] and Jean-Marc Egly[1,4,*]
[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UdS, BP 163, 67404 Illkirch Cedex, C. U. Strasbourg, France
[2]Cell Biology Department, PharmaMar, Avda. de los Reyes, 1 Pol. Ind. La Mina, 28770 Colmenar Viejo, Madrid, Spain
[3]Departmento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain
[4]Institut de Recherche contre les Cancers de l'Appareil Digestif, 1 place de l'Hôpital, 67000 Strasbourg, France
[5]These authors contributed equally to this work
*Correspondence: egly@igbmc.fr
DOI 10.1016/j.chembiol.2011.06.007

SUMMARY

Trabectedin and Zalypsis are two potent anticancer tetrahydroisoquinoline alkaloids that can form a covalent bond with the amino group of a guanine in selected triplets of DNA duplexes and eventually give rise to double-strand breaks. Using well-defined in vitro and in vivo assays, we show that the resulting DNA adducts stimulate, in a concentration-dependent manner, cleavage by the XPF/ERCC1 nuclease on the strand opposite to that bonded by the drug. They also inhibit RNA synthesis by: (1) preventing binding of transcription factors like Sp1 to DNA, and (2) arresting elongating RNA polymerase II at the same nucleotide position regardless of the strand they are located on. Structural models provide a rationale for these findings and highlight the similarity between this type of DNA modification and an interstrand crosslink.

INTRODUCTION

One major challenge in cancer therapeutics is to characterize the altered biochemical pathway in the tumor cell that can be selectively targeted by drugs. Many clinically used antineoplastic agents cause DNA damage, and their action is influenced by their ability to specifically disturb fundamental DNA metabolic processes such as replication, repair, and transcription (Bustamante et al., 2011).

Trabectedin (Yondelis, ecteinascidin-743), a compound originally extracted from the marine tunicate *Ecteinascidia turbinata* (Rinehart et al., 1990) and commercially prepared nowadays by semisynthesis from microbially produced cyanosafracin (Cuevas and Francesch, 2009), and Zalypsis, a molecule structurally related to jorumycin and renieramycins (Scott and Williams, 2002), are two tetrahydroisoquinoline alkaloids (Figures 1A and 1B) endowed with potent anticancer activities. Trabectedin is currently used for the treatment of advanced soft tissue sarcoma and, in combination with pegylated liposomal doxorubicin, relapsed platinum-sensitive ovarian cancer (D'Incalci and Galmarini, 2010). Zalypsis is particularly effective against leukemia and stomach tumor cell lines (Leal et al., 2009), is in phase II clinical trials (Oku et al., 2003), and appears to be promising for the treatment of multiple myelomas (Ocio et al., 2009).

Both drugs bind covalently to the exocyclic 2 amino group of a central guanine (G) in selected triplet sequences of double-stranded DNA (dsDNA). Thus, the adducts are located in the DNA minor groove (Leal et al., 2009; Pommier et al., 1996; Zewail-Foote and Hurley, 1999), where the drugs also establish additional noncovalent interactions with the sugar-phosphate backbone and other nucleobases in both the strand containing the adduct (AS) and the complementary, opposite strand (OS), of the same duplex. As a result, the drug-induced lesion has been suggested to functionally mimic an interstrand crosslink (ICL), a type of DNA damage that is highly effective in blocking replication and transcription (Casado et al., 2008; De Silva et al., 2000). ICL resolution occurs through the coordinated action of multiple DNA repair pathways (McHugh et al., 2001), including nucleotide excision repair (NER) and homologous recombination (HR), in a process that involves double-strand breaks (DSBs). Interestingly, exposure of cells to trabectedin (Casado et al., 2008; Guirouilh-Barbat et al., 2008; Soares et al., 2007; Tavecchio et al., 2008) or Zalypsis (Guirouilh-Barbat et al., 2009; Leal et al., 2009; Ocio et al., 2009) results in the generation of DSBs, as assessed by γ-H2AX and Rad51 foci formation. Moreover, HR-deficient Fanconi anemia cells are extremely sensitive to trabectedin (Casado et al., 2008). On the contrary, trabectedin is more cytotoxic in NER-proficient cells than in their NER-deficient counterparts, particularly those harboring deficiencies in the XPG and XPF endonucleases, XPB helicase, and/or ERCC1, the partner of XPF (Damia et al., 2001; Takebayashi et al., 2001; Zewail-Foote et al., 2001). Furthermore, it was shown that pharmacological concentrations of trabectedin can inhibit expression of many NF-Y targeted genes including those encoding the multidrug resistance efflux pump MDR1 (Jin et al., 2000) and the molecular chaperone HSP70 (Minuzzo et al., 2000) as well as several cell cycle regulators such as the Sp1-regulated *p21* gene (Friedman et al., 2002). It is currently unknown whether Zalypsis affects RNA synthesis.

To gain further insight into the mechanism of action of these two antitumor tetrahydroisoquinolines, we investigated their

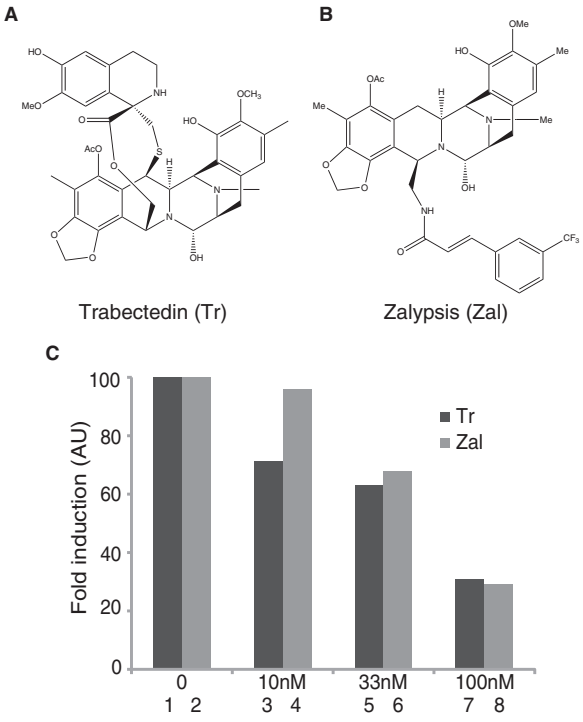


Figure 1. Cytotoxic Effects of Trabectedin and Zalypsis
Structures of trabectedin (A) and Zalypsis (B).
(C) Real-time PCR analysis of *RARβ2* induction by t-RA from HeLa cells pretreated with different concentrations of trabectedin (dark gray) and Zalypsis (light gray). The fold activation (noninduced versus induced) is shown, and *RARβ2* mRNA was normalized to *GAPDH* mRNA levels.

effect on transcription and DNA repair both in vitro and in vivo. By using a bead-immobilized SV40 promoter-containing DNA template, we explored whether trabectedin can compete with the transcription factor Sp1 for binding to its responsive elements and, therefore, prevent the accurate positioning of the transcription machinery around the promoter. We next assessed, by means of another bead-immobilized tailor-made oligonucleotide containing a single CGG triplet (a favored site for trabectedin and Zalypsis covalent bond formation) (Pommier et al., 1996), whether these drugs can inhibit RNA synthesis by preventing the formation of the preinitiation transcription complex and/or by blocking RNA polymerase II (RNA pol II) elongation (independently of the strand to which the drug is bonded), and we sought confirmation in HeLa cells. We also tested whether adducts formed by these drugs can be recognized by defined components of the NER machinery. Finally, computer modeling and molecular dynamics (MD) simulations help us to rationalize our experimental observations.

RESULTS

Antitumor Activities and Effects on Gene Expression of Trabectedin and Zalypsis

The growth inhibitory activity of the two drugs was evaluated in a panel of sarcoma (SW872, A673, and Saos-2), breast (MCF7), and ovarian (A2780) cancer cell lines. All the cancer cell lines tested were very sensitive to trabectedin and Zalypsis (Table 1), with IC_{50} values in the high picomolar/low nanomolar range (0.1–3.0 and 0.4–1.8 nM, respectively). Their effect on gene expression was studied in HeLa cells by monitoring transcription of the *RARβ2* gene, a retinoic acid-responsive gene whose expression was induced with 10^{-5} M all-trans retinoic acid (t-RA). We observed that 1 hr of exposure to increasing concentrations of trabectedin and Zalypsis before incubation with t-RA over 6 hr led to a dose-dependent inhibition of *RARβ2* gene expression. At 100 nM, 70% of *RARβ2* transcription was inhibited (Figure 1C).

Trabectedin Competes with the DNA-Binding Transcription Factor Sp1

DNA-binding transcription factors convey information by associating to their responsive elements. Thus, we investigated how these drugs could affect the binding of Sp1 (Dynan and Tjian, 1983), a ubiquitous transcription factor that specifically recognizes GC-rich sequences in many gene promoters. The Sp1-responsive element-containing SV40 promoter (Mathis and Chambon, 1981) was ^{32}P labeled, immobilized on magnetic beads, and incubated with increasing amounts of either Sp1-containing HeLa nuclear extracts (NEs) or trabectedin for 30 min before being washed to remove any nonspecific DNA-binding molecules. DNase I footprinting showed that increasing amounts of Sp1 protected the SV40 promoter between −106 and −45 as well as a stretch closer to the transcription start site, as visualized by the disappearance of the G_{-9} and C_{-10} hypersensitive sites (HSs) (Figure 2A, lanes 1–4). Trabectedin, which is known to possess high affinity for several GC-rich triplets (Pommier et al., 1996), also targets the Sp1-binding sites on the SV40 promoter in a dose-dependent manner, in addition to DNA sequences located at positions C_{-116} or C_{-119} and G_{-3} or G_{-7}, with a DNase I HS being visible at position G_{-60} (lanes 5–8).

Table 1. Growth Inhibitory Activities of Trabectedin and Zalypsis on Different Cancer Cell Lines

		IC_{50} (nM)[a]	
Tissue Type	Cell Line	Trabectedin	Zalypsis
Liposarcoma	SW872	0.5	0.8
Ewing sarcoma	A673	1.0	0.4
Osteosarcoma	Saos-2	0.1	0.9
Breast cancer	MCF7	2.6	1.8
Ovarian cancer	A2780	3.1	1.1

Tumor cells were seeded in 96-well trays and incubated with serial dilutions of each drug for 72 hr.
[a] Half-maximal inhibitory concentration.

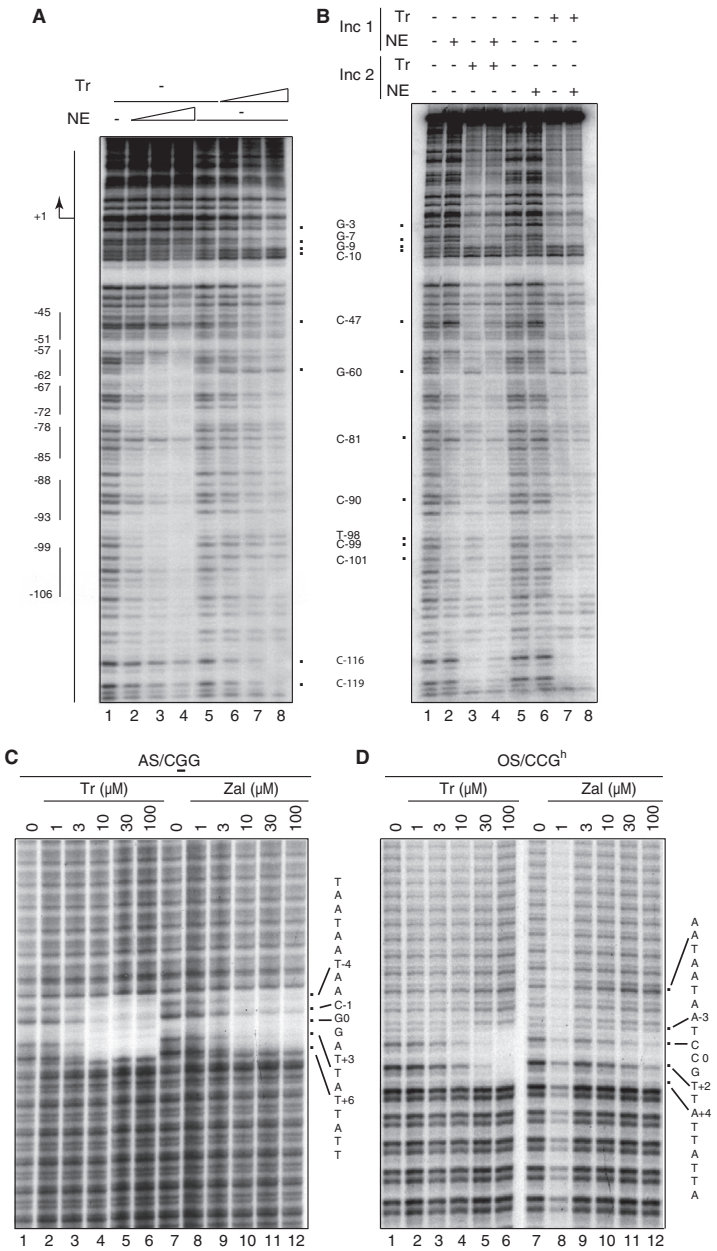


Figure 2. DNase I Footprinting Assays on the SV40 Promoter and on DNA Templates Containing a Unique Drug-Binding Site

(A) The ^{32}P-labeled and immobilized SV40 promoter was incubated with increasing amounts of either HeLa NE or trabectedin before being digested by DNase I. The Sp1-binding sites are shown at the right of the gel.

(B) Different rounds of incubation of either HeLa NE or trabectedin, as indicated at the top of the panel, were performed on the bead-immobilized SV40 promoter-containing DNA followed by DNase I treatment.

The AS/CGG (C) or OS/CCG^h (D) was incubated with increasing drug concentrations and treated with DNase I. The positions of the HS (A and B) or the protected nucleotides (C and D) are indicated (nb: for C and D the G0 is the guanine that is covalently bonded to the drug).

See also Figure S1.

To further investigate whether the DNA-binding factor and the drug compete for similar target sites, we set up a challenge assay based on sequential rounds of incubation of either Sp1 or trabectedin. When the DNA template was first incubated with Sp1-containing HeLa NE and then with the drug, we observed an extended protection originated by Sp1 itself (Figure 2B, lanes 4 and 2) and the absence of the drug-specific HS located at positions G_{-60}, T_{-98}, C_{-99}, and C_{-101} (compare lanes 4 and 3). This suggests that the drug does not displace Sp1 from its binding sites on the DNA template. The converse experiments, in which the DNA template was first incubated with the drug and then with Sp1, showed the same footprinting pattern as when the template was incubated with trabectedin alone (lanes 7 and 8). Thus, under the conditions employed, trabectedin competes with the transcription factor Sp1 for targeting specific DNA sequences (see Figure S1 available online).

Trabectedin and Zalypsis Protect from DNase I Digestion and Inhibit RNA pol II Transcription

To learn more about how these anticancer drugs affect RNA synthesis, we designed a bead-immobilized ^{32}P-labeled DNA substrate (Riedl et al., 2003) containing a unique, high-affinity, adduct-forming site (5′-CGG-3′) in just one strand flanked by TAA and ATT repeats on each side. After 30 min of exposure to increasing drug concentrations to saturate the high-affinity site, followed by extensive washing to remove any unbound molecules, this substrate was treated with DNase I. The AS/CGG was protected from DNase I digestion from nucleotides (nt) T_{-4} to T_{+6}, being G_0 the guanine to which trabectedin is covalently bonded (Figure 2C, lanes 1–6). The Zalypsis adduct led to a similar but weaker protection (lanes 7–12). An

identical concentration-dependent protection of the OS/CCG^h, which associates with either trabectedin or Zalypsis only through hydrogen bonds and van der Waals interactions involving the CCG^h triplet, was likewise detected from nt A_{-3} to A_{+4} (Figure 2D). These data show that the presence of the covalently bonded drugs in the minor groove of the DNA double helix specifically protects a CGG triplet and the surrounding nucleotides from DNase I access to both DNA strands. The use of 3- and 10-fold excesses of drug allowed us to confirm that this was a unique drug-binding site on the whole polynucleotide chain because no other footprints were apparent at these very high concentrations.

The effect of these drugs on RNA synthesis was investigated by performing an in vitro transcription assay using as template a 5′ end bead-immobilized DNA molecule containing the adenovirus major late promoter (AdMLP) in front of a G-less cassette (Figure 3A). Following preincubation with the indicated drug concentrations and washes to remove the excess of unbound drug, the beads were incubated with highly purified RNA pol II and the basal transcription factors TBP, TFIIB, TFIIE, TFIIF, and TFIIH, together with nucleotide triphosphates (NTPs) with the exception of GTP. Drug exposure resulted in decreased synthesis of RNA transcripts compared to controls (Figure 3B). In these experimental conditions we did not observe the presence of any shorter RNA transcripts due to premature RNA synthesis arrest (data not shown). Of note, Zalypsis completely inhibited transcription at a concentration of 1 μM (lane 10), whereas a much higher concentration of trabectedin was necessary to abolish transcription entirely (lane 4). This may indicate that Zalypsis is more potent than trabectedin in inhibiting RNA synthesis. Altogether, the above experiments suggest that some drug molecules target the promoter and thereby prevent formation of the transcriptional preinitiation complex, thus inhibiting RNA synthesis.

To assay the effects of trabectedin and Zalypsis on transcription elongation, we used another 5′ end bead-immobilized substrate containing, in addition to the AdMLP and the G-less cassette of 96 nt, a 127 nt long DNA sequence encompassing a CGG triplet at position +113 surrounded by AAT or ATT repeats (Figure 3A). The CGG drug-binding site was placed either on the transcribed strand (TS) or the nontranscribed strand (NTS). By following this strategy one trabectedin or Zalypsis adduct was placed on either the TS or the NTS DNA strand during RNA pol II elongation. The CGG/TS template was first incubated with all transcription components in the absence of GTP. In these conditions RNA pol II paused at the end of the G-less cassette (+96/+99). After several washes to remove both the remaining preinitiation complexes and the nonspecifically bound factors, the immobilized template containing only the elongating RNA pol II (Riedl et al., 2003; Zawel et al., 1995) was incubated with increasing drug concentrations. After washing off the excess of drugs, addition of all NTPs including GTP allowed RNA pol II to further elongate. We observed that both drugs inhibited full-length transcription (FLT: +223; Figure 3C, lanes 1–6 and 8–13, respectively). Zalypsis-dependent FLT inhibition was complete at a much lower concentration than that required by trabectedin (lanes 11–12 and 4–5). Furthermore, RNA pol II was blocked by trabectedin a few nucleotides before the CGG site targeted by the drug, i.e., at positions +108, +109, and +110 (lane 6), whereas Zalypsis arrested RNA pol II at +108 and +109 (lane 13). Another difference was the detection of a Zalypsis-specific RNA pol II stop site at position +98 (lane 13). This could be rationalized because this drug, but not trabectedin, binds to the 5′-TGT-3′ triplet and stabilizes a duplex containing this sequence (Leal et al., 2009).

Strikingly, when we investigated whether RNA pol II elongation could also be affected by the presence of a trabectedin or Zalypsis adduct in the CGG/NTS, FLT inhibition was observed not only at the same drug concentrations but also at the same nucleotide positions that were identified previously on the CGG/TS construct (Figures 3C and 3D, compare lanes 5 and 11). This result strongly supports the view that the adducts formed by either trabectedin or Zalypsis mimic ICLs.

Trabectedin and Zalypsis Promote the 5′ XPF/ERCC1-Dependent Incision

We next investigated whether the binding of trabectedin and Zalypsis to DNA would interfere with NER, given the established connection between these drugs and several NER factors (Takebayashi et al., 2001). To examine whether the NER machinery could eliminate this type of DNA damage, we incubated internally labeled polynucleotides containing a single drug-binding site with the drugs. After extensive washing, we added the highly purified NER factors XPC/HR23B, TFIIH, XPA, RPA, XPG, and XPF/ERCC1 (hereafter XPF) and ATP. In these conditions we did not observe removal of the damaged oligonucleotide (data not shown), but rather unexpectedly, we detected that the OS/CCG^h strand was cut at positions A_{-4} and A_{-3}, whereas the AS/CGG strand (drug adduct at the underlined guanine) was slightly incised at positions T_{-7} and A_{-5} (Figure 4A, lanes 2 and 6). XPF endonuclease was responsible for this cleavage because the incision signal was lost in its absence (lane 7), whereas the OS/CCG^h substrate was still cleaved when XPG was omitted (lane 8). Furthermore, the OS/CCG^h pretreated with either trabectedin (Figure 4C) or Zalypsis (Figure 4E) was better cut by XPF in a concentration-dependent manner, but the AS/CGG template was not significantly cleaved (Figures 4B and 4D).

We reasoned that the weak XPG-uncoupled XPF-dependent cut on the drug-free DNA template (Figures 4B and 4D, lanes 1–5) must be due to the particular choice of TAA and ATT repeats on each side of the CGG target triplet giving rise to a sequence-dependent deformation of the DNA helix (most likely melting) that is partially recognized by XPF. This interpretation was supported by results from MD simulations of a solvated 15-mer in 0.1 M NaCl at 400 K, i.e., under conditions that favor strand separation. Indeed, initial fraying of the oligomer's ends was followed by splaying and complete melting of the two strands after ~200 ns, whereas the base pairs making up the central CGG triplet in the drug-modified DNAs remained hydrogen bonded throughout the whole simulation (Figure S2; unpublished data).

We also found that in the presence of replication protein A (RPA), which plays a role in all DNA metabolic processes involving ssDNA, the XPF-dependent incision on the OS/CCG^h — that is enhanced by trabectedin — was stimulated still further (Figure 4F, lanes 15–18). Remarkably, the weak XPF-dependent cuts observed in the absence of drug on the AS/CGG template at positions T_{-7} and A_{-5} (lanes 5 and 9), and on

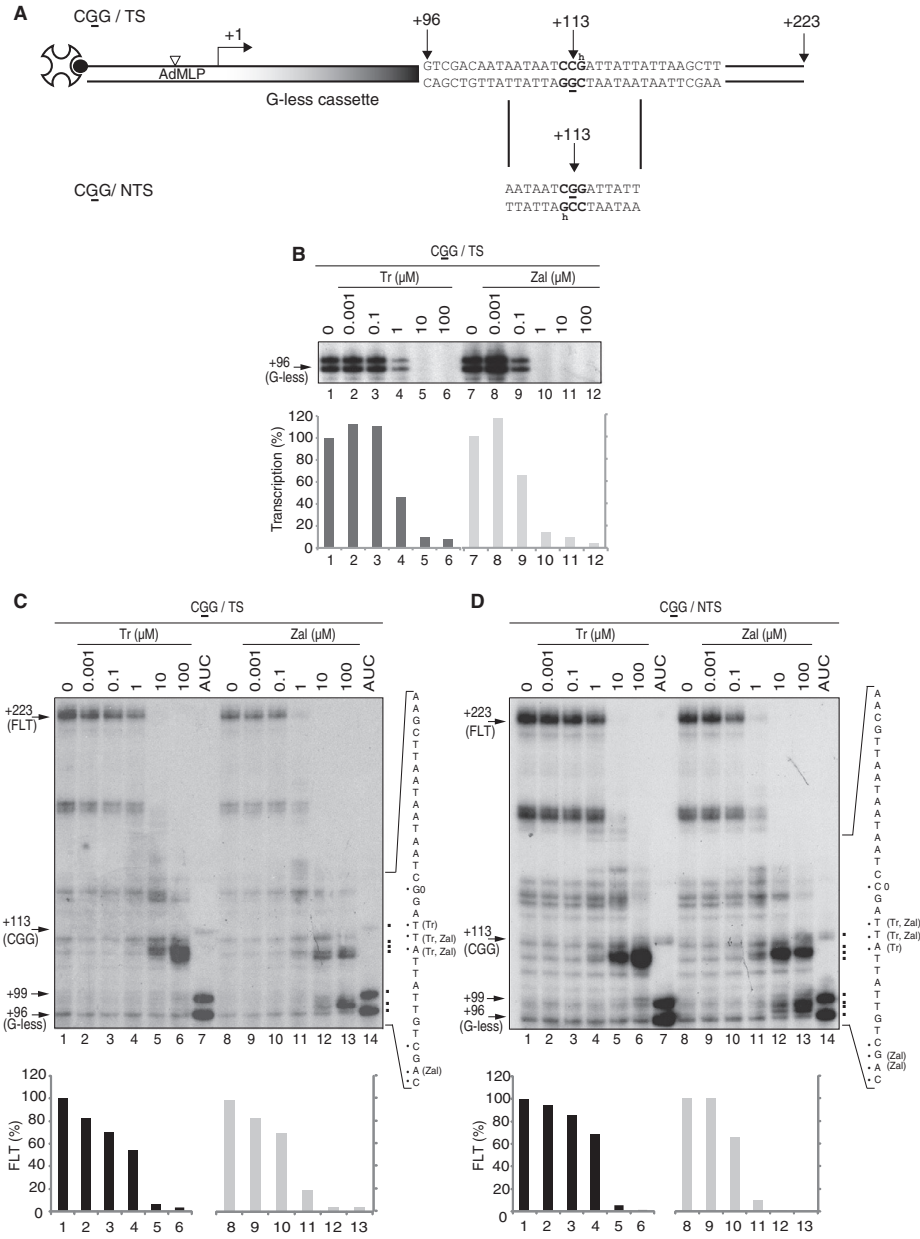


Figure 3. Effects of Trabectedin and Zalypsis on Transcription

(A) Scheme of the 5′ end-immobilized DNA containing the unique binding site 5′-CGG-3′. The drug-modified guanine (G) is positioned at +113 bp of the transcription start site (+1) either on the CGG/TS or the CGG/NTS.

(B) The DNA template was incubated with amounts of drug as indicated. Quantification of the +96 and +99 nt long transcripts was performed using the ImageJ software.

The CGG/TS (C) and the CGG/NTS (D) DNA templates were transcribed in the presence of NTPs except GTP, leading to RNA pol II arrest at the end of the G-less-cassette. After incubation with increasing amounts of trabectedin and Zalypsis and washing off the unbound drug, transcription was resumed in the presence of

the OS/CCG^h at position C_{-1} (lanes 15 and 17), became undetectable at high drug concentrations. This demonstrates how the drug-modified DNA becomes a suitable substrate for the XPF endonuclease. In the presence of Zalypsis, we also observed a dose-dependent stimulation of the OS/CCG^h cleavage (Figure 4F, lanes 20–23). However, the incision located at position A_{-3} on this strand predominated over that appearing at position A_{-4} (lanes 22–23) when compared to the results obtained in the presence of trabectedin (lane 17). Interestingly, in the case of Zalypsis, RPA partially inhibited the XPF-dependent incision on the OS/CCG^h strand. This illustrates that our system is able to translate the structural differences between the two drugs into macroscopic observables. Altogether, the above data show how trabectedin and Zalypsis promote the formation of a DNA structure that allows a specific 5′ XPF-dependent incision on the OS/CCG^h DNA, i.e., the strand opposite to that containing the drug adduct.

Trabectedin and Zalypsis Induce DNA Strand Breaks In Vivo

The trabectedin- and Zalypsis-dependent induction of DNA strand breaks was evaluated in HeLa cells using the comet assay, which is based on the alkaline lysis of labile DNA at sites of damage (Collins et al., 2008). DNA from cells that have accumulated damage appears as fluorescent comets with tails of DNA fragmentation or unwinding, whereas normal, undamaged DNA does not migrate far from the origin (Figure 5A). We next investigated how the silencing of XPF (using HeLa cells that stably express siRNA against XPF, HeLa siXPF) would affect the generation of DNA breaks. Both HeLa siCtrl and HeLa siXPF cells were treated with trabectedin and Zalypsis at 10 and 100 nM for 15 hr. We observed a clear concentration-dependent increase in DNA strand breaks in both cell lines following treatment with either drug (Figures 5B and 5C). Of note, HeLa siXPF cells showed a lower amount of DNA breaks after drug treatment compared to HeLa siCtrl cells. With trabectedin these differences were observed only at 100 nM ($p < 0.01$), whereas with Zalypsis, they were observed at the two drug concentrations tested (10 and 100 nM; $p < 0.01$). Notably, no major differences in formation of γ-H2AX foci, a DSB marker, were found upon drug treatment in siCtrl and siXPF cells (Figure S3), consistent with earlier reports for mitomycin C (Niedernhofer et al., 2004). Altogether, our results suggest that XPF endonuclease is involved in the formation of drug-induced single-strand DNA breaks (SSBs).

DISCUSSION

We and others showed that both trabectedin and Zalypsis inhibit the growth of a variety of cancer cell lines and tumor cells (D'Incalci and Jimeno, 2003; Leal et al., 2009; Ocio et al., 2009) (Table 1). However, the cell type-sensitivity pattern for trabectedin is different from that observed for Zalypsis because the former drug shows selectivity against sarcoma, breast, ovarian, and lung tumor cell lines, with head-and-neck and colon tumor cells being slightly less sensitive, whereas the latter is more potent against leukemia and stomach tumor cell lines, with breast cancer cells being slightly less sensitive than the average (Koeppel et al., 2004; Poindessous et al., 2003). The aim of the present work is to improve our understanding of the particular and specific mode of action of these drugs and highlight any differences between the two.

Trabectedin and Zalypsis Mimic "DNA Interstrand Crosslinkers"

The first step in the binding of a typical bifunctional agent to DNA is the formation of a monoadduct, and this is followed by a second reaction, on either the same strand (intrastrand) or the complementary strand (interstrand), leading to the formation of a crosslink (Muniandy et al., 2010). By covalently linking the two complementary strands of the double helix, ICLs prevent DNA melting, transcription, and replication: hence, they are considered as the most toxic lesions induced by chemotherapeutic agents (Ben-Yehoyada et al., 2009). Because the thermal stability of a DNA duplex increases substantially by the presence of a single bonded trabectedin or Zalypsis molecule (Casado et al., 2008; Leal et al., 2009), this type of monoadduct could represent the same complex challenge to DNA repair mechanisms, as does an ICL. Here, we produce additional evidence that the adducts formed by both trabectedin and Zalypsis functionally mimic a typical ICL despite the fact that these drugs bind covalently to only one DNA strand, their interaction with the OS being only through van der Waals contacts and hydrogen bonds. These ICL-mimetic properties were first illustrated by footprinting assays that demonstrated how the drugs specifically protect the high-affinity CGG sequence, as well as the complementary triplet in the OS, from DNase I digestion (Figure 2). The observed differences in the DNase I footprinting patterns, together with subtly distinct effects in the transcription and repair assays (Figures 3 and 4), reflect the structural variations in the drug substituent that protrudes out of the minor groove in the covalent complexes formed with dsDNA, i.e., a tetrahydroisoquinoline for trabectedin (García-Nieto et al., 2000) and a trifluorocinnamic moiety for Zalypsis (Leal et al., 2009). Because these drugs "interact," albeit in different ways, with both DNA strands, it is not surprising that NER was ineffective in removing the adducts (data not shown), contrary to what is seen with UV- or cisplatin-damaged DNA: neither XPC, XPE, nor XPA would recognize and unwind the damaged DNA (Aboussekhra et al., 1995). Likewise, XPC is inefficient in recognizing the highly tumorigenic benzo[a]pyrene diol epoxide DNA adducts that are known to weaken Watson-Crick hydrogen bonding and base-base stacking interactions and give rise to local thermal destabilization of the double helix (Schinecker et al., 2003).

Trabectedin and Zalypsis Promote DNA Cleavage by XPF

Further support for the ICL-mimetic properties of the drug-DNA adducts studied was obtained when we discovered the

all NTPs including GTP. The RNA transcripts were analyzed on a 5% denaturing urea gel, and quantifications of the full-length transcripts (FLT, +223 bp) were performed. Trabectedin or Zalypsis-mediated stops were detected close to the 5′-CGG-3′ containing region (+113 nt). The drug-induced RNA pol II arrests are indicated in brackets on the sequence on the right of each gel.

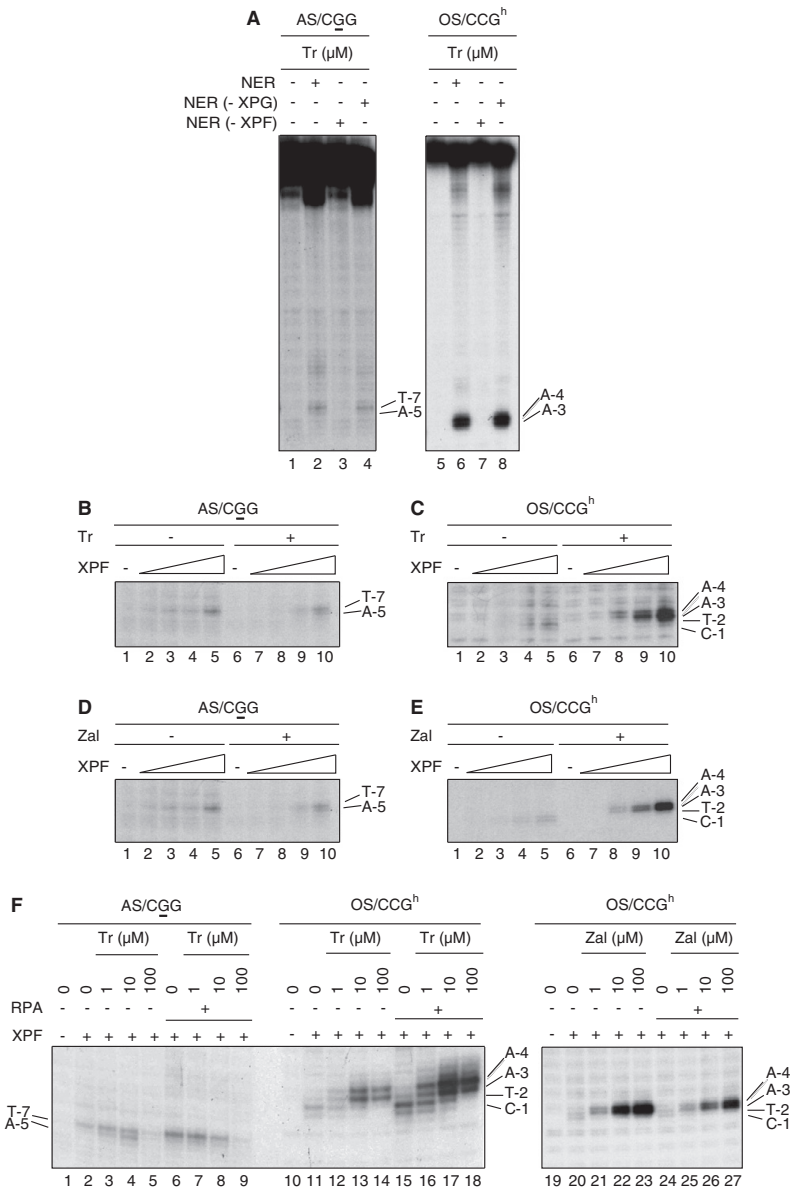


Figure 4. Trabectedin- and Zalypsis-Dependent Effects on DNA Incision

(A) Characterization of the incision on the AS/CGG and the OS/CCG^h pretreated with trabectedin and incubated with NER factors in the presence or absence of XPG and XPF.

XPF-dependent incision on either the AS/CGG or the OS/CCG^h in the presence or absence of either trabectedin (B and C) or Zalypsis (D and E). (F) RPA-mediated effect on the trabectedin (right panel)- and Zalypsis (left panel)-dependent incision using the AS/CGG and the OS/CCG^h substrates.

See also Figure S2.

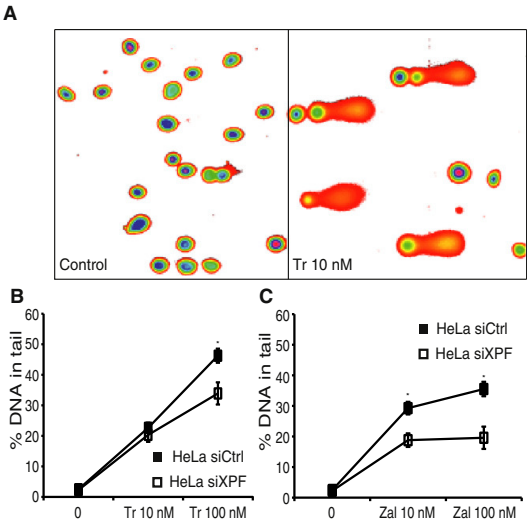


Figure 5. Trabectedin and Zalypsis Induce DNA Breaks In Vivo

(A) Representative images of damaged DNA in the comet assay in untreated and trabectedin-treated (for 12 hr) HeLa cells. HeLa siCtrl and HeLa siXPF cells were analyzed through comet assay followed by exposure to (B) trabectedin and (C) Zalypsis at 10 and 100 nM during 15 hr. Cells were treated, washed with PBS, detached mechanically, and included in low melting point agarose. After lysis, cells were subjected to electrophoresis, and the comets were stained with SYBR Green. DNA quantitation in the comets was performed with TriTek CometScore Freeware version 1.5 ($p < 0.001$). Pictures were taken with a Leica DM IRM fluorescence microscope equipped with a DFC 340 FX digital camera (Leica, Wetzlar, Germany). A minimum of 200 cells was analyzed, and the experiments were repeated in duplicate. The Mann-Whitney U test was used for the statistical analysis. See also Figures S3 and S4.

enhanced cutting of the DNA substrates containing trabectedin or Zalypsis by XPF, a structure-specific endonuclease that preferentially cleaves DNA duplexes adjacent to a 3′ ss flap (Mocquet et al., 2007) and that has been shown to recognize partially unwound structures near a psoralen-induced ICL (de Laat et al., 1998). Drug-induced DNA strand breaks in living cells were assessed by the comet assay and γ-H2AX staining (Figure 5; Figure S3). The finding that SSBs decreased following silencing of XPF expression strongly suggests that the drug-bonded DNA provides an accurate 3D substrate that is recognized and acted upon by the XPF endonuclease, a key player in multiple steps of ICL repair (Kuraoka et al., 2000; Soares et al., 2007).

The extreme sensitivity to trabectedin of HR-deficient cells, which is similar to that evoked by the minor groove-binding and ICL-forming mitomycin C (Rahn et al., 2010), is indeed suggestive that the type of monoadducts formed by these tetrahydroisoquinolines can functionally behave like an ICL. In fact when DNA strand separation is effectively prevented, replication and transcription forks stall, and this event is known to trigger the recruitment of XPF/ERCC1, together with RPA, for 5′ cleavage of the damaged DNA. Support for this interpretation was obtained from MD simulations results, which clearly showed that the melting of the target oligonucleotide at, and on both sides of, the central CGG triplet is hampered in the presence of the bonded drugs (Figure S2). These observations imply that trabectedin and Zalypsis contribute to extending the length of the Watson-Crick base-paired central region and, therefore, enhance XPF recognition and binding by stabilizing a ds/ss discontinuity and one or two 3′ DNA flaps.

To account for the positions of the incisions catalyzed by XPF when either trabectedin or Zalypsis is bonded to the CGG target site, we built molecular models of human XPF/ERCC1 in complex with 5′-$(TAA)_2$CGG$(ATT)_2$-3′. The DNA-binding domain (DBD) of ERCC1 recognizes dsDNA, and the nuclease domain of XPF cuts one strand of this dsDNA at the 5′ side of a junction with ssDNA. The exact cleavage position varies from 2 to 8 nt away from the junction (Mu et al., 2000), and it appears that the ssDNA arm protruding in the 3′ direction (the "3′-flap" or "3′-overhang") is the only requisite for the positioning of the incisions carried out by XPF/ERCC1. Our modeled structures indicate that two different cleavable complexes need to be formed to account for the cuts on either the AS/CGG or OS/CCG^h of the DNA molecule, but the observation of an incision only at the OS/CCG^h indicates that only one is feasible in the presence of the drugs (Figure 4). We note that to cleave the AS/CGG, the required ssDNA region favored by the TA surroundings could start forming at the first A located 3′ to the central CGG dsDNA stretch. However, the presence of a drug molecule covalently bonded to the middle (underlined) guanine in this triplet hampers DNA melting and, hence, cutting of this strand by the nuclease, which cannot find the mandatory 3′ flap. On the contrary, the ssDNA region in the OS can start forming following the T placed 3′ to the CCG^h triplet, whereas the dsDNA stretch will comprise the central CGG/CCG^h plus a variable number of paired bases 5′ to the first C (in CCG^h) that is longer for the trabectedin adduct, and shorter for the Zalypsis adduct (Figures 6A and 6B). Of note, the structural differences between the two drugs studied here are also found in this region, with trabectedin establishing an additional H bond between the phenolic oxygen in the protruding tetrahydroisoquinoline and the phosphodiester backbone (García-Nieto et al., 2000) and Zalypsis sticking out a trifluorocinnamic moiety into the solvent/protein environment.

Although the comet assay showed that the SSBs depended on XPF (Figure 5), the silencing of individual NER factors (XPG or XPF/ERCC1) rendered the cells neither resistant nor hypersensitive to these drugs (Figure S4). These observations suggest that inducing the XPF-dependent DNA cuts is not the main cytotoxic mechanism harnessed by these drugs and that their activity should follow an alternative pathway.

Trabectedin and Zalypsis Inhibit RNA pol II Transcription

Finally, extra support for the ICL-mimetic hypothesis was gained when we showed that DNA damage caused by both drugs obstructs the expression of activated genes (Figure 1). The binding of these drugs to particular sequence triplets, located within certain responsive elements of the promoter of a given gene, might prevent the formation of the transcription complex. This was documented by challenge experiments between trabectedin and the transcription factor Sp1 competing to bind to GC-rich sites (Figure 2; Figure S1). Our results additionally

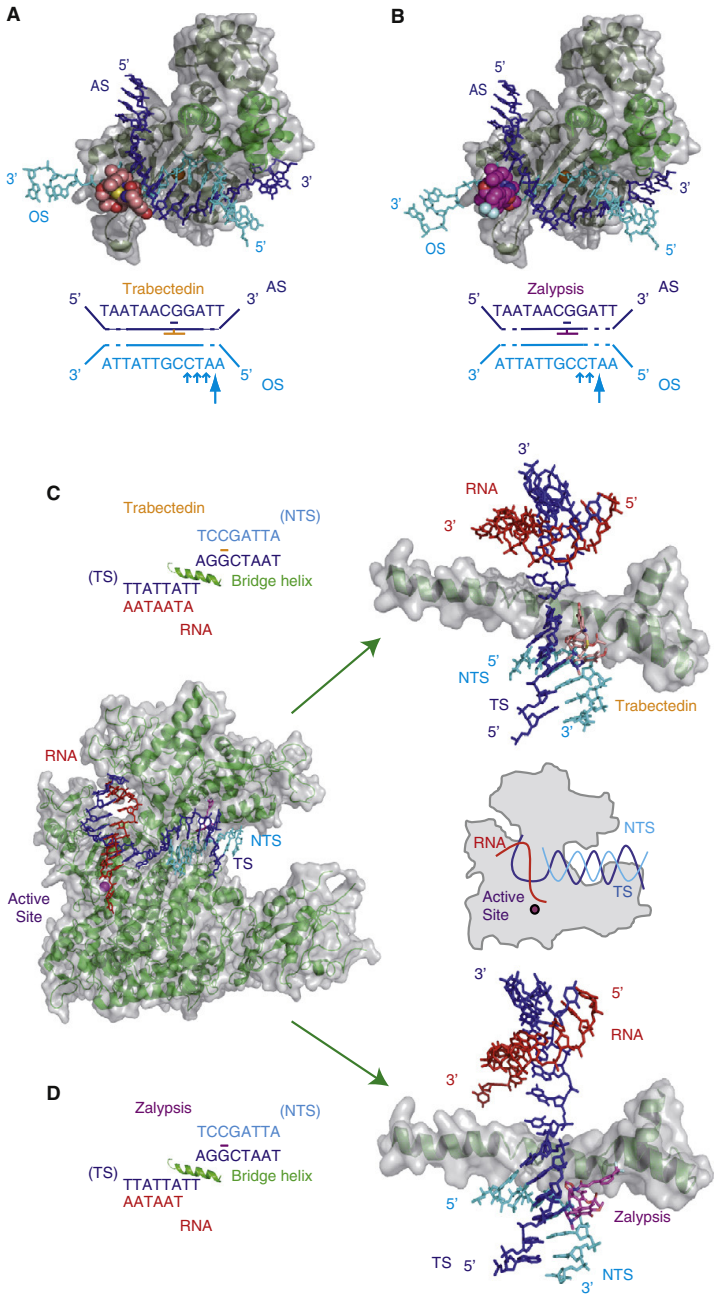


Figure 6. Impact of Trabectedin and Zalypsis on DNA Incision and Transcription Processes

Trabectedin (A) and Zalypsis (B) stimulate an XPF-dependent incision on the OS/CCG^h, i.e., the strand opposite to that harboring the drug adducts. Trabectedin (C) and Zalypsis (D) block RNA pol II progression when the drug is located either on the TS or the NTS. Only the complex structure containing the longest possible RNA transcript is represented for each drug. The schematic representation of RNA pol II is based on Figure 4 from Cramer (2004).

demonstrate that trabectedin and Zalypsis cause RNA pol II stalling regardless of whether the covalent adduct is present on the TS or the NTS (Figure 3). Such property supports the ICL-mimetic behavior of the drug-DNA adducts, which likely underlies the potent cytotoxicity of these compounds, and is in stark contrast with observations made for other adducts (de Laat et al., 1998), which do not have such an effect on transcription elongation when placed on the NTS. It must be kept in mind that the foremost DNA adduct for cisplatin is the 1,2-GG intrastrand crosslink and that less than 8% of cisplatin lesions consist of ICL (Frit et al., 2002; Schinecker et al., 2003). Therefore, the major difference between these two types of adducts is that the key lesion involving cisplatin affects the major groove of two contiguous guanines in the same strand, whereas in the case of the minor groove-binding tetrahydroisoquinolines studied here, the presence of the covalently bonded drug affects both strands and stabilizes the duplex structure. As a consequence, DNA strand separation is hampered regardless of which strand bears the lesion, and RNA pol II is arrested during transcription elongation. This finding could be relevant to the mechanism of cytotoxicity exerted by these drugs.

From the structural point of view, we provide an explanation in atomic detail for the differential arrest of RNA pol II synthesis (Figures 6C and 6D) in light of available crystal structures of RNA pol II elongation complexes (Huang et al., 1995; Stehlikova et al., 2002). Indeed, the number of base pairs between the drug adducts and the ssDNA-RNA hybrid past the polymerase active site was found to be slightly different in the modeled complexes involving each separate drug and to be in good accordance with the distinct sizes obtained for the RNA transcripts. Thus, three alternative models for the trabectedin-DNA-RNA pol II complex are possible,

depending on the relative location of the covalently modified guanine with respect to the active site, but only two in the case of Zalypsis due to the extended protrusion of the trifluorocinnamic moiety giving rise to steric clashes with the protein and precluding further movement of RNA pol II along the DNA template. For simplicity we present only one structure of the putative complex containing the DNA-bonded drug and the longest possible RNA transcript (Figures 6C and 6D). Because translocation over the bridge helix is prevented due to the presence of the drug on either strand, completion of the next nucleotide addition cycle is not feasible, and the RNA transcript is prematurely terminated.

In conclusion our study shows that the anticancer compounds trabectedin and Zalypsis, upon binding to DNA, give rise to a lesion that functionally mimics an ICL even though these drugs are covalently linked to only one DNA strand. These drugs promote XPF/ERCC1-dependent DNA breaks, inhibit RNA synthesis regardless of which DNA strand contains the adduct, and may prevent binding of Sp1, as shown here, and/or other transcription factors. However, it still remains to be seen whether these ICL-mimicking adducts can also preclude or stimulate the recruitment of one or more of the HR proteins that are likely to participate in the removal of these covalent modifications from damaged DNA.

SIGNIFICANCE

Exposure of cells to trabectedin or Zalypsis eventually gives rise to DNA double-strand breaks that do not originate directly from drug binding to the amino group of selected guanines in the minor groove of DNA but are formed during the processing of the resulting covalent adducts by the DNA repair machinery. Intriguingly, cells defective in homologous recombination (HR) repair are particularly sensitive to these agents, whereas cells deficient in nucleotide excision repair (NER) are more resistant to trabectedin than their NER-proficient counterparts. Using well-defined in vitro and in vivo assays, we now show that the DNA adducts formed by these two potent anticancer agents are not removed by the NER machinery, although the XPF/ERCC1 nuclease is able to cleave the strand opposite the lesion in a drug-dependent way. Furthermore, these adducts compete with DNA-binding factors like Sp1 for their specific DNA-binding sites and also arrest elongating RNA polymerase II at the site of the lesion regardless of whether the compounds are covalently attached to the template or the nontemplate strand. All of these results, which are supported by structural models, are consistent with the view that a single adduct formed by any of these drugs in just one DNA strand can functionally mimic an interstrand crosslink insofar as it can hamper or prevent repair and strand separation very effectively. Therefore, it would seem that stalling transcription and replication forks is likely to be a hallmark of the antiproliferative action of these tetrahydroisoquinoline alkaloids. Moreover, identification of the genes involved in their mechanism of action, and particularly in the DNA damage response, that are deregulated in a given tumor may help to select the subsets of patients that will benefit the most from this type of chemotherapeutic agent.

EXPERIMENTAL PROCEDURES

Activated Transcription

HeLa cells were grown to 50% confluency before being subjected to media containing charcoal-treated FCS (10%) and devoid of phenol red for 16 hr. Then, cells were pretreated for 1 hr with either drug, before t-RA induction (10 μM for 6 hr). Reverse transcription was done using 1 μg of total RNA (GenElute Mammalian Total RNA Miniprep Kit; Sigma-Aldrich), oligo-$dT_{(15)}$, and SuperScript II Reverse Transcriptase (Invitrogen).

Design of the DNA Templates

The SV40 promoter was PCR amplified from pGL3 vector (Promega) using a 5′ radioactively labeled primer with Optikinase (USB Corp.) and 100 nM γ^{32}P-dATP (6000 Ci/mmol), a biotinylated primer, and resolved on a 5% native 1× TBE polyacrylamide (PAA) gel. The DNA substrates containing a single CGG site for drug binding were generated by heating the forward ($A[TAA]_{20}$**CGG**ATTATT) and the reverse ($AA[TAA]_{20}$**CCG**TTATT) primers at 95°C for 5 min and slowly cooling down to 4°C. The Klenow fragment was used for fill-in synthesis during 30 min at 25°C in the presence of 1 mM dNTPs. The purified amplicon (QIAGEN PCR) was cloned into pGEM-T plasmid (Invitrogen) leading to pGEM-T 1xCGG containing the following sequence: $AT(TAA)_{13}$**CGG**$(ATT)_{19}T$. The AS/CGG and the OS/CCG^h-labeled templates were generated by digesting 100 μg pGEM-T 1xCGG using SpeI for AS/CGG, and NcoI for OS/CCG^h (50 U for 3 hr). Biotinylated and radioactive nucleotides were filled in using 30 U Klenow fragment, 50 μM dATP/dGTP, 20 μM Biotin-16-dUTP (Roche Diagnostics), and 130 nM α^{32}P-dCTP (3000 Ci/mmol) (GE Healthcare) at 37°C, followed by a second enzyme digestion (NcoI for AS/CGG and SpeI for OS/CCG^h) overnight at 37°C. The probe was resolved on an 8% native PAA gel.

The pBL-CAX TS.500 (Charlet-Berguerand et al., 2006) was digested by XbaI and SalI. A G-less cassette was PCR amplified (Sawadogo and Roeder, 1985) using High Fidelity polymerase (Roche), the XbaI-restriction site containing specific primers, and then digested by XbaI and SalI, gel purified, and ligated using NEB's Quick Ligation Kit protocol. The remaining G in the XbaI restriction site was changed to a C by performing site-directed mutagenesis using Phusion polymerase (Finnzymes) and specific primers. After gel purification, the PCR amplicon was ligated using T4 ligase (NEB) leading to the pBL-Gless96-CAX TS.500.

To insert the single drug-binding site into the pBL-G-less96-CAX TS.500, we deleted the sequence located after the G-less cassette using primers with the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Then, we inserted the single drug-binding triplet into the plasmid using primers leading to the pBL-Gless96-CAX TS.500 CGG/TS and pBL-Gless96-CAX TS.500 CGG/NTS.

DNase I Footprinting Assays

The 5′ end ^{32}P-labeled (20,000 cpm) SV40 promoter template was immobilized on 2.01×10^5 magnetic beads (Dynabeads), incubated for 30 min at RT in buffer A (Tris/HCl 10 mM [pH 7.6], glycerol 10%, EDTA 1 mM, DTT 0.5 mM, KCl 50 mM, and 0.05% NP40), and then incubated either with 1–5 μg/μl of HeLa Sp1 containing NE or with 10–20 μM trabectedin for 30 min. For the competition assay the above substrate was incubated first with either 3 μg/μl of NE or 20 μM trabectedin for 30 min, and vice versa in a second incubation. Low-salt concentration washes were then performed to remove any DNA-unbound molecules before subjecting the DNA template to digestion using 1 ng/μl DNase I (Sigma-Aldrich) for 45 s at RT in buffer A supplemented with 5 mM $MgCl_2$, and 20 μg/ml polydI-dC.

The 3′ end-labeled (20,000 cpm) pGEM-T 1xCGG-derived AS/CGG or the OS/CCG^h substrates were bound to magnetic beads and treated with the indicated drug concentrations. Unspecific bound molecules were removed by gently washing with buffer A before DNase I digestions. The reaction was stopped, and the purified nucleic acids were resolved on an 8% denaturing urea-PAA gel.

In Vitro Transcription

The immobilized pBL-Gless96-CAX TS.500 CGG/TS or the pBL-Gless96-CAX TS.500 CGG/NTS PCR-amplified product was incubated with the drug. After extensive washing, the template was incubated with all the basal transcription factors including RNA pol II for 30 min at 25°C (Charlet-Berguerand et al.,

2006); then 300 μM ATP/UTP, 10 μM cold CTP, and 5 μCi α-^{32}P-CTP (3000 Ci/mmol) were added for 30 min at 25°C.

For the investigation of the drugs' effects on transcription elongation, the template was first incubated with all the basal transcription factors and rNTPs except GTP. After extensive washes (Riedl et al., 2003), the ternary elongation complex was incubated for 30 min with the drug in buffer A containing rNTPs (in the absence of GTP). After washes with buffer A, 300 μM rNTPs and 6.5 mM $MgCl_2$ were added. Nucleic acids were analyzed on a 5% denaturing urea-PAA gel.

Comet Assay

For the comet assay a single-cell gel electrophoresis assay was used (Trevigen's CometAssay) following the manufacturer's instructions after treatment of cells for 12 hr with the appropriate concentration of trabectedin and Zalypsis. Pictures were taken with a Leica DM IRM fluorescence microscope equipped with a DFC 340 FX digital camera (Leica, Wetzlar, Germany). Quantitation of the DNA in the tails of the comets was performed with Adobe Photoshop CS3 (Adobe Systems Inc., San Jose, CA). For each condition 30 cells were analyzed, and the experiments were repeated several times.

Molecular Modeling

XPF/ERCC1 in Complex with Drug-Bound DNA

The nuclease domain of human XPF was modeled using the coordinates of *Pyrococcus furiosus* Hef as a template (PDB ID: 1J23) and the sequence alignment (Kelley and Sternberg, 2009). The resulting structure and the DBDs of XPF and ERCC1 (PDB ID: 2A1J) were superimposed onto their DBD counterparts in the *Aeropyrum pernix* XPF homodimer in complex with dsDNA (PDB ID: 2BGW). The root-mean-square deviation between the Cα traces of human and crenarchaeal XPF nuclease domain was 1.1 Å, whereas that between their DBDs was 1.26 Å (53 atoms). The *A. pernix* XPF-dsDNA structure comprises two XPF protomers and one DNA duplex, bound to the nuclease and $(HhH)_2$ domains of the first protomer (site I). Because the $(HhH)_2$ domain of the second protomer (which corresponds to ERCC1) is also potentially capable of binding DNA (site II), a second dsDNA molecule was built at 90° to the experimentally observed dsDNA using a symmetry operator on the DNA in the unit cell as reported in Newman et al. (2005). The DNA-drug adducts were modeled as reported for trabectedin in PDB/OCA entry 1EZH (García-Nieto et al., 2000).

Drug-DNA-RNA Pol II Elongation Complex

The crystal structure of the RNA pol II elongation complex (Sydow et al., 2009), containing both a DNA template and an RNA transcript (PDB ID: 3HOV), was used upon appropriate substitution of DNA bases so as to model the oligonucleotide employed in our experiments. The position of the CGG triplet to which trabectedin and Zalypsis bind in relation to the polymerase active site was systematically varied stepwise to define those locations in which the drug would give rise to steric clashes with the protein. All primer sequences are available upon request.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at doi:10.1016/j.chembiol.2011.06.007.

ACKNOWLEDGMENTS

We dedicate this work to our collaborator Jonathan Gintz. We also acknowledge Professor J.M. Fernández-Sousa, Dr. Carmen Cuevas, and Dr. Frédéric Coin for fruitful discussions. We are grateful to C. Braun, A. Larnicol, B. Prève, and G. Santamaria Nunez for purification of repair and transcription factors and cell toxicity assays. We want also to thank Denis Biard for the HeLa siXPF. This work was supported by l'Association de la Recherche contre le Cancer (ARC n°3153), la Ligue contre le Cancer, INCA, and an ERC Advanced grant (to J.M.E.), and Comisión Interministerial de Ciencia y Tecnología (SAF2006-12713-C02-02 and SAF2009-13914-C02-02) and Comunidad de Madrid (S-BIO/0214/2006) (to F.G.). S.F. was supported by an ARC and Marie Curie fellowship, and C.G. is a recipient of a fellowship from the Ministère de la Recherche and ARC. M.M.-D. is an employee of PharmaMar, and C.M.G. is an employee and shareholder of Pharma Mar.

Received: March 12, 2011
Revised: June 3, 2011
Accepted: June 17, 2011
Published: August 25, 2011

REFERENCES

Aboussekhra, A., Biggerstaff, M., Shivji, M.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protić, M., Hübscher, U., Egly, J.M., and Wood, R.D. (1995). Mammalian DNA nucleotide excision repair reconstituted with purified protein components. Cell *80*, 859–868.

Ben-Yehoyada, M., Wang, L.C., Kozekov, I.D., Rizzo, C.J., Gottesman, M.E., and Gautier, J. (2009). Checkpoint signaling from a single DNA interstrand crosslink. Mol. Cell *35*, 704–715.

Bustamante, C., Cheng, W., and Mejia, Y.X. (2011). Revisiting the central dogma one molecule at a time. Cell *144*, 480–497. Erratum: (2011). Cell *145*, 160.

Casado, J.A., Río, P., Marco, E., García-Hernández, V., Domingo, A., Pérez, L., Tercero, J.C., Vaquero, J.J., Albella, B., Gago, F., and Bueren, J.A. (2008). Relevance of the Fanconi anemia pathway in the response of human cells to trabectedin. Mol. Cancer Ther. *7*, 1309–1318.

Charlet-Berguerand, N., Feuerhahn, S., Kong, S.E., Ziserman, H., Conaway, J.W., Conaway, R., and Egly, J.M. (2006). RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. EMBO J. *25*, 5481–5491.

Collins, A.R., Oscoz, A.A., Brunborg, G., Gaivão, I., Giovannelli, L., Kruszewski, M., Smith, C.C., and Stetina, R. (2008). The comet assay: topical issues. Mutagenesis *23*, 143–151.

Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. Curr. Opin. Genet. Dev. *14*, 218–226.

Cuevas, C., and Francesch, A. (2009). Development of Yondelis (trabectedin, ET-743). A semisynthetic process solves the supply problem. Nat. Prod. Rep. *26*, 322–337.

Damia, G., Silvestri, S., Carrassa, L., Filiberti, L., Faircloth, G.T., Liberi, G., Foiani, M., and D'Incalci, M. (2001). Unique pattern of ET-743 activity in different cellular systems with defined deficiencies in DNA-repair pathways. Int. J. Cancer *92*, 583–588.

de Laat, W.L., Appeldoorn, E., Jaspers, N.G.J., and Hoeijmakers, J.H.J. (1998). DNA structural elements required for ERCC1-XPF endonuclease activity. J. Biol. Chem. *273*, 7835–7842.

De Silva, I.U., McHugh, P.J., Clingen, P.H., and Hartley, J.A. (2000). Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. Mol. Cell. Biol. *20*, 7980–7990.

D'Incalci, M., and Galmarini, C.M. (2010). A review of trabectedin (ET-743): a unique mechanism of action. Mol. Cancer Ther. *9*, 2157–2163.

D'Incalci, M., and Jimeno, J. (2003). Preclinical and clinical results with the natural marine product ET-743. Expert Opin. Investig. Drugs *12*, 1843–1853.

Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79–87.

Friedman, D., Hu, Z., Kolb, E.A., Gorfajn, B., and Scotto, K.W. (2002). Ecteinascidin-743 inhibits activated but not constitutive transcription. Cancer Res. *62*, 3377–3381.

Frit, P., Kwon, K., Coin, F., Auriol, J., Dubaele, S., Salles, B., and Egly, J.M. (2002). Transcriptional activators stimulate DNA repair. Mol. Cell *10*, 1391–1401.

García-Nieto, R., Manzanares, I., Cuevas, C., and Gago, F. (2000). Increased DNA binding specificity for antitumor ecteinascidin 743 through protein-DNA interactions? J. Med. Chem. *43*, 4367–4369.

Guirouilh-Barbat, J., Redon, C., and Pommier, Y. (2008). Transcription-coupled DNA double-strand breaks are mediated via the nucleotide excision repair and the Mre11-Rad50-Nbs1 complex. Mol. Biol. Cell *19*, 3969–3981.

Guirouilh-Barbat, J., Antony, S., and Pommier, Y. (2009). Zalypsis (PM00104) is a potent inducer of gamma-H2AX foci and reveals the importance of the

 ©2011 Elsevier Ltd All rights reserved

C ring of trabectedin for transcription-coupled repair inhibition. Mol. Cancer Ther. *8*, 2007–2014.

Huang, H., Zhu, L., Reid, B.R., Drobny, G.P., and Hopkins, P.B. (1995). Solution structure of a cisplatin-induced DNA interstrand cross-link. Science *270*, 1842–1845.

Jin, S., Gorfajn, B., Faircloth, G., and Scotto, K.W. (2000). Ecteinascidin 743, a transcription-targeted chemotherapeutic that inhibits MDR1 activation. Proc. Natl. Acad. Sci. USA *97*, 6775–6779.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.

Koeppel, F., Poindessous, V., Lazar, V., Raymond, E., Sarasin, A., and Larsen, A.K. (2004). Irofulven cytotoxicity depends on transcription-coupled nucleotide excision repair and is correlated with XPG expression in solid tumor cells. Clin. Cancer Res. *10*, 5604–5613.

Kuraoka, I., Kobertz, W.R., Ariza, R.R., Biggerstaff, M., Essigmann, J.M., and Wood, R.D. (2000). Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. J. Biol. Chem. *275*, 26632–26636.

Leal, J.F., García-Hernández, V., Moneo, V., Domingo, A., Bueren-Calabuig, J.A., Negri, A., Gago, F., Guillén-Navarro, M.J., Avilés, P., Cuevas, C., et al. (2009). Molecular pharmacology and antitumor activity of Zalypsis in several human cancer cell lines. Biochem. Pharmacol. *78*, 162–170.

Mathis, D.J., and Chambon, P. (1981). The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature *290*, 310–315.

McHugh, P.J., Spanswick, V.J., and Hartley, J.A. (2001). Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. Lancet Oncol. *2*, 483–490.

Minuzzo, M., Marchini, S., Broggini, M., Faircloth, G., D'Incalci, M., and Mantovani, R. (2000). Interference of transcriptional activation by the antineoplastic drug ecteinascidin-743. Proc. Natl. Acad. Sci. USA *97*, 6780–6784.

Mocquet, V., Kropachev, K., Kolbanovskiy, M., Kolbanovskiy, A., Tapias, A., Cai, Y., Broyde, S., Geacintov, N.E., and Egly, J.M. (2007). The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[a]pyrenyl-DNA lesions. EMBO J. *26*, 2923–2932.

Mu, D., Bessho, T., Nechev, L.V., Chen, D.J., Harris, T.M., Hearst, J.E., and Sancar, A. (2000). DNA interstrand cross-links induce futile repair synthesis in mammalian cell extracts. Mol. Cell. Biol. *20*, 2446–2454.

Muniandy, P.A., Liu, J., Majumdar, A., Liu, S.T., and Seidman, M.M. (2010). DNA interstrand crosslink repair in mammalian cells: step by step. Crit. Rev. Biochem. Mol. Biol. *45*, 23–49.

Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P.P., White, M.F., and McDonald, N.Q. (2005). Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. EMBO J. *24*, 895–905.

Niedernhofer, L.J., Odijk, H., Budzowska, M., van Drunen, E., Maas, A., Theil, A.F., de Wit, J., Jaspers, N.G., Beverloo, H.B., Hoeijmakers, J.H., and Kanaar, R. (2004). The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. Mol. Cell. Biol. *24*, 5776–5787.

Ocio, E.M., Maiso, P., Chen, X., Garayoa, M., Alvarez-Fernández, S., San-Segundo, L., Vilanova, D., López-Corral, L., Montero, J.C., Hernández-Iglesias, T., et al. (2009). Zalypsis: a novel marine-derived compound with potent antimyeloma activity that reveals high sensitivity of malignant plasma cells to DNA double-strand breaks. Blood *113*, 3781–3791.

Oku, N., Matsunaga, S., van Soest, R.W., and Fusetani, N. (2003). Renieramycin J, a highly cytotoxic tetrahydroisoquinoline alkaloid, from a marine sponge Neopetrosia sp. J. Nat. Prod. *66*, 1136–1139. Erratum: (2004). J. Nat. Prod. *67*, 526.

Poindessous, V., Koeppel, F., Raymond, E., Comisso, M., Waters, S.J., and Larsen, A.K. (2003). Marked activity of irofulven toward human carcinoma cells: comparison with cisplatin and ecteinascidin. Clin. Cancer Res. *9*, 2817–2825.

Pommier, Y., Kohlhagen, G., Bailly, C., Waring, M., Mazumder, A., and Kohn, K.W. (1996). DNA sequence- and structure-selective alkylation of guanine N2 in the DNA minor groove by ecteinascidin 743, a potent antitumor compound from the Caribbean tunicate *Ecteinascidia turbinata*. Biochemistry *35*, 13303–13309.

Rahn, J.J., Adair, G.M., and Nairn, R.S. (2010). Multiple roles of ERCC1-XPF in mammalian interstrand crosslink repair. Environ. Mol. Mutagen. *51*, 567–581.

Riedl, T., Hanaoka, F., and Egly, J.M. (2003). The comings and goings of nucleotide excision repair factors on damaged DNA. EMBO J. *22*, 5293–5303.

Rinehart, K.L., Holt, T.G., Fregeau, N.L., Keifer, P.A., Wilson, G.R., Perun, T.J., Jr., Sakai, R., Thompson, A.G., Stroh, J.G., Shield, L.S., et al. (1990). Bioactive compounds from aquatic and terrestrial sources. J. Nat. Prod. *53*, 771–792.

Sawadogo, M., and Roeder, R.G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell *43*, 165–175.

Schinecker, T.M., Perlow, R.A., Broyde, S., Geacintov, N.E., and Scicchitano, D.A. (2003). Human RNA polymerase II is partially blocked by DNA adducts derived from tumorigenic benzo[c]phenanthrene diol epoxides: relating biological consequences to conformational preferences. Nucleic Acids Res. *31*, 6004–6015.

Scott, J.D., and Williams, R.M. (2002). Chemistry and biology of the tetrahydroisoquinoline antitumor antibiotics. Chem. Rev. *102*, 1669–1730.

Soares, D.G., Escargueil, A.E., Poindessous, V., Sarasin, A., de Gramont, A., Bonatto, D., Henriques, J.A., and Larsen, A.K. (2007). Replication and homologous recombination repair regulate DNA double-strand break formation by the antitumor alkylator ecteinascidin 743. Proc. Natl. Acad. Sci. USA *104*, 13062–13067.

Stehlikova, K., Kostrhunova, H., Kasparkova, J., and Brabec, V. (2002). DNA bending and unwinding due to the major 1,2-GG intrastrand cross-link formed by antitumor cis-diamminedichloroplatinum(II) are flanking-base independent. Nucleic Acids Res. *30*, 2894–2898.

Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol. Cell *34*, 710–721.

Takebayashi, Y., Pourquier, P., Zimonjic, D.B., Nakayama, K., Emmert, S., Ueda, T., Urasaki, Y., Kanzaki, A., Akiyama, S.I., Popescu, N., et al. (2001). Antiproliferative activity of ecteinascidin 743 is dependent upon transcription-coupled nucleotide-excision repair. Nat. Med. *7*, 961–966.

Tavecchio, M., Simone, M., Erba, E., Chiolo, I., Liberi, G., Foiani, M., D'Incalci, M., and Damia, G. (2008). Role of homologous recombination in trabectedin-induced DNA damage. Eur. J. Cancer *44*, 609–618.

Zawel, L., Kumar, K.P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. Genes Dev. *9*, 1479–1490.

Zewail-Foote, M., and Hurley, L.H. (1999). Ecteinascidin 743: a minor groove alkylator that bends DNA toward the major groove. J. Med. Chem. *42*, 2493–2497.

Zewail-Foote, M., Li, V.S., Kohn, H., Bearss, D., Guzman, M., and Hurley, L.H. (2001). The inefficiency of incisions of ecteinascidin 743-DNA adducts by the UvrABC nuclease and the unique structural feature of the DNA adducts can be used to explain the repair-dependent toxicities of this antitumor agent. Chem. Biol. *8*, 1033–1049.

Chemistry & Biology *18*

Supplemental Information

XPF-Dependent DNA Breaks and RNA Polymerase II Arrest Induced by Antitumor DNA Interstrand Crosslinking-Mimetic Alkaloids

Sascha Feuerhahn, Christophe Giraudon, Marta Martínez-Díez, Juan A. Bueren-Calabuig, Carlos M. Galmarini, Federico Gago, and Jean-Marc Egly

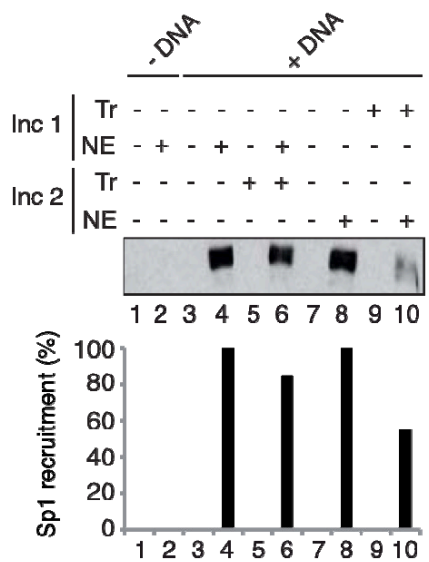


Figure S1

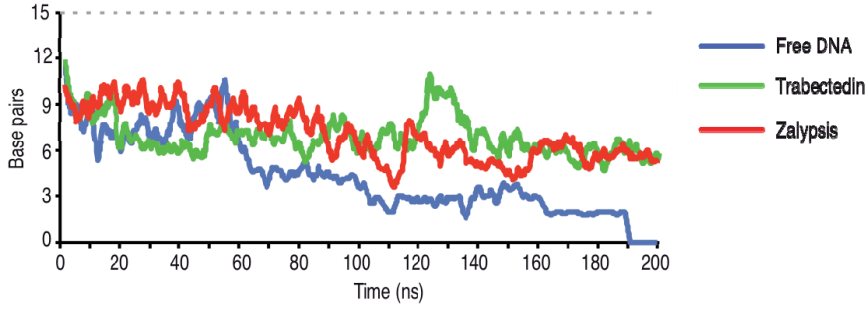

15
12
9
6
3
0
Base pairs
0
20
40
60
80
100
120
140
160
180
200
Time (ns)
Free DNA
Trabectedin
Zalypsis

Figure S2

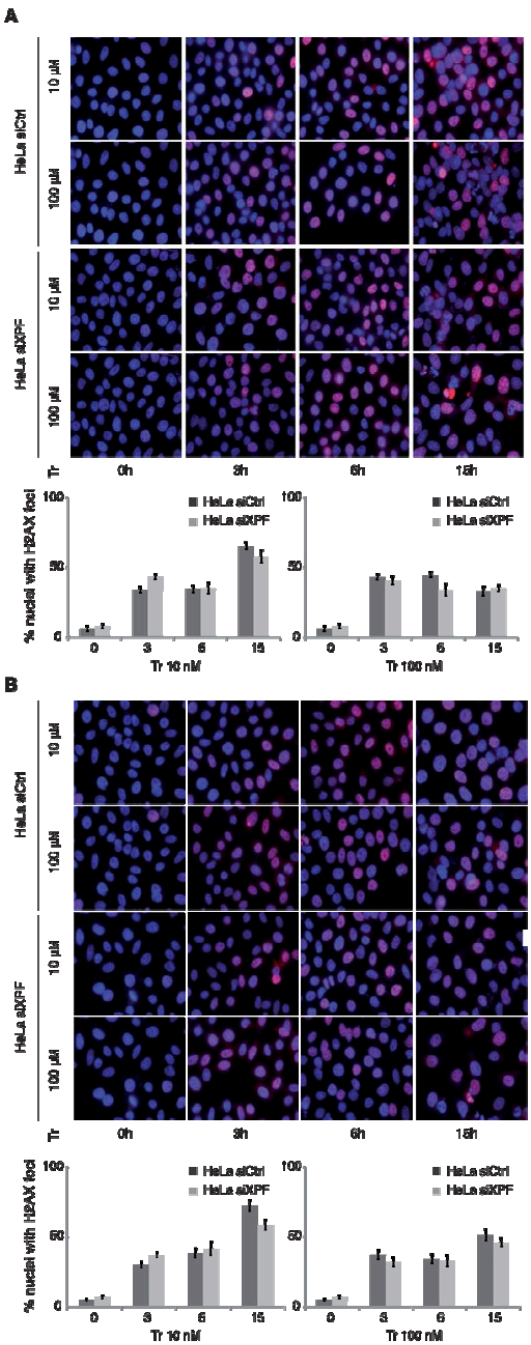

A
HeLa siCtrl
HeLa siXPF
10 µM
100 µM
Tr
0h
3h
6h
15h
% nuclei with H2AX foci
HeLa siCtrl
HeLa siXPF
Tr 10 nM
Tr 100 nM
B

Figure S3

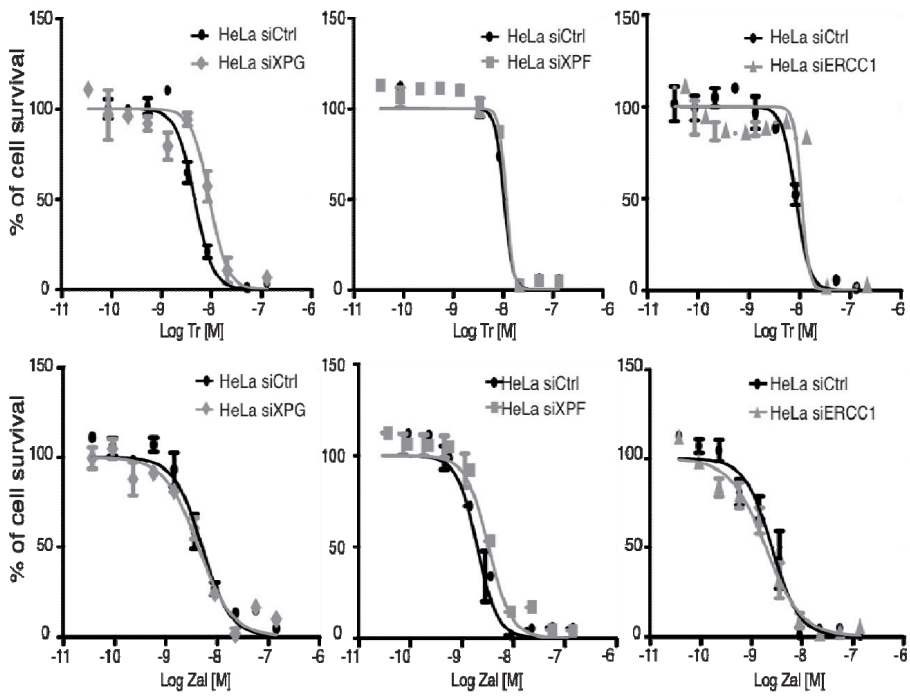

% of cell survival
HeLa siCtrl
HeLa siXPG
HeLa siXPF
HeLa siERCC1
Log Tr [M]
Log Zal [M]

Figure S4

Figure S1, related to Figure 2: Sp1 recruitment on its binding elements-containing SV40 promoter template. 50 ng of the biotin-containing SV40 promoter DNA was bound to magnetic beads (Dynabeads) for 30 min at RT in buffer A (Experimental Procedure section). Then the template was incubated in the same conditions as the challenge footprint assay. Finally Sp1 was detected by western blot technique using a specific antibody (Santa Cruz). When the SV40 promoter template is incubated first HeLa NE and then with 20 μM Trabectedin, Sp1 transcription factor is recruited (compare lanes 4 and 6). However, the SV40 promoter DNA previously treated with Trabectedin strongly decreases Sp1 recruitment (compare lanes 8 and 10).

Figure S2, related to Figure 4: DNA base-pair disruption along the molecular dynamics simulations of free and drug-bound DNA in 0.1 M NaCl at 400 K. The stabilities of the explicitly solvated free DNA duplex or the duplexes containing a Trabectedin or Zalypsis adduct in the central region were assessed during 200 ns of unrestrained molecular dynamics simulations at 400 ºK in the presence of 0.1M NaCl using the AMBER force field (parmbsc0) and program PMEMD (http://ambermd.org/; UCSF, version 10). The trajectories were analyzed in terms of double-strand fraying and peeling by searching for Watson-Crick (WC) paired bases. Our criteria for WC H-bonded pairs were (i) a distance between H-bond donor and acceptor atoms shorter than 3.2 Å and (ii) two H-bonds between paired bases irrespective of their being A:T or G:C. Note the complete loss of all the original base pairs only in free DNA as well as the melting and reannealing events observed in the DNA molecules containing the drug adducts which keep maintaining the central region Watson-Crick paired at the end of the simulation.

Figure S3, related to Figure 5: Induction of γ-H2AX foci in HeLa siCtrl and siXPF cells upon Trabectedin treatment. HeLa siCtrl and siXPF cells were continuously treated with 10nM and 100nM Trabectedin during 15 h (A) or or treated during 1 h with 10nM and 100nM Trabectedin, washed and then incubated during 15 h with culture media without drug (B). Cells were fixed (4% paraformaldehyde), permeabilized (0.5% Triton X-100) and incubated with the primary anti-γ-H2AX monoclonal antibody (Upstate, Temecula, CA) for 1 h at 37 ºC. Then, cells were washed and incubated with the secondary anti-mouse AlexaFluor 594 (Invitrogen, Carlsbad, CA, USA) for 30 min

at 37ºC. Finally the slides were incubated with Hoesch 33342 (Sigma, St Louis, MO, USA) and mounted with Mowiol mounting medium. Pictures were taken with a Leica DM IRM fluorescence microscope equipped with a 100x oil immersion objective and a DFC 340 FX digital camera (Leica, Wetzlar, Germany).

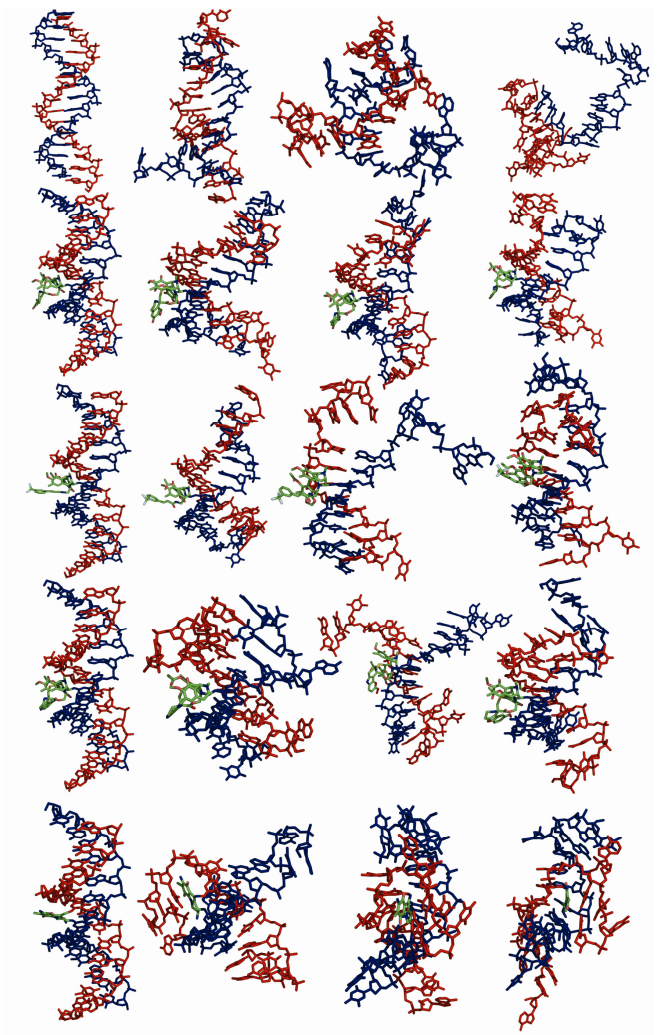
We further assessed whether Trabectedin induced-DNA adducts could eventually give rise to double-strand breaks (DSBs) in living cells by analyzing the formation of γ-H2AX foci as a surrogate indicator of DSB formation. Figure S3 shows typical results found in HeLa siCtrl and HeLa siXPF cells treated with Trabectedin at 10 and 100 nM. Untreated cells displayed a homogenous nuclear staining, with a small fraction of cells containing few nuclei with H foci (Fig. S3). In contrast, after drug treatment, a higher percentage of cells (approximately 50%) were stained for γ-H2AX and each of the positive cells carried a large number of foci, indicating that several DSBs per cell were induced at these drug concentrations. As shown in Figure S3, equimolar concentrations of Trabectedin caused similar levels of DSBs in both cell lines (siCtrl and siXPF).

Figure S4, related to Figure 5: Proliferation inhibitory effects of Trabectedin and Zalypsis on different HeLa cells. HeLa siControl, HeLa siXPG, HeLa siXPF and HeLa siERCC1 were seeded in 96-well trays and treated with dilutions of the drugs (in DMSO) during 72 hours. After addition of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) (Sigma, St Louis, MO, USA), the formed formazan crystals dissolved in DMSO were measured (OD=540nM) with a POLARStar Omega Reader (BMG Labtech, Offenburg, Germany). Determination of IC_{50} values was performed by iterative non-linear curve fitting with the Prism 5.0 statistical software (GraphPad, La Jolla, CA, USA). There were no significant differences between IC_{50} values of HeLa cells used as controls and the silenced cell lines. The mean±SD IC_{50} values for HeLa siXPG, were 9.9±2.2 nM for Trabectedin and 5±0.7 nM for Zalypsis, while for HeLa siXPF cells, they were 6.8±1.9 nM for Trabectedin and 2.6±0.9 nM for Zalypsis and for HeLa siERCC1, 9.5±0.8 nM for Trabectedin and 2.7±0.9 nM for Zalypsis.

ARTÍCULO IV.

"Temperature-induced melting of double-stranded DNA in the absence and presence of covalently bonded antitumour drugs: insight from molecular dynamics simulations"

***Nucleic Acids Res.* 2011 Jul 3. [Epub ahead of print]**



En una molécula estándar de ADN, dos cadenas complementarias de polinucleótidos interactúan entre sí a través de típicos enlaces de hidrógeno de Watson-Crick. La estabilidad de esta doble hélice depende de la naturaleza de las bases y de la secuencia, y puede ser alterada por variaciones en el pH, en la temperatura y en el disolvente, así como por la unión a proteínas. La separación de las hebras también puede verse afectada por la presencia de agentes antitumorales que se unen al ADN como la mitomicina C (MMC), que provoca entrecruzamientos intercatenarios y las tetrahidroisoquinolinas monofuncionales trabectedina (Yondelis®), Zalypsis® y PM01183, todas ellas capaces de formar enlaces covalentes con el grupo amino de determinadas guaninas en el surco menor. Al simular el comportamiento dinámico de un oligodesoxinucleótido de 15 residuos en una solución acuosa de 0,1 M de NaCl a 400 K, tanto en ausencia como en presencia de un aducto de un fármaco en la región central, se observaron importantes diferencias en la cinética de desnaturalización de la hélice que dan fe de la capacidad demostrada de estos fármacos de estabilizar el ADN de doble cadena, horquillas de replicación y transcripción, y reclutar proteínas de unión al ADN. Por otra parte, la estabilización producida por los aductos monofuncionales, cuantificada en términos de pares de bases inalterados, fue mayor que la obtenida por MMC lo que apoya nuestra tesis de que estos agentes se comportan funcionalmente como miméticos de un típico entrecruzador intercatenario de ADN.

Nucleic Acids Research Advance Access published July 3, 2011

*Nucleic Acids Research, 2011, **1–10***
doi:10.1093/nar/gkr512

Temperature-induced melting of double-stranded DNA in the absence and presence of covalently bonded antitumour drugs: insight from molecular dynamics simulations

Juan A. Bueren-Calabuig[1], Christophe Giraudon[2], Carlos M. Galmarini[3], Jean Marc Egly[2] and Federico Gago[1,*]

[1]Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain, [2]Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UdS, BP 163, 67404 Illkirch Cedex, C. U. Strasbourg, France and [3]Cell Biology Department, PharmaMar, Avda. de los Reyes, 1 - Pol. Ind. La Mina, 28770 Colmenar Viejo, Madrid, Spain

Received May 10, 2011; Revised June 3, 2011; Accepted June 7, 2011

ABSTRACT

The difference in melting temperature of a double-stranded (ds) DNA molecule in the absence and presence of bound ligands can provide experimental information about the stabilization brought about by ligand binding. By simulating the dynamic behaviour of a duplex of sequence 5′-d(TAATAACGG ATTATT)·5′-d(AATAATCCGTTATTA) in 0.1 M NaCl aqueous solution at 400 K, we have characterized in atomic detail its complete thermal denaturation profile in <200 ns. A striking asymmetry was observed on both sides of the central CGG triplet and the strand separation process was shown to be strongly affected by bonding in the minor groove of the prototypical interstrand crosslinker mitomycin C or the monofunctional tetrahydroisoquinolines trabectedin (*Yondelis*®), *Zalypsis*® and PM01183®. Progressive helix unzipping was clearly interspersed with some reannealing events, which were most noticeable in the oligonucleotides containing the monoadducts, which maintained an average of 6 bp in the central region at the end of the simulations. These significant differences attest to the demonstrated ability of these drugs to stabilize dsDNA, stall replication and transcription forks, and recruit DNA repair proteins. This stabilization, quantified here in terms of undisrupted base pairs, supports the view that these monoadducts can functionally mimic a DNA interstrand crosslink.

INTRODUCTION

The stability of a DNA double helix containing two complementary polynucleotide chains interacting with each other via Watson–Crick (WC) hydrogen bonds largely depends on its length, base composition, sequence and presence of mismatches, but it can also be strongly perturbed either globally or locally by alterations in pH, temperature and solvent (1) as well as by the binding of proteins such as helicases, polymerases or topoisomerases in living cells (2). In fact, local melting leading to strand separation is necessary for such vital processes as replication, transcription, recombination or repair (3) and can be affected by the presence of non-covalently bound [e.g. mono- and bis-intercalators (4)] or covalently bonded drugs [e.g. cisplatin and nitrogen mustards (5)]. Within the latter category we have focused on several anti-tumour agents (Figure 1) that form a covalent bond with the exocyclic N^2 of select guanines in the DNA minor groove: (i) on the one hand, the well-known mitomycin C (MMC), which upon reductive activation is capable of physically connecting the two strands by simultaneously bonding to the two guanines present at a CpG step in a double-stranded (ds) DNA molecule (6), and (ii) on the other hand, the tetrahydroisoquinoline derivatives (7) *Yondelis*®, *Zalypsis*® and *PM01183*®, which form monoadducts in the minor groove that have been shown in fluorescence-based assays to increase substantially the melting temperature of selected DNA oligonucleotides (8–10). An interstrand crosslink (ICL) is considered a highly damaging lesion in DNA and clinically relevant for anti-neoplastic action because it represents a complex challenge to the DNA repair machinery (11).

*To whom correspondence should be addressed. Tel: +34 918 854 514; Fax: +34 918 854 591; Email: federico.gago@uah.es

© The Author(s) 2011. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

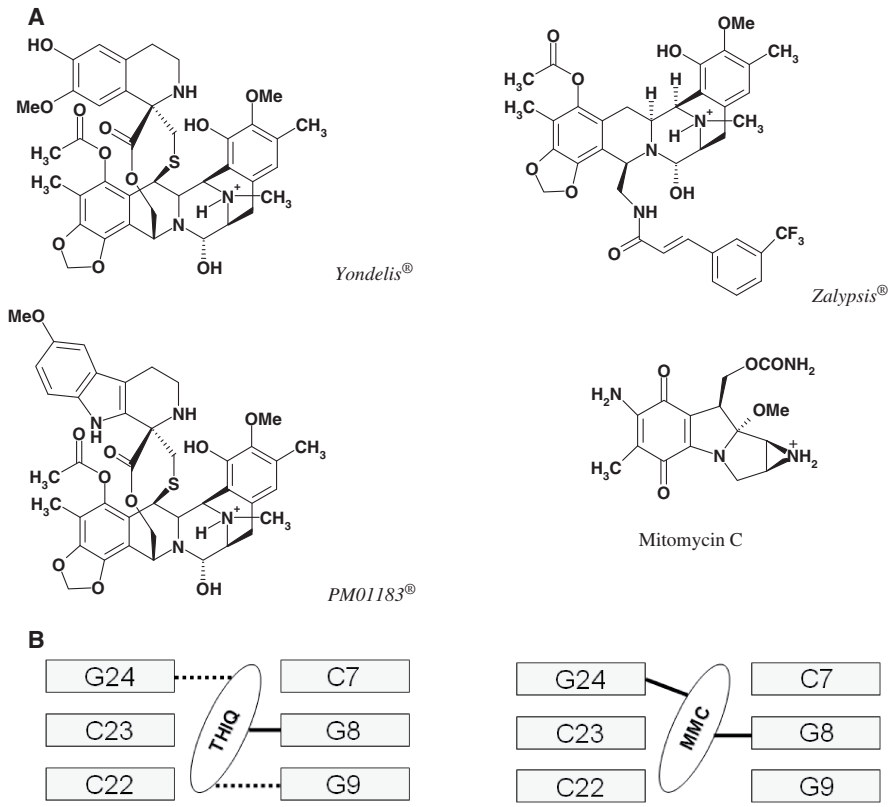


Figure 1. (**A**) Chemical structures of the *N*-protonated forms of the drugs used in the melting simulations. In the case of the tetrahydroisoquinolines, dehydration of the hemiaminal (carbinolamine) yields the reactive iminium intermediate that reacts with the 2-amino group of a guanine (7). In the case of MMC, two-electron reduction of the quinone ring facilitates methoxide elimination, formation of a leuco-aziridinomitosene and opening of the aziridine ring to provide the initial alkylating agent that leads to a DNA monoadduct with the 2-amino group of a guanine; the second alkylating centre is the iminium ion that forms upon the reverse Michael elimination of carbamic acid from the monoadduct and undergoes a nucleophilic attack by the 2-amino group of the guanine in the opposite strand at a CpG step (6). (**B**) Schematic of the binding modes of the tetrahydroisoquinolines *Yondelis®*, *Zalypsis®* and *PM-01183®* (THIQ, left) and MMC (right) to the central CGG triplet of the DNA 15-mer studied. Covalent and hydrogen bonds are shown, respectively, as solid and dotted lines. Base numbers refer to their position in the 15-mer: strand bearing the adduct = 1–15; complementary strand = 16–30.

Advances in computer power, force field development (12) and simulation methodologies applied to the study of nucleic acid structures have led to a maturity in the field that enables theoretical results to fill in some of the missing features that the experimental tools cannot provide (13,14). Thus simulations describing the thermal melting properties of several DNA duplexes have been carried out using both simplified coarse-grained models in the absence of ions and solvent molecules (15–18) and atomistic models in either a continuum solvent model (19) or explicit water (20–22). A remaining limitation in modelling this process is that the time scale usually covered in most MD simulations (tens or hundreds of nanoseconds) is much shorter, with some notable recent exceptions (19), than the time over which melting actually takes place. The workaround has been either to study very short DNA oligomers [e.g. CGC (20) or CGCG (18)] or to carry out the simulations under conditions that favour WC hydrogen bond disruption and strand separation, as for example, increased salt concentration and elevated temperature (21) or the presence of miscible pyridine in the aqueous solvent (19). Another less explored alternative has been to use a combination of replica exchange and metadynamics to simulate the helix to coil transition of three different hexanucleotides (23). Interestingly, by comparing results from melting simulations of Dickerson's dodecamer in the presence of only water or a water–pyridine mixture, it was concluded that strand separation is actually triggered by thermal energy and water molecules, with pyridine's major role being to stabilize partially disrupted microstates through stacking with the solvent-exposed extrahelical nucleobases (19).

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

Recent work from our labs (24) showed that a bead-immobilized DNA template containing a central CGG triplet surrounded by A:T pairs on both sides was weakly sensitive to XPF, a structure-specific endonuclease that preferentially cleaves one strand of duplex DNA at the 5′ side of a junction with single-stranded (ss) DNA (25). Our finding that the cleavage intensity increased, in a concentration-dependent manner, if the polynucleotide was pre-treated with either *Yondelis*® or *Zalypsis*®, led us to think that XPF, which is known to play a key role in multiple steps of ICL repair (26) and cuts the strand having the unpaired 3′ flap, was probably recognizing a ds/ss discontinuity brought about by the particular choice of trinucleotide repeats (TAA and ATT) on each side of the drug-targeted CGG triplet. Our reasoning was that the enhanced XPF endonucleolytic activity observed in the presence of the drug adducts was likely to arise from increased stabilization of the A:T-rich region in which the central CGG triplet was embedded.

With the aim of gaining further insight into the effects of a covalently bonded drug on the process of DNA-strand separation, we have used all-atom molecular dynamics (MD) simulations in saline solutions at high temperature to examine and compare the thermal denaturation of an A:T-rich DNA 15-mer containing a central CGG triplet in the absence and presence of a drug covalently bonded to the underlined guanine. Complete loss of WC base pairing was observed for the drug-free oligonucleotide after 190 ns of simulation, whereas the central regions of the drug-bonded duplexes remained hydrogen bonded to different extents at the final time of 200 ns. Our findings thus support the view that the three tetrahydroisoquinoline adducts studied functionally mimic an ICL, despite the fact that these drugs are covalently bonded to only one strand. Moreover, several of the DNA conformational states that were characterized during the melting transitions were unique to the MMC bis-adduct, and this observation might account for some of the differences observed experimentally regarding the repair of these lesions in mammalian cells (8).

MATERIALS AND METHODS

Construction of the deoxyoligonucleotides and force field parameters for the non-standard residues

An initial model for the drug-free 15-mer of sequence 5′-d(TAATAACGGATTATT)•5′-d(AATAATCCGTTATTA) was built using the *nucgen* module in the AMBER suite of programs (http://ambermd.org/) and optimized parameters for B-DNA (27). The structures of *Yondelis*® (8), *Zalypsis*® (9) and *PM01183*® (10) covalently bonded to the central guanine in the CGG triplet were built as previously reported. The N^2(G)-MMC-N^2(G) bisadduct with 1″-α stereochemistry (6) was modelled by changing the connectivity and subsequent energy refinement of the MMC-N^2(G) monoadduct structure deposited in the Protein Data Bank [PDB (28)] with code 199D and then transferring the ICL to the 15-mer by best-fit superposition on the central CpG step. Electrostatic potential-derived (ESP) point charges for the atoms making up the ICL were assigned from the wavefunction calculated using a 6–31 G(d) basis set, as implemented in the *ab initio* quantum chemistry program Gaussian03 (29), and subsequent fitting employing the RESP methodology (30). Parameters derived for drug atoms were consistent with those present for polynucleotides and proteins in the AMBER force field (31), currently incorporating corrections (parmbsc0) for an improved description of DNA conformations on a multinanosecond time scale (11).

MD simulations

Each molecular system was immersed in a truncated octahedron of TIP3P water molecules (32) that extended 15 Å away from any solute atom so that the resulting volume was enough to accommodate the separation of the strands upon DNA melting. Incorporation of the appropriate number of sodium ions (33) at random locations ensured electrical neutrality. A salt concentration of 0.1 M NaCl was then achieved by addition to each complex of equivalent amounts of Na^+ and Cl^- ions. The cut-off distance for the non-bonded interactions was 9 Å and periodic boundary conditions were used. Electrostatic interactions were represented using the smooth particle-mesh Ewald method (34) with a grid spacing of 1 Å. The SHAKE algorithm (35) was applied to all bonds involving hydrogens, and an integration step of 2.0 fs was used throughout. The simulation protocol made use of the *pmemd* module in AMBER 10 (36) and consisted of a number of steps. First, solvent molecules and counterions were relaxed by energy minimization and allowed to redistribute around the positionally restrained solute (25 kcal•mol^{-1}•Å^{-2}) during 50 ps of MD at constant temperature (300 K) and pressure (1 atom) essentially as described previously (37). These initial harmonic restraints were gradually reduced in a series of progressive energy minimizations until they were completely removed. The resulting systems were then heated from 100 to 400 K during 25 ps, equilibrated at 400 K for 1 ns in the absence of any restraints and further simulated under the same conditions up to a total time of 200 ns during which system coordinates were collected every 20 ps for further analysis. Each complete simulation took ~2000 h of CPU time running in parallel on 32 IBM Power PC 970MP processors at the Barcelona Supercomputer Centre (MareNostrum).

Analysis of the MD trajectories

Three-dimensional structures and trajectories were visually inspected using the computer graphics program PyMOL (38). Interatomic or centroid–centroid distances, angles and root mean square deviations (rmsd) from a given structure were monitored using the *ptraj* module in AMBER. Our criteria for WC hydrogen-bonded pairs were (i) a distance between H-bond donor and acceptor atoms shorter than 3.2 Å, and (ii) the existence of at least two H-bonds between paired bases irrespective of their being A:T or G:C.

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

RESULTS AND DISCUSSION

Progressive loss of WC base pairing in the duplexes

General considerations. MD simulations of DNA duplexes at room temperature routinely show fraying and reannealing of terminal base pairs independently of the force field used (39). In line with these findings, the A:T base pairs closest to the ends of the duplexes studied here were lost soon after initiation of the MD production runs (Figure 2). This strand dissociation process was progressively followed by 'peeling', that is, untwisting motions favoured by nucleic base competition with water molecules and ions, as well as by non-canonical hydrogen bonding of the purine and pyrimidine base edges with the sugar-phosphate backbone. Our choice of A:T base pairs surrounding the central CGG triplet clearly had an impact on the speed of base pair disruption (Figure 3) as only two hydrogen bonds need to be broken compared with three in the case of G:C base pairs, and they also present a smaller surface area for stacking interactions. Likewise, the selection of non-homopolymeric sequences prevented further base complementarity searching to yield a change of registry, as reported for $d(A_{12})\cdot d(T_{12})$ (21) or $d(CGCG)_2$ duplexes (18).

As the simulations progressed it became apparent that the stability of the central CGG triplet was crucial in the reaction coordinate of the melting process. Thus, for the drug-free DNA duplex this triplet remained WC hydrogen-bonded for not less than ~160 ns. At this point dehybridization took over rehybridization and at ~190 ns no standard WC hydrogen-bonded bases remained between the two strands. At 200 ns the simulation was stopped and we considered this time as the deadline for comparison with the other duplexes (Figure 4). Rather unexpectedly, the oligonucleotide containing the MMC bisadduct followed a very similar course, the main difference with respect to the drug-free DNA being that the two strands could not physically separate due to the existence of the ICL between the two guanines. But only the two G:C base pairs making up the CpG step remained WC hydrogen-bonded at the end of the simulation. In stark contrast, the duplexes containing the tetrahydroisoquinoline drug adducts in only one strand displayed a greatly increased stabilization in their central region, with up to 6 bp still hydrogen-bonded after 200 ns. Moreover, further dehybridization after ~110 ns was not only stopped but also reversed in the oligos containing *Yondelis*® or *Zalypsis*®. These reannealing events, which translate into a decreased percentage of time in the unpaired state for the base pairs on the 3′-side of the central CGG triplet relative to drug-free DNA (see shoulder in Figure 5), strongly suggest that the presence of the bulky drug in the minor groove helps stabilize the duplex on both sides of the covalent adduct.

Drug-free DNA complex. The rather dramatic decrease in the number of WC base pairs (from 15 down to 6) during the first ~12 ns of the MD production run at 400 K was mostly due to terminal base hydration, loss of stacking interactions (see below), and the existence of fast and reversible microevents of A:T base pair mismatching (Figure 4) following shearing of the canonical base pairs. This was followed by a period of ~40 ns during which an alternation of reversible dehybridization and rehybridization processes involving the central base pairs of the duplex maintained the double helical structure. The well-preserved central CGG triplet helped nucleate some extra dsDNA on both sides during this first part of the simulation forcing the closest AT base pairs to be temporarily reannealed (Figure 3). During the next 60 ns of simulation, the progressive and irreversible melting of all the A:T base pairs was observed. The result was that by 120 ns only the central and more stable CGG triplet remained WC hydrogen bonded. After this point, random hydrogen bonds between mismatched bases could be detected, but strong competition with the solvent prevented stabilization of any of these transition intermediates. Despite the occurrence of some occasional rehybridization microevents between 140 and 160 ns, the remaining two G:C base pairs were clearly not enough to promote further helicity and they broke up at 190 ns. At this time, we considered that the 15-mer was completely and irreversibly melted. Therefore, the three G:C base pairs making up the central triplet remained hydrogen bonded for a total length of about 190 ns.

The time evolution of the rmsd of the drug-free duplex with respect to the initial structure (Supplementary Figure S1) correlated strongly with the fraying-peeling process described above. The results of this simulation allowed us to understand the pathway of double helix disruption under denaturing conditions and were found to be in consonance with results from previous studies using different sequences (19,21). However, by introducing a central CGG triplet, we were able to assess more clearly the much higher stability of these three base pairs with respect to the surrounding AT pairs, which nicely highlights the base-pair dependence of the melting transition. Furthermore, examination of the evolution of the unpaired state of individual base pairs over the course of the simulation revealed some melting asymmetry. Thus, the ATT stretch located 3′ to the central CGG sequence behaved in an all-or-none fashion and melted as a block rather than progressively as did the TAA stretch located on the 5′ side (Figure 3). Interestingly, when TAA and ATT stretches on the 5′ and 3′ sides of the central CGG triplet were swapped, the asymmetry was even more accentuated (Supplementary Figure S2), with $(ATT)_2$ on the 5′-side melting much faster and smoothly than the $(TAA)_2$ block on the 3′-side, which was more resilient to melting and whose base pairs were also disrupted synchronously. Although more work is necessary to account for these differences, it seems clear that ATT triplets are more easily destabilized than TAA triplets, at least when located on the 5′-side of a CGG site. Our next goal was to examine the effect on this melting process of a drug-bonded guanine in the central triplet.

Drug-bonded DNA duplexes. The presence of the bonded drugs had a prominent effect on the denaturing profile of the DNA oligonucleotide (Figure 5). Most noticeable was the greater occurrence of reannealing events that took place over the course of the simulation, with some very

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

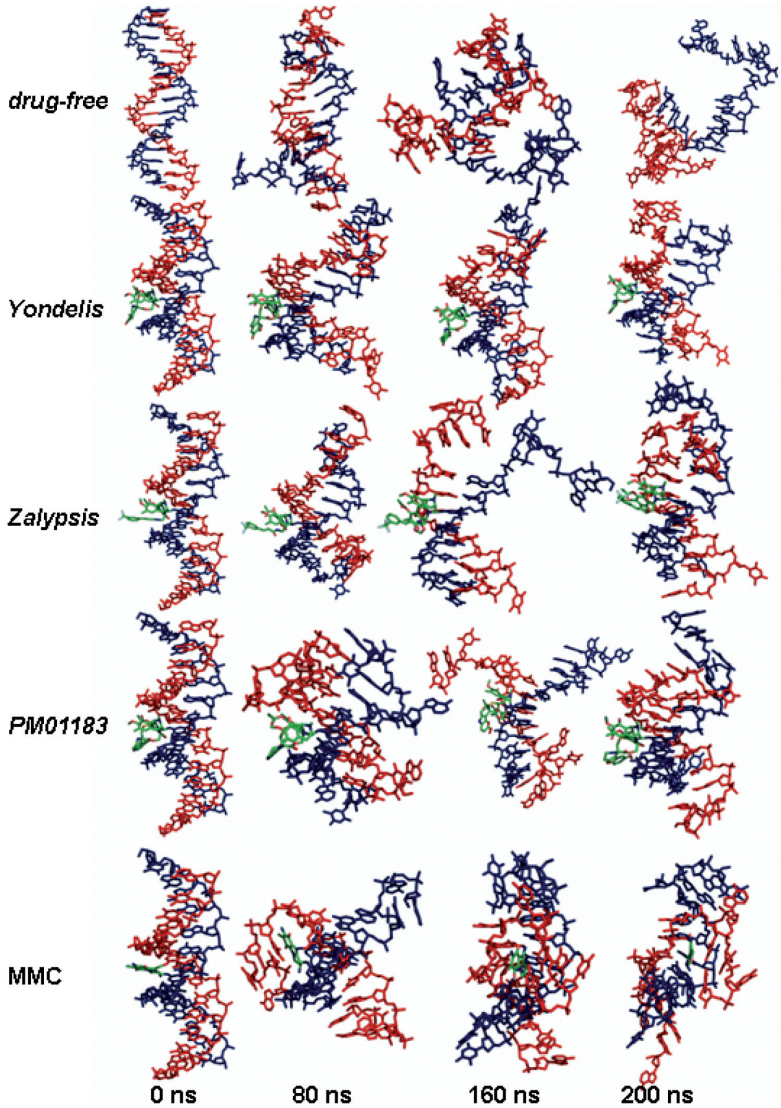


Figure 2. Representative snapshots showing the time evolution of the DNA duplexes during the MD simulations in the absence of any bonded drug (top) and in the presence of *Yondelis*®, *Zalypsis*® or *PM01183*® covalently bonded to the blue strand, or MMC (bottom) covalently bonded to both blue and red strands. Drug carbon atoms are coloured in green. For more detailed visualization see the video in Supplementary Data.

dramatic examples showing up at ~120 ns (Figure 4). Indeed, six WC hydrogen-bonded base pairs still remained in the complexes containing the three tetrahydroisoquinolines midway the simulations and for *Yondelis*® this number went up to 11 at ~125 ns. In other words, the double helical structure was preserved at this time to virtually the same extent that it was at the beginning of the MD production run. These continuous dehybridization–rehybridization microevents kept taking place until the very end of the simulation. At this final time, the duplex structures were clearly maintained in the central region since the number of WC hydrogen-bonded base pairs was very close to six in the three complexes.

Remarkably, these critical reannealing events were not so perceptible in the MMC–DNA complex despite the fact that the two strands were physically linked through this drug, which effectively prevented strand separation (Figure 4). Therefore, the most likely explanation for this distinct behaviour is that the smaller size of this

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

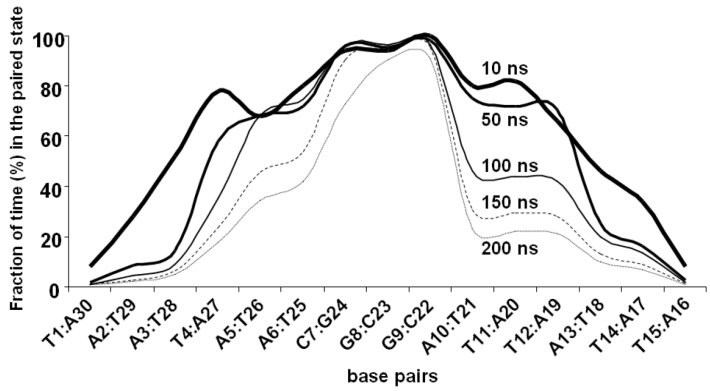


Figure 3. Fraction of the MD simulation time (measured at 10, 50, 100, 150 and 200 ns) during which individual base pairs remained WC hydrogen bonded in the drug-free DNA 15-mer of sequence 5′-d(ATTATTCGGTAATAA)•5′-d(TTATTACCGAATAAT). Hydrogen bonds were rapidly lost for those base pairs found at the strand termini whereas those making up the central CGG triplet were the most durable. Note the asymmetry in the melting profiles of the DNA stretches located 5′ and 3′ to the central CGG triplet. This profile is to be compared with that obtained when these TAA and ATT stretches were swapped (Supplementary Figure S2).

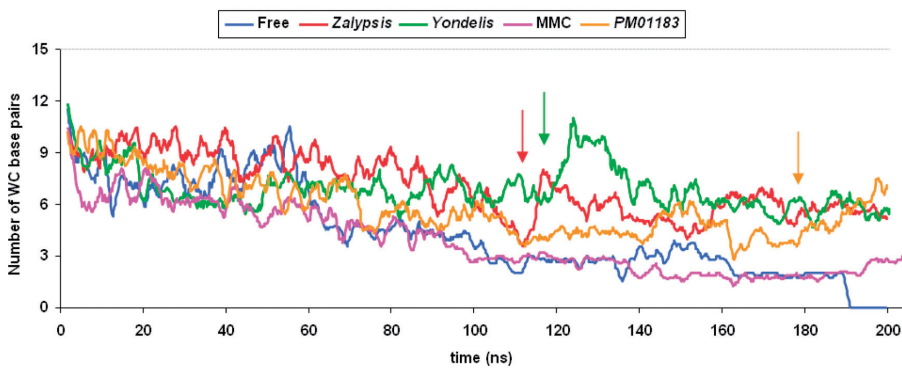


Figure 4. DNA base-pair disruption along the production phase of the MD simulations of the free and drug-bonded DNA 15-mers. All the original base pairs were progressively lost in the drug-free DNA, the last ones belonging to the central CGG triplet. In all the drug-bonded complexes, however, this triplet remained H-bonded at the end of the simulation. Furthermore, in the DNA molecules containing the bonded tetrahydroisoquinolines, the WC base-paired region was extended further due to the occurrence of some conspicuous reannealing events (marked by a vertical arrow), which were more marked for *Zalypsis*® and *Yondelis*® than for *PM01183*®.

molecule, which fits into the minor groove causing only minimal distortion in the DNA (6), does not give rise to significant non-covalent (van der Waals and electrostatic) interactions with the DNA strands that could help stabilize the duplex. The two G:C base pairs making up the CpG step remained hydrogen bonded all along, with occasional extension to the remaining G:C base pair during the second half of the simulation, but this did not impinge on the stability of the neighbouring A:T base pairs. Interestingly, the last base pair to lose the WC hydrogen bonds in the simulation of the free DNA was G9:C22, whereas the most recalcitrant was C7:G24 when G8 and G24 were linked together by means of the bonded MMC.

Our results also showed, in common with previous evidence (19,21), that disrupted interstrand WC base pairing during the melting process, as described above, was accompanied by increased anomalous pairings in the duplex (Figure 2). Nonetheless, the solvent exposure of the polar atoms that make up the WC hydrogen bonds in a standard DNA double helix consistently increased during the simulated melting process of the drug-free duplex and reached a maximum after the two strands were separated (Supplementary Figure S3A). As expected, this increased exposure was much less marked in the duplexes containing the drug adducts, particularly in the case of *Zalypsis*®, in good correspondence with the better maintained WC base pairs discussed above.

Base stacking and hydrophobic effect

Base stacking plays an even greater role than hydrogen bonding in the stabilization of DNA duplexes in aqueous solutions due to a combination of the hydrophobic effect (40,41) and optimal dipolar interactions between the planar heteroaromatic systems (42). During the

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

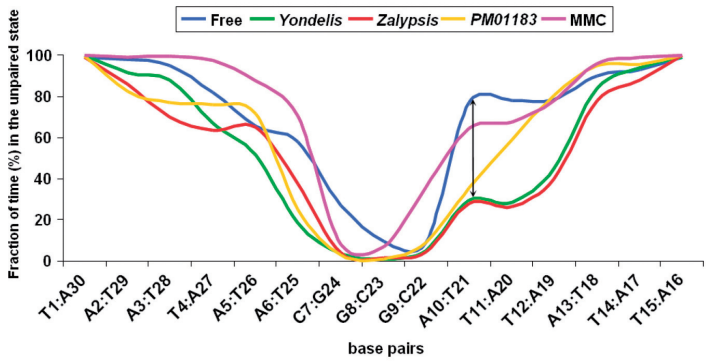


Figure 5. Fraction of total MD simulation time during which a particular base pair was disrupted in the DNA 15-mer (note that for drug-free DNA this plot is the reverse of that shown in Figure 3). The lower frequency of the unpaired state in the central CGG triplet was extended to the ATT sequence 3′ to it (i.e. A10:T21 and T11:A20 base pairs) in the duplexes containing bonded *Yondelis*® and *Zalypsis*® (vertical double-headed arrow) as a consequence of the reannealing events shown in Figure 4.

simulations, canonical intrastrand base stacking was largely maintained and, when lost, it was partially compensated by both non-contiguous intrastrand base stacking events and occasional interstrand face-to-face and edge-to-face interactions (data not shown). The overall result was that the individual strands showed a relatively high degree of compactness over the whole melting process (Figure 2). Nonetheless, despite the partial collapse of the individual strands, the DNA non-polar atoms were more exposed to the solvent at the end of the simulations than at the beginning (Supplementary Figure S3B) as a consequence of the increased disorder and loss of stacking interactions. Comparison of initial and final states for the different duplexes shows that those containing the adducts initially display a reduced solvent-accessible surface area (SASA) of their polar atoms relative to drug-free DNA due to partial occlusion by the bonded drug and also reduced solvent exposure at the end of the simulations as a consequence of the larger number of WC hydrogen-bonded base pairs. These differences are not as striking in the case of non-polar atoms because the extent of avoidance of interactions with water is much more similar in all the duplexes.

Implications for the mode of action of these drugs

The relevance of the present results is supported by (i) renewed interest in DNA minor groove binders as anti-cancer therapeutic agents (7,43), (ii) the established role of MMC in the therapy of bladder tumours (44) and other types of cancer, (iii) the recent approval of trabectedin (*Yondelis*®) (45) for the treatment of patients with advanced or metastatic soft tissue sarcoma and relapsed platinum-sensitive ovarian carcinoma and (iv) the ongoing clinical development of *Zalypsis*® (9) and *PM01183*® (10) for the treatment of solid tumours and haematological malignancies.

The three tetrahydroisoquinolines studied here have in common a fused pentacyclic skeleton that is mostly responsible for DNA recognition and binding to the amino group of a suitably embedded guanine (7). The sequence specificity of the three drugs depends on the establishment of highly specific hydrogen bonds with the nucleotides on both sides of the guanine that will be covalently bonded (46,47). Besides, non-covalent interactions within the minor groove help to stabilize the double helical structure of the duplex to which they bind, as inferred from experiments demonstrating notable increases (~20°C) in the melting temperature of dsDNA containing a single drug adduct (8–10) and the MD simulation results presented herein. Thus, although these three compounds are covalently bonded to only one strand of the DNA double helix, they have been viewed (8) as functional interstrand crosslinkers insofar as they prevent strand separation very effectively. To compare this action with that elicited by a true interstrand crosslinker, we chose MMC as a prototypical representative because of the similarities detected in cells exposed to all of these drugs regarding (i) phosphorylation of histone H2AX, a surrogate marker of double-strand breaks, (ii) Rad51 foci formation in the nucleus, which is known to facilitate DNA repair via homologous recombination and (iii) participation of the Fanconi anaemia (FA) pathway in the processing of the DNA damage (8). Nonetheless, a notable and yet unexplained difference regarding the effects of MMC and *Yondelis*® on the cell cycle of FA-deficient cells is that trabectedin, unlike MMC, did not induce any significant accumulation of cells in G_2-M. Among the possible reasons for this differential effect on cell-cycle progression, the nature of the interaction of these drugs with DNA was pinpointed (8). The present results reinforce this possibility and provide a clear atomistic rationale for this interpretation even though more work will be necessary to clarify this point categorically.

CONCLUSIONS

DNA melting in the absence and presence of bonded drugs can provide information about the duplex

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

stabilization brought about by drug binding. Performing the simulations under conditions similar to those used in the real-life experiments, i.e. progressively increasing the temperature until complete separation of the two strands is detected, is currently unfeasible given the time scale of this process. However, by employing a constant elevated temperature from the outset as a control parameter for overcoming energy barriers more easily and using high salt concentrations, we were able to detect complete thermal denaturation of the drug-free oligodeoxynucleotide in <200 ns, still a formidable time for simulating large macromolecular systems such as these but now made affordable by the combined use of optimized parallel algorithms and available supercomputing power.

Initially, the duplex under study had fifteen WC hydrogen-bonded base pairs. During the initial stages of the simulation the frayed nucleobases at the ends started looping back to make transient hydrogen bonds with the sugar-phosphate backbone and this resulted in both untwisting and peeling apart of the DNA strands, essentially as described for similar duplexes (19,21). Although the number of intact base pairs consistently decreased with time, a striking asymmetry was observed on both sides of the central CGG triplet, as well as considerable fluctuations over the full course of the denaturation process. Furthermore, the progressive unzipping was clearly interspersed with some reannealing events, which were most noticeable in the oligonucleotides containing the tetrahydroisoquinoline monoadducts. After 200 ns of MD simulation, all the original WC hydrogen bonds were lost in the free oligonucleotide, whereas an extended WC-paired region embedding the central CGG triplet was still maintained in all drug–DNA complexes. This finding attests to the ability of these drugs to stabilize dsDNA and stall replication and transcription forks. On the other hand, in the MMC–DNA complex (Figure 6), but not in those with *Yondelis*®, *Zalypsis*® or *PM01183*®, several unique arrangements between the drug and different DNA bases in the ss region favoured the transient formation of distinct secondary structures in which MMC appeared pseudointercalated (Supplementary Figure S4). These hydrophobically collapsed intermediates, together with the existence of a true ICL physically connecting the two DNA strands, are likely to pose distinct challenges to the DNA repair machinery and may account for the differences in cellular response that have been detected experimentally (8).

Taken together, our results provide an atomistic view of the process of DNA melting in the presence and absence of several bonded anti-tumour drugs, and they are found to be in agreement with the large stabilization that single *Yondelis*®, *Zalypsis*® and *PM01183*® monoadducts have been shown to impose on the DNA double helix (8–10). On the contrary, the prototypical interstrand crosslinker MMC does prevent strand separation because of the fact that it forms a covalent bond with each DNA strand, but its effect on duplex stabilization does not appear to extend beyond the CpG step to which it is covalently attached. Therefore, this is another example of how theory and experiment can be successfully merged to help us improve our understanding in atomic detail of anti-tumour drugs acting on such a vital macromolecule as DNA.

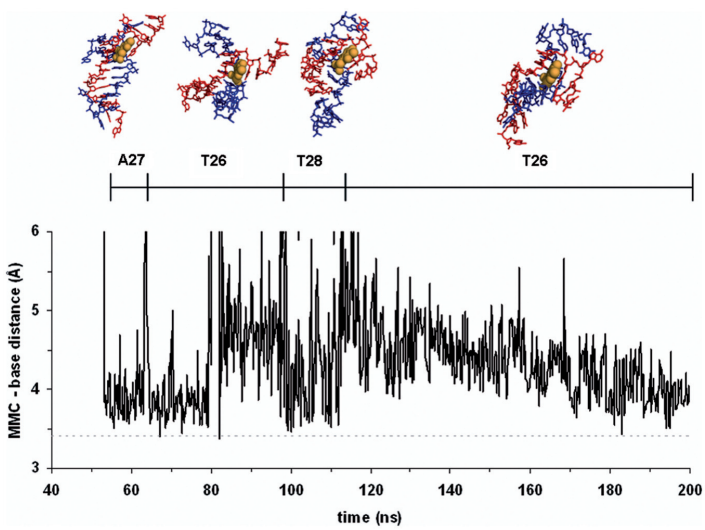


Figure 6. Distances between the ring centroids of MMC and different DNA nucleobases during the simulated melting process. The drug's chromophore (CPK spheres coloured in orange) served as a platform for hydrophobic contacts with different nucleobases from one of the melted strands (that coloured in red), in the chronological order shown at the top of the figure. The dotted line given as a reference at 3.4 Å represents the canonical distance between stacked nucleobases in a standard B-DNA double helix.

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thankfully acknowledge the generous allowance of computer resources, technical expertise and assistance provided by the Red Española de Supercomputación at the Barcelona Supercomputing Center (MareNostrum).

FUNDING

This work was partially supported by l'Association de la Recherche contre le Cancer (ARC n°3153), la Ligue contre le Cancer, INCA and an ERC Advanced grant (to J.M.E.), and Comisión Interministerial de Ciencia y Tecnología (SAF2009-13914-C02-02) (to F.G.). C.G. is a recipient of a fellowship from the Ministère de la Recherche and ARC. J.A.B.-C. enjoys a research fellowship from PharmaMar S.A.U. Funding for open access charge: Comisión Interministerial de Ciencia y Tecnología (SAF2009-13914-C02-02).

Conflict of Interest statement: C.M.G. is an employee and shareholder of PharmaMar S.A.U. No other potential conflicts of interest were disclosed.

REFERENCES

1. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
2. Stucki,M., Stagljar,I., Jónsson,Z.O. and Hübscher,U. (2001) A coordinated interplay: proteins with multiple functions in DNA replication, DNA repair, cell cycle/checkpoint control, and transcription. *Prog. Nucleic Acid. Res. Mol. Biol.*, **65**, 261–298.
3. Bustamante,C., Cheng,W. and Meija,Y.X. (2011) Revisiting the central dogma one molecule at a time. *Cell*, **144**, 480–497.
4. Negri,A., Marco,E., García-Hernández,V., Domingo,A., Llamas-Saiz,A.L., Porto-Sandá,S., Riguera,R., Laine,W., David-Cordonnier,M.H., Bailly,C. *et al.* (2007) Antitumor activity, X-ray crystal structure, and DNA binding properties of thiocoraline A, a natural bis-intercalating thiodepsipeptide. *J. Med. Chem.*, **50**, 3322–3333.
5. Lawley,P.D. and Phillips,D.H. (1996) DNA adducts from chemotherapeutic agents. *Mutat. Res.*, **355**, 13–40.
6. Tomasz,M. (1995) Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.*, **2**, 575–579.
7. Manzanares,I., Cuevas,C., García-Nieto,R., Marco,E. and Gago,F. (2001) Advances in the chemistry and pharmacology of ecteinascidins, a promising new class of anticancer agents. *Curr. Med. Chem. Anti-Canc. Agents*, **1**, 257–276.
8. Casado,J.A., Río,P., Marco,E., García-Hernández,V., Domingo,A., Pérez,L., Tercero,J.C., Vaquero,J.J., Albella,B., Gago,F. *et al.* (2008) The relevance of the Fanconi anemia pathway in the response of human cells to trabectedin. *Mol. Cancer Ther.*, **7**, 1309–1318.
9. Leal,J.F., García-Hernández,V., Moneo,V., Domingo,A., Bueren-Calabuig,J.A., Negri,A., Gago,F., Guillén-Navarro,M.J., Avilés,P., Cuevas,C. *et al.* (2009) Molecular pharmacology and antitumor activity of Zalypsis in several human cancer cell lines. *Biochem. Pharmacol.*, **78**, 162–170.
10. Leal,J.F., Martínez-Díez,M., García-Hernández,V., Moneo,V., Domingo,A., Bueren-Calabuig,J.A., Negri,A., Gago,F., Guillén-Navarro,M.J., Avilés,P. *et al.* (2010) PM01183, a new DNA minor groove covalent binder with potent in vitro and in vivo anti-tumour activity. *Brit. J. Pharmacol.*, **161**, 1099–1110.
11. McHugh,P.J., Spanswick,V.J. and Hartley,J.A. (2001) Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *Lancet Oncol.*, **2**, 483–490.
12. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
13. Schatz,G.C. (2007) Using theory and computation to model nanoscale properties. *Proc. Natl. Acad. Sci. USA*, **104**, 6885–6892.
14. Orozco,M., Noy,A. and Pérez,A. (2008) Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.*, **18**, 185–193.
15. Drukker,K. and Schatz,G.C. (2000) A model for simulating dynamics of DNA denaturation. *J. Phys. Chem. B*, **104**, 6108–6111.
16. Drukker,K., Wu,G. and Schatz,G.C. (2001) Model simulations of DNA denaturation dynamics. *J. Chem. Phys.*, **114**, 579–590.
17. Tepper,H.L. and Voth,G.A. (2005) A coarse-grained model for double-helix molecules in solution: spontaneous helix formation and equilibrium properties. *J Chem. Phys.*, **122**, 124906.
18. Linak,M.C. and Dorfman,K.D. (2010) Analysis of a DNA simulation model through hairpin melting experiments. *J. Chem. Phys.*, **133**, 125101.
19. Kannan,S. and Zacharias,M. (2009) Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, **11**, 10589–10595.
20. Pérez,A. and Orozco,M. (2010) Real-time atomistic description of DNA unfolding. *Angew. Chem. Int. Ed.*, **49**, 4805–4808.
21. Hagan,M.F., Dinner,A.R., Chandler,D. and Chakraborty,A.K. (2003) Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *Proc. Natl Acad. Sci. USA*, **100**, 13922–13927.
22. Wong,K.Y. and Pettitt,B.M. (2008) The pathway of oligomeric DNA melting investigated by molecular dynamics simulations. *Biophys. J.*, **95**, 5618–5626.
23. Piana,S. (2007) Atomistic simulation of the DNA helix-coil transition. *J. Phys. Chem. A*, **111**, 12349–12354.
24. Feuerhahn,S., Giraudon,C., Martínez-Díez,M., Bueren-Calabuig,J.A., Galmarini,C.M., Gago,F. and Egly,J.M. XPF-dependent DNA breaks and RNA polymerase II arrest induced by antitumor DNA interstrand crosslinking-mimetic tetrahydroisoquinoline alkaloids. *Chem. Biol.*, (in preparation).
25. de Laat,W.L., Appeldoorn,E., Jaspers,N.G.J. and Hoeijmakers,J.H.J. (1998) DNA structural elements required for ERCC1-XPF endonuclease activity. *J. Biol. Chem.*, **273**, 7835–7842.
26. Rahn,J.J., Adair,G.M. and Nairn,R.S. (2010) Multiple roles of ERCC1-XPF in mammalian interstrand crosslink repair. *Environ. Mol. Mutagen*, **51**, 567–581.
27. Arnott,S. and Hukins,D.W. (1972) Optimised parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Comm.*, **47**, 1504–1509.
28. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
29. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Zakrzewski,V.G., Montgomery,J.A. Jr, Stratmann,R.E., Burant,J.C. *et al.* (2003) *Gaussian 03, Revision B. 04*. Gaussian, Inc., Pittsburgh, PA.
30. Bayly,C.I., Cieplak,P., Cornell,W.D. and Kollman,P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: the RESP model. *J. Phys. Chem.*, **97**, 10269–10280.
31. 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. and Kollman,P.A. (1995) A second-generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
32. Jorgensen,W.L., Chandrasekhar,J. and Madura,J.D. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
33. Åqvist,J. (1990) Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.*, **94**, 8021–8024.

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

34. Darden,T.A., York,D. and Pedersen,L.G. (1993) Particle mesh Ewald: an N*log(N) method for computing Ewald sums. *J. Chem. Phys.*, **98**, 10089–10092.
35. Ryckaert,J.P., Ciccoti,G. and Berendsen,H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, **23**, 327–341.
36. Case,D.A., Darden,T.A., Cheatham,T.E. III, Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Crowley,M., Walker,W., Kollman,P.A. *et al.* (2008) *AMBER 10*. University of California, San Francisco.
37. Marco,E., Negri,A., Luque,F.J. and Gago,F. (2005) Role of stacking interactions in the binding sequence preferences of DNA bis-intercalators: insight from thermodynamic integration free energy simulations. *Nucleic Acids Res.*, **33**, 6214–6224.
38. Lano,W.D. (2006) PyMOL version 0.99. DeLano Scientific LLC, http://www.pymol.org/ (10 May 2011, date last accessed).
39. Feig,M. and Pettitt,B.M. (1998) Structural equilibrium of DNA represented with different force fields. *Biophys. J.*, **75**, 134–149.
40. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, NY.
41. Friedman,R.A. and Honig,B. (1995) A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.*, **69**, 1528–1535.
42. Gago,F. (1998) Stacking interactions and intercalative DNA binding. *Methods*, **14**, 277–292.
43. Susbielle,G., Blattes,R., Brevet,V., Monod,C. and Kas,E. (2005) Target practice: aiming at satellite repeats with DNA minor groove binders. *Curr. Med. Chem.*, **5**, 409–420.
44. Shelley,M.D., Mason,M.D. and Kynaston,H. (2010) Intravesical therapy for superficial bladder cancer: a systematic review of randomised trials and meta-analyses. *Cancer Treat Rev.*, **36**, 195–205.
45. D'Incalci,M. and Galmarini,C.M. (2010) A review of trabectedin (ET-743): a unique mechanism of action. *Mol. Cancer Ther.*, **9**, 2157–2163.
46. García-Nieto,R., Manzanares,I., Cuevas,C. and Gago,F. (2000) Bending of DNA upon binding of ecteinascidin 743 and phthalascidin 650 studied by unrestrained molecular dynamics simulations. *J. Am. Chem. Soc.*, **122**, 7172–7182.
47. Marco,E., David-Cordonnier,M.H., Bailly,C., Cuevas,C. and Gago,F. (2006) Further insight into the DNA recognition mechanism of trabectedin from the differential affinity of its demethylated analog ecteinascidin ET729 for the triplet binding site CGA. *J. Med. Chem.*, **49**, 6925–6929.

Downloaded from nar.oxfordjournals.org at Universidad de Alcala on September 16, 2011

© 2011 *Nucleic Acids Research* 15 pages

SUPPLEMENTARY MATERIAL

Title: Temperature-induced melting of double-stranded DNA in the absence and presence of covalently bonded antitumour drugs: insight from molecular dynamics simulations

Authors: Juan A. Bueren-Calabuig, Christophe Giraudon, Carlos M. Galmarini, Jean Marc Egly, and Federico Gago

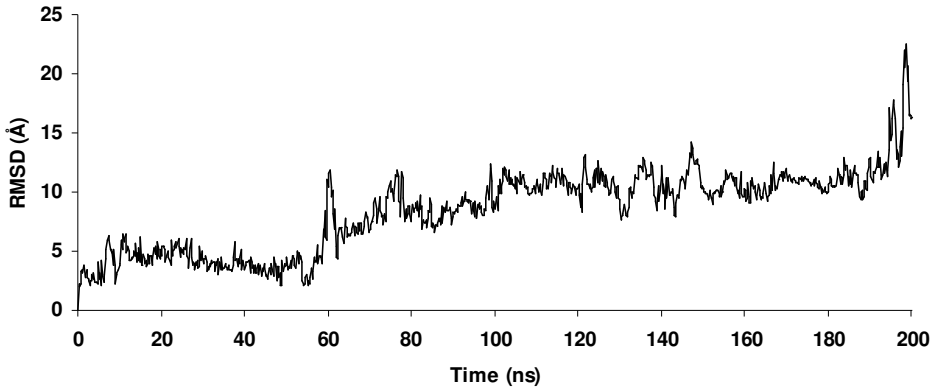


Figure S1. Time evolution of the root-mean-square deviation from the initial structure of the drug-free DNA duplex.

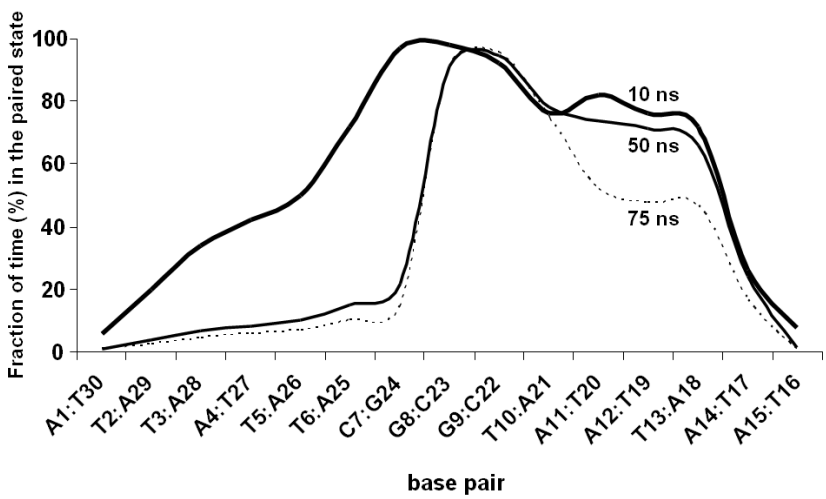
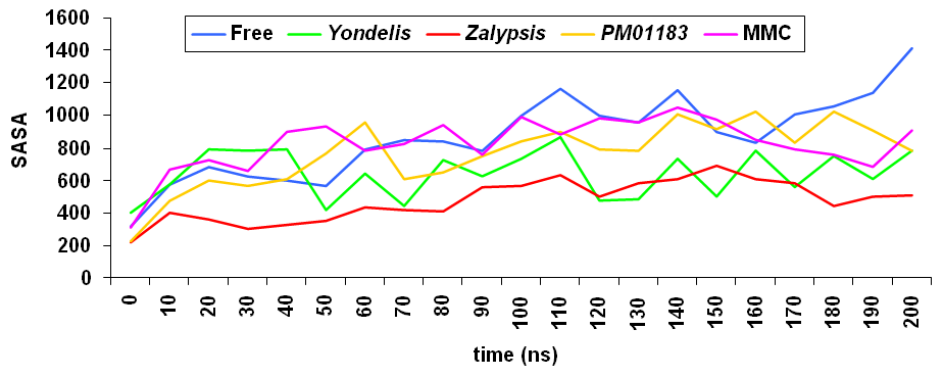


Figure S2. Fraction of MD simulation time (measured at 10, 50 and 75 ns out of a total of 80 ns) during which individual base pairs remained WC hydrogen-bonded in a drug-free DNA 15mer of sequence 5'-d(ATTATTCGGTAATAA)·5'-d(TTATTACCGAATAAT). In common with the results presented in Figure 3, hydrogen bonds were rapidly lost for those base pairs found at the strand termini whereas those making up the central CGG triplet were the most durable. The asymmetry of the profile is also apparent, with the $(ATT)_2$ blocks on the 5' side of the central CGG triplet melting much faster than the $(TAA)_2$ blocks on the 3' side.

A



B

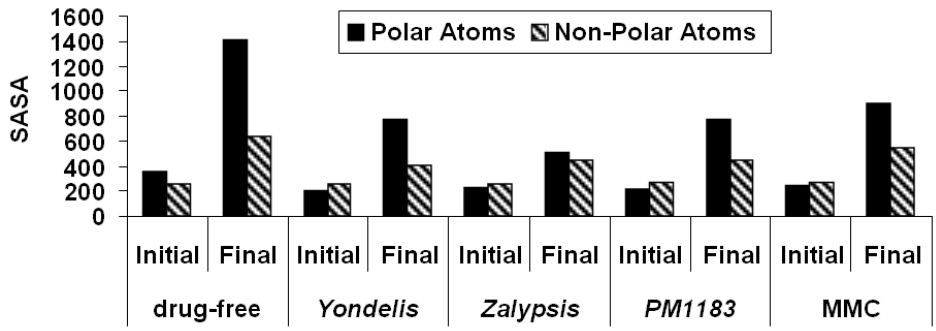


Figure S3. (**A**) Time evolution of the solvent-accessible surface area (SASA, Å^2) of DNA polar atoms involved in Watson-Crick base pairing in the different systems. (**B**) SASA of DNA polar and non-polar atoms at the beginning and end of the MD simulations for drug-free and drug-bonded duplexes. Non-polar atoms include C2, C4, C5, C6 and C7 (for A and G) and C2, C4, C5 and C6 (for C and T).

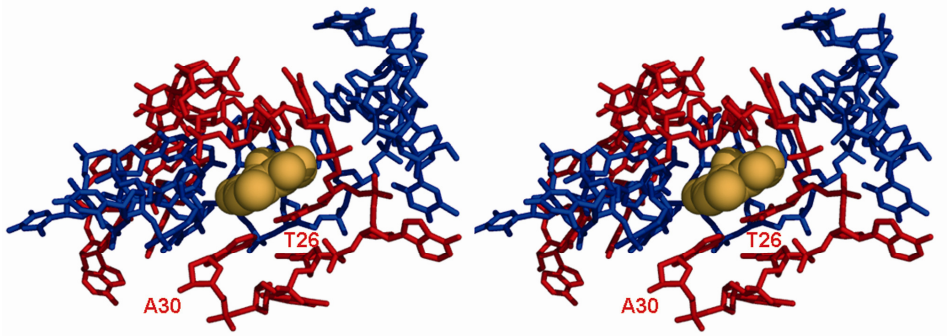


Figure S4. Stereoview of a detail of the MMC-DNA complex at the end of the simulation showing the pseudointercalation of the drug moiety, which is covalently attached to both G8 and G24, between the T26:A30 reverse Hoogsteen base-pair and the sugar-phosphate backbone on both sides of G24.

MOVIE "Melting of a DNA molecule containing a central CGG triplet" / "Reannealing of a partially melted DNA molecule containing a Zalypsis adduct in the central CGG triplet" / "Reannealing of a partially melted DNA molecule containing a Yondelis adduct in the central CGG triplet": NAR-01115-2011_movie.wmv

AMBER PREP file for *Yondelis*®, *Zalypsis*®, PM01183 and MMC (note that A, B and C do not refer to structural subunits).

```
0   0   2

Modified Guanine for adducts with A-Yondelis®, A-Zalypsis® and A-PM01183®
eta.data
ECG  INT     1
CORR OMIT DU   BEG
   0.00000
   1  DU1   DU     M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DU2   DU     M   1   0  -1   1.7321   0.0000   0.0000   0.00000
   3  DU3   DU     M   2   1   0   1.4142 144.7356 -90.0000   0.00000
   4   P    P      M   2   1   3   2.4703  50.1888  61.6477   1.1659
   5   O1P  O2     E   4   2   1   1.4763 144.7274  47.2144  -0.7761
   6   O2P  O2     E   4   2   1   1.4821  62.8724 309.9299  -0.7761
   7   O5'  OS     M   4   2   1   1.5948 102.4912 204.2703  -0.4954
   8   C5'  CT     M   7   4   2   1.4493 118.8416 263.2188  -0.0069
   9   H5'1 H1     E   8   7   4   1.0890 109.8293  93.9650   0.0754
  10   H5'2 H1     E   8   7   4   1.0898 109.7241 333.9517   0.0754
  11   C4'  CT     M   8   7   4   1.5109 109.7919 214.0418   0.1629
  12   H4'  H1     E  11   8   7   1.0892 107.8247 156.3735   0.1176
  13   O4'  OS     M  11   8   7   1.4640 108.5952 279.1056  -0.3691
  14   C1'  CT     M  13  11   8   1.4144 109.8389 106.3765   0.0358
  15   H1'  H2     E  14  13  11   1.0899 107.9098 115.7314   0.1746
  16   N9   N*     S  14  13  11   1.4860 107.8006 232.5199  -0.008847
  17   C8   CK     B  16  14  13   1.3799 129.3211  82.0803   0.116717
  18   H8   H5     E  17  16  14   1.0894 122.9311   0.0000   0.177271
  19   N7   NB     S  17  16  14   1.3084 114.1040 179.8877  -0.524723
  20   C5   CB     S  19  17  16   1.3965 104.0411   0.1827   0.162978
  21   C6   C      B  20  19  17   1.4120 130.2586 179.8255   0.526401
  22   O6   O      E  21  20  19   1.2311 129.0633   0.0000  -0.515010
  23   N1   NA     B  21  20  19   1.4066 111.4536 179.9325  -0.443218
  24   H1   H      E  23  21  20   1.0301 117.4544 179.6740   0.317620
  25   C2   CA     B  23  21  20   1.3790 125.0207 359.6645   0.420088
  26   N2   N2     B  25  23  21   1.3364 116.1465 180.2635  -0.433801
  27   H21  H      E  26  25  23   1.0294 106.9340  60.0461   0.336348
  28   C1   CT     B  26  25  23   1.4796 119.9660 180.0141   0.013125
  29   H2   HC     E  28  26  25   1.0904 105.6685 304.1576   0.179603
  30   C3   CT     B  28  26  25   1.5603 105.6791  55.8639  -0.089666
  31   H4   H1     E  30  28  26   1.1317 106.6857 313.8982   0.122702
  32   N5   N3     3  30  28  26   1.5067 110.6052  70.0845   0.092835
  33   H6   H      E  32  30  28   1.0293 106.9270 304.2391   0.206181
  34   C7   CT     3  32  30  28   1.4795 106.8390 184.2761  -0.107390
  35  H71   HC     E  34  32  30   1.0888 109.5107 180.0612   0.108842
  36  H72   HC     E  34  32  30   1.0905 109.4499 300.0685   0.108842
  37  H73   HC     E  34  32  30   1.0902 109.5153  59.9898   0.108842
  38   C11  CT     B  32  30  28   1.5227 106.8317  64.2877   0.012519
  39  H12   HC     E  38  32  30   1.1352 107.4078 186.1215   0.157478
  40  C13   CA     S  38  32  30   1.4963 105.7016  70.0665  -0.083898
  41  C14   C      B  40  38  32   1.4022 120.8386 134.9914   0.212355
  42  O15   OH     S  41  40  38   1.3717 117.9248   1.9597  -0.585719
  43  H16   HO     E  42  41  40   0.9738 107.9749 182.9521   0.451391
  44  C17   CA     B  41  40  38   1.4125 119.9951 184.0290   0.065299
  45  O18   OM     S  44  41  40   1.3896 116.2935 175.2938  -0.323304
  46  C19   CT     3  45  44  41   1.4317 113.9478 108.1402  -0.035596
  47  H20   H1     E  46  45  44   1.1194 103.3890 176.6840   0.095662
  48  H23   H1     E  46  45  44   1.1170 109.9585 295.4014   0.095662
  49  H22   H1     E  46  45  44   1.1157 110.5985  57.8354   0.095662
  50  C23   CA     B  44  41  40   1.4002 120.6241   1.4926   0.053147
  51  C24   CT     3  50  44  41   1.4809 121.3482 178.8403  -0.142522
  52  H25   HC     E  51  50  44   1.1190 109.9140  94.3331   0.074748
  53  H26   HC     E  51  50  44   1.1189 110.7278 213.8631   0.074748
  54  H27   HC     E  51  50  44   1.1185 110.7967 334.4048   0.074748
  55  C28   CA     B  50  44  41   1.4036 118.7298 359.6637  -0.231978
  56  H29   HA     E  55  50  44   1.1007 119.1213 179.8199   0.168359
  57  C30   CA     S  55  50  44   1.3933 121.1542 359.6603  -0.057758
```

```
  58  C31   CT     B  57  55  50   1.4899 119.5215 179.4211  -0.144867
  59  H32   HC     E  58  57  55   1.1240 109.3309 319.1414   0.102911
  60  H33   HC     E  58  57  55   1.1332 107.6475  75.4368   0.102911
  61  N3    NC     S  25  23  21   1.3308 123.5859   0.7777  -0.427006
  62  C4    CB     E  61  25  23   1.3587 112.3112 359.4417   0.099071
  63  C2'   CT     M  14  13  11   1.5234 107.8896 355.8350  -0.0854
  64  H2'1  HC     E  63  14  13   1.0900 112.3093 144.2909   0.0718
  65  H2'2  HC     E  63  14  13   1.0909 112.2146 265.4816   0.0718
  66  C3'   CT     M  63  14  13   1.5200 102.1265  24.8682   0.0713
  67  H3'   H1     E  66  63  14   1.0898 111.9942 205.6875   0.0985
  68  O3'   OS     M  66  63  14   1.4259 109.1307  84.4189  -0.5232

LOOP
 C4' C3'
 C4  C5
 C4  N9
 C31 C3
 C13 C30

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 N7   N9   C8   H8
 N1   N3   C2   N2

DONE
STOP
```

```
0   0   2

B-Yondelis®
etb.data
EBP  INT     1
CORR OMIT DU   BEG
   0.00000
   1  DU1   DU     M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DU2   DU     M   1   0  -1   1.7321   0.0000   0.0000   0.00000
   3  DU3   DU     M   2   1   0   2.2361 140.7685 -90.0000   0.00000
   4   S1   S      M   3   2   1   2.8252  89.6372 247.4061  -0.233548
   5   C2   CT     M   4   3   2   1.7679  71.6890 229.9646  -0.241637
   6   H3   H1     E   5   4   3   1.1216 104.8484 156.2402   0.214012
   7   H4   H1     E   5   4   3   1.1185 110.5377 272.0579   0.181152
   8   C5   CT     M   5   4   3   1.5508 115.7728  37.2204   0.076269
   9   C6   C      B   3   2   1   1.5224 128.6055 147.2134   0.669597
  10   O7   O      E   9   3   2   1.2306  65.1922 294.1509  -0.525712
  11   O8   OS     E   9   3   2   1.3765 107.3918 181.0905  -0.257188
  12   C9   CA     M   8   5   4   1.5254 108.4555 165.2585   0.022390
  13  C10   CA     M  12   8   5   1.4033 119.2136  46.2909  -0.285658
  14  H11   HA     E  13  12   8   1.1015 119.5380   3.0877   0.126184
  15  C12   CA     M  13  12   8   1.3921 120.7120 181.6717   0.185421
  16  O13   OM     S  15  13  12   1.3775 125.1511 180.1584  -0.253968
  17  C14   CT     3  16  15  13   1.4233 116.1084 352.1952  -0.073517
  18  H15   H1     E  17  16  15   1.1195 103.2364 184.6011   0.090158
  19  H16   H1     E  17  16  15   1.1169 110.9504 303.2670   0.090158
  20  H17   H1     E  17  16  15   1.1169 110.4872  66.2787   0.090158
  21  C18   C      M  15  13  12   1.4219 119.4977 359.3408   0.330420
  22  O19   OH     S  21  15  13   1.3712 116.9131 179.8343  -0.666080
  23  H20   HO     E  22  21  15   0.9696 107.8348 178.5601   0.476309
  24  C21   CA     M  21  15  13   1.3926 119.5564 359.4825  -0.408728
  25  H22   HA     E  24  21  15   1.1005 119.8911 180.7655   0.172632
  26  C23   CA     M  12   8   5   1.3993 121.0679 226.3069   0.051134
  27  C24   CT     M  26  12   8   1.4874 122.1520 358.7466   0.023490
  28  H25   HC     E  27  26  12   1.1237 108.9215 107.1643   0.032940
  29  H26   HC     E  27  26  12   1.1223 109.9636 224.4217   0.016746
  30  C27   CT     M  27  26  12   1.5249 111.4835 345.4285   0.016629
  31  H28   HC     E  30  27  26   1.1252 109.4265 164.9416   0.055146
  32  H29   HC     E  30  27  26   1.1283 109.3266 283.1164   0.099782
  33  N30   N3     S  30  27  26   1.4471 114.8116  43.8175  -0.644849
  34  H31   H      E  33  30  27   1.0037 110.7201  65.1486   0.368700

LOOP
 C9  C23
 N30 C5

IMPROPER
 C10 C23 C9  C5
 C9  C21 C23 C24

DONE
STOP
```

```
0   0   2

B-Zalypsis®
etg.data
ZSC  INT     1
CORR OMIT DU   BEG
   0.00000
   1   DU   DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2   DU   DU    M   1   0  -1   3.0000   0.0000   0.0000   0.00000
   3   DU   DU    M   2   1   0   1.0000  90.0000 -90.0000   0.00000
   4   N1   N     M   3   2   1   1.4115 128.2046 152.6264  -0.489819
   5   H2   H     E   4   3   2   1.0900 119.4173  50.5454   0.319854
   6   C3   C     M   4   3   2   1.3841 121.1203 230.5565   0.706672
   7   O4   O     E   6   4   3   1.4299 123.5333 359.9648  -0.574691
   8   C5   C2    M   6   4   3   1.5529 112.9425 180.0406  -0.438300
   9   H6   HA    E   8   6   4   1.0901 125.6320  10.6025   0.160270
  10   C7   C2    M   8   6   4   1.5487 108.7955 190.5617   0.014744
  11   H8   HA    E  10   8   6   1.0892 120.0165 359.9439   0.131658
  12   C9   CA    M  10   8   6   1.5394 119.9741 180.0342   0.017582
  13  C10   CA    M  12  10   8   1.3994 119.9957 352.2553  -0.081395
  14  H11   HA    E  13  12  10   1.0802 120.0923   0.0000   0.141680
  15  C12   CA    M  13  12  10   1.4007 119.9472 180.1008  -0.210119
  16  H13   HA    E  15  13  12   1.0806 120.0067 180.0051   0.163857
  17  C14   CA    M  15  13  12   1.4002 120.0266 359.9228  -0.091230
  18  H15   HA    E  17  15  13   1.0792 120.0073 180.0570   0.151678
  19  C16   CA    M  17  15  13   1.4002 119.9556   0.1319  -0.064273
  20  C17   CT    3  19  17  15   1.5400 119.9212 179.9508   0.546921
  21  F18   F     E  20  19  17   1.3799 109.5118  81.2203  -0.191482
  22  F19   F     E  20  19  17   1.3798 109.4827 201.2046  -0.191482
  23  F20   F     E  20  19  17   1.3794 109.5407 321.1846  -0.191482
  24  C21   CA    M  19  17  15   1.3990 120.0271 359.8793  -0.132499
  25  H22   HA    E  24  19  17   1.0807 119.9760 180.0895   0.158320

LOOP
 C9 C21

IMPROPER
 N1 C5 C3 O4
 C3 C7 C5 H6
 C5 C9 C7 H8
 C9 C12 C10 H11
 C10 C14 C12 H13
 C12 C16 C14 H15
 C14 C21 C16 C17
 C16 C9 C21 H22

DONE
STOP
```

```
0    0    2

B-PM01183®
CSB.data
CSB INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU    M   1   0  -1   1.9401   0.0000   0.0000   0.00000
   3  DUMM  DU    M   2   1   0   1.3357  59.8192 -90.0000   0.00000
   4  O1    OS    M   2   1   3   3.1738 111.2214  42.9335  -0.185378
   5  C2    C     M   4   2   1   1.3552  57.1667 238.2327   0.655149
   6  O3    O     E   5   4   2   1.2023 121.8433  41.4872  -0.534981
   7  C4    CT    M   5   4   2   1.5203 112.0809 223.3956  -0.082233
   8  C5    CT    3   7   5   4   1.5318 111.7201 313.2411  -0.369968
   9  H6    HC    E   8   7   5   1.0738 112.0839  61.5723   0.202728
  10  H7    HC    E   8   7   5   1.0762 107.5028 183.1687   0.202728
  11  S8    S     E   8   7   5   1.8952 112.3475 300.0645  -0.238389
  12  C9    CB    M   7   5   4   1.5143 108.2095  73.8198   0.026962
  13  C10   CB    M  12   7   5   1.4540 131.0002 291.1659  -0.027619
  14  C11   CA    M  13  12   7   1.3887 134.6917   5.3252  -0.171541
  15  H12   HA    E  14  13  12   1.0648 121.7123   2.2333   0.125407
  16  C13   CA    M  14  13  12   1.3787 119.3358 180.4984   0.206464
  17  O14   OS    S  16  14  13   1.3818 115.6175 179.9303  -0.322386
  18  C15   CT    3  17  16  14   1.4329 120.7140 180.7068  -0.041083
  19  H16   HC    E  18  17  16   1.0776 105.6597 179.4255   0.080427
  20  H17   HC    E  18  17  16   1.0844 111.5530 298.2055   0.080427
  21  H18   HC    E  18  17  16   1.0835 111.5605  60.6218   0.080427
  22  C19   CA    M  16  14  13   1.3940 120.9386 359.3847  -0.158858
  23  H20   HA    E  22  16  14   1.0689 120.5364 180.1704   0.166018
  24  C21   CA    M  22  16  14   1.3800 120.5338 359.7933  -0.278449
  25  H22   HA    E  24  22  16   1.0720 120.2509 180.4373   0.180788
  26  C23   CB    M  24  22  16   1.3794 118.6211   0.4497   0.115320
  27  N24   NB    M  26  24  22   1.3847 130.9036 180.4461  -0.432363
  28  H25   H     E  27  26  24   0.9951 125.7544   0.6184   0.352484
  29  C26   CB    M  12   7   5   1.3529 121.7854 115.5170   0.116695
  30  C27   CT    M  29  12   7   1.4960 125.6208 355.2344  -0.053649
  31  H28   HC    E  30  29  12   1.0842 110.5987 219.7478   0.058208
  32  H29   HC    E  30  29  12   1.0842 109.9693  99.2014   0.058208
  33  C30   CT    M  30  29  12   1.5354 107.2085 340.5377  -0.054935
  34  H31   HC    E  33  30  29   1.0827 110.1850 287.9878   0.082237
  35  H32   HC    E  33  30  29   1.0819 109.9842 168.0377   0.082237
  36  N33   NT    M   7   5   4   1.4638 110.3625 192.8341  -0.449255
  37  H34   H     E  36   7   5   1.0046 111.0457  54.5224   0.316177

IMPROPER
 O1   O3   C2   C4
 C4   C26  C9   C10
 C9   C23  C10  C11
 C10  H12  C11  C13
 C11  O14  C13  C19
 C13  H20  C19  C21
 C19  H22  C21  C23
 C21  C10  C23  N24
 C23  H25  N24  C26
 N24  C9   C26  C27

LOOP
 C9  C26
 C10 C23
 C4  N33
```

```
0   0   2

C-Yondelis® and C-PM01183®
etc.data
EAP  INT     1
CORR OMIT DU   BEG
   0.00000
   1  DU1   DU     M    0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DU2   DU     M    1   0  -1   1.7321   0.0000   0.0000   0.00000
   3  DU3   DU     M    2   1   0   1.4142 144.7356 -90.0000   0.00000
   4   C1   CT     M    3   2   1   4.5092  99.1131  96.6553  -0.071289
   5   H2   H1     E    4   3   2   1.1190 143.6557 187.9485   0.095758
   6   H3   H1     E    4   3   2   1.1248 105.0016 357.1723   0.095758
   7   C4   CT     M    4   3   2   1.5519  61.2800 100.7672   0.010826
   8   H5   HC     E    7   4   3   1.1381 106.0020 202.8745   0.067203
   9   C6   CA     M    7   4   3   1.5065 109.3189 316.9118  -0.117999
  10   C7   CA     M    9   7   4   1.3822 120.2523 297.8663   0.261116
  11   O8   OS     M   10   9   7   1.3905 128.3380 359.5376  -0.351784
  12   C9   CT     M   11  10   9   1.4496 106.9184 180.5403   0.200592
  13  H10   H2     E   12  11  10   1.1183 108.5688 242.3847   0.115380
  14  H11   H2     E   12  11  10   1.1171 108.6191 117.0935   0.095791
  15  O12   OS     M   12  11  10   1.4473 107.4825 359.7093  -0.358820
  16  C13   CA     M   15  12  11   1.3884 107.0656   0.3940   0.139095
  17  C14   CA     M   16  15  12   1.3831 128.1410 179.0926   0.017994
  18  C15   CT     3   17  16  15   1.4756 122.8561   0.7596  -0.110000
  19  H16   HC     E   18  17  16   1.1209 109.7535 118.5516   0.060705
  20  H17   HC     E   18  17  16   1.1192 110.1449 237.7233   0.060705
  21  H18   HC     E   18  17  16   1.1179 110.8324 358.2742   0.060705
  22  C19   CA     M   17  16  15   1.4208 114.7930 180.9824   0.032043
  23  O20   OS     S   22  17  16   1.3926 118.3751 173.0642  -0.352833
  24  C21   C      B   23  22  17   1.3865 119.1251  77.3707   0.705103
  25  O22   O      E   24  23  22   1.2271 118.4742   2.3209  -0.506715
  26  C23   CT     3   24  23  22   1.4875 111.6411 182.3908  -0.294508
  27  H24   HC     E   26  24  23   1.1185 108.7044 121.5081   0.096999
  28  H25   HC     E   26  24  23   1.1186 108.6725 240.0117   0.096999
  29  H26   HC     E   26  24  23   1.1159 111.3011   0.8052   0.096999
  30  C27   CA     M   22  17  16   1.4040 123.3773 359.2568  -0.089816
  31  C28   CT     M   30  22  17   1.4879 121.3734 175.9710  -0.031982
  32  H29   H1     E   31  30  22   1.1294 110.1951 330.5218   0.165564
  33  C30   CT     B   31  30  22   1.5489 106.6851 214.5237   0.102017
  34  H31   H1     E   33  31  30   1.1425 105.4199  52.4422   0.159013
  35  N32   N3     E   33  31  30   1.4567 109.4527 294.9026  -0.117754

LOOP
 C7  C13
 C27 C6
 N32 C4

IMPROPER
 C6  C13 C7  O8
 C7  C14 C13 O12
 C6  C19 C27 C28
 C7  C27 C6  C4
 O20 C23 C21 O22


DONE
STOP
```

```
0   0   2

C-Zalypsis®
zap.data
ZAP  INT     1
CORR OMIT DU   BEG
   0.00000
   1    DU   DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2    DU   DU    M   1   0  -1   1.6310   0.0000   0.0000   0.00000
   3    DU   DU    M   2   1   0   1.3006 115.3522 -90.0000   0.00000
   4    C1   CT    M   3   2   1   1.5009 129.2206  17.0884  -0.071289
   5    H2   H1    E   4   3   2   1.1192 118.9823 245.4602   0.095758
   6    H3   H1    E   4   3   2   1.1249  91.2406   0.0000   0.095758
   7    C4   CT    M   4   3   2   1.5521 113.8619 111.2733   0.010826
   8    H5   HC    E   7   4   3   1.1379 106.0072 199.9143   0.067203
   9    C6   CA    M   7   4   3   1.5065 109.3149 313.9921  -0.117999
  10    C7   CA    M   9   7   4   1.3824 120.2383 297.7807   0.261116
  11    O8   OS    M  10   9   7   1.3906 128.3515 359.6065  -0.351784
  12    C9   CT    M  11  10   9   1.4491 106.9260 180.4766   0.200592
  13   H10   H2    E  12  11  10   1.1176 108.6035 242.3799   0.115380
  14   H11   H2    E  12  11  10   1.1171 108.6160 117.0655   0.095791
  15   O12   OS    M  12  11  10   1.4479 107.4586 359.7366  -0.358820
  16   C13   CA    M  15  12  11   1.3876 107.0918   0.3628   0.139095
  17   C14   CA    M  16  15  12   1.3826 128.1386 179.1171   0.017994
  18   C15   CT    3  17  16  15   1.4755 122.8830   0.7717  -0.110000
  19   H16   HC    E  18  17  16   1.1209 109.7801 118.5520   0.060705
  20   H17   HC    E  18  17  16   1.1197 110.1173 237.7142   0.060705
  21   H18   HC    E  18  17  16   1.1179 110.8328 358.2415   0.060705
  22   C19   CA    M  17  16  15   1.4213 114.7949 180.9866   0.032043
  23   O20   OS    S  22  17  16   1.3926 118.3860 173.0939  -0.352833
  24   C21   C     B  23  22  17   1.3860 119.1502  77.3482   0.705103
  25   O22   O     E  24  23  22   1.2268 118.4974   2.3007  -0.506715
  26   C23   CT    3  24  23  22   1.4875 111.6382 182.4124  -0.294508
  27   H24   HC    E  26  24  23   1.1180 108.7414 121.5120   0.096999
  28   H25   HC    E  26  24  23   1.1190 108.6615 239.9955   0.096999
  29   H26   HC    E  26  24  23   1.1165 111.3147   0.7940   0.096999
  30   C27   CA    M  22  17  16   1.4036 123.3731 359.2444  -0.089816
  31   C28   CT    M  30  22  17   1.4883 121.3853 175.9972  -0.031982
  32   H29   H1    E  31  30  22   1.0904 106.6561  94.5165   0.080000
  33   H30   H1    E  31  30  22   1.0899 106.6577 334.4927   0.080000
  34   C31   CT    M  31  30  22   1.5486 106.6763 214.5100   0.102017
  35   H32   H1    E  34  31  30   1.1423 105.4350  52.4035   0.159013
  36   N33   N3    M  34  31  30   1.4564 109.4634 294.9089  -0.117754

LOOP
 C7  C13
 C27 C6
 N33 C4

IMPROPER
 C6  C13 C7  O8
 C7  C14 C13 O12
 C6  C19 C27 C28
 C7  C27 C6  C4
 O20 C23 C21 O22


DONE
STOP
```

```
0     0     2

MMC covalently bonded to first Guanine
GMM.data
GMM INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU    M   1   0  -1   1.5502   0.0000   0.0000   0.00000
   3  DUMM  DU    M   2   1   0   1.4595 109.5041 -90.0000   0.00000
   4  P     P     M   3   2   1   1.6098 109.5229 171.7957   1.07350
   5  O1P   O2    E   4   3   2   1.4760 107.8942 128.9165  -0.75450
   6  O2P   O2    E   4   3   2   1.4829 107.0576   3.4646  -0.75450
   7  O5'   OS    M   4   3   2   1.5945 105.9806 246.7548  -0.41280
   8  C5'   CI    M   7   4   3   1.4498 118.7822 312.9643  -0.01110
   9  H5'1  H1    E   8   7   4   1.0900 109.7613  93.9934   0.06620
  10  H5'2  H1    E   8   7   4   1.0895 109.7582 333.9654   0.06620
  11  C4'   CT    M   8   7   4   1.5109 109.7677 213.9796   0.07210
  12  H4'   H1    E  11   8   7   1.0901 107.7608 156.3397   0.08870
  13  O4'   OS    S  11   8   7   1.4632 108.6489 279.1130  -0.31240
  14  C1'   CT    B  13  11   8   1.4145 109.7846 106.4011   0.15570
  15  H1'   H2    E  14  13  11   1.0901 107.9285 115.7842   0.08860
  16  N9    N*    S  14  13  11   1.4877 107.7679 232.4945  -0.038777
  17  C8    CK    B  16  14  13   1.3792 129.2830  82.1486   0.123243
  18  H8    H5    E  17  16  14   1.0899 122.9320 359.9051   0.154475
  19  N7    NB    S  17  16  14   1.3081 114.1026 179.8800  -0.559339
  20  C5    CB    S  19  17  16   1.3965 103.9975   0.1633   0.294664
  21  C6    C     B  20  19  17   1.4133 130.2191 179.8644   0.388537
  22  O6    O     E  21  20  19   1.2309 129.0731   0.0000  -0.530049
  23  N1    NA    B  21  20  19   1.4059 111.4324 179.8800  -0.216712
  24  H1    H     E  23  21  20   1.0303 117.4688 179.6679   0.276784
  25  C2    CA    B  23  21  20   1.3798 125.0383 359.6854   0.105208
  26  N2    N2    B  25  23  21   1.3366 116.1488 180.2365  -0.309644
  27  H22   H     E  26  25  23   1.0305 120.0123   0.0000   0.274286
  28  C1    CT    B  26  25  23   1.5127 117.6538 173.7370   0.050471
  29  H11   HC    E  28  26  25   1.0809 104.2807 342.1348   0.038004
  30  C21   CT    3  28  26  25   1.5724 112.3552 223.5178   0.203713
  31  H2    HC    E  30  28  26   1.0762 115.3011  18.2558   0.087449
  32  N31   N3    3  30  28  26   1.5410 107.4922 145.5955  -0.463291
  33  H31   H     E  32  30  28   1.0190 111.3881 177.5966   0.351639
  34  H32   H     E  32  30  28   1.0219 107.6278  58.7999   0.351639
  35  H33   H     E  32  30  28   1.0178 112.0462 299.1398   0.351639
  36  C41   CT    3  30  28  26   1.5552 105.7249 256.8189   0.002681
  37  H41   HC    E  36  30  28   1.0775 115.2474 210.6396   0.082141
  38  H42   HC    E  36  30  28   1.0818 109.9666  87.2135   0.082141
  39  N5    N2    S  36  30  28   1.4549 100.0325 330.6351  -0.001023
  40  C61   CB    S  39  36  30   1.3479 134.6292 187.9487  -0.119607
  41  C7    C     B  40  39  36   1.4678 126.0789  16.9804   0.401268
  42  O8    O     E  41  40  39   1.2278 119.0215 359.8077  -0.474053
  43  C9    CM    B  41  40  39   1.4451 114.6522 180.0688  -0.063269
  44  C10   CT    3  43  41  40   1.5103 117.5083 179.3860  -0.156271
  45  H101  HC    E  44  43  41   1.0870 111.3755 119.0958   0.068978
  46  H102  HC    E  44  43  41   1.0853 111.4504 240.3782   0.068978
  47  H103  HC    E  44  43  41   1.0803 109.5025 359.8728   0.068978
  48  C11   CM    B  43  41  40   1.3533 121.5368 359.2821   0.014983
  49  N12   N2    B  48  43  41   1.3384 124.9855 180.3719  -0.673605
  50  H121  H     E  49  48  43   0.9994 118.7702 180.1443   0.372075
  51  H122  H     E  49  48  43   0.9966 122.2374 359.5462   0.372075
  52  C13   C     B  48  43  41   1.5272 123.0227 359.7784   0.625178
  53  O14   O     E  52  48  43   1.2130 119.7557 181.9962  -0.450562
  54  C15   CB    S  40  39  36   1.3576 107.7085 196.8374  -0.150074
  55  C16   CC    B  54  40  39   1.4293 108.6265   2.0357  -0.036816
  56  C17   CT    B  55  54  40   1.4987 127.5583 175.6767   0.029926
  57  H171  HC    E  57  55  54   1.0774 108.9511  23.8400   0.086068
  58  H172  HC    E  57  55  54   1.0821 109.0602 143.5287   0.086068
  59  C18   CC    E  55  54  40   1.3565 105.4075 357.8175  -0.110424
  60  N3    NC    S  25  23  21   1.3304 123.5537   0.7501  -0.211589
  61  C4    CB    E  60  25  23   1.3591 112.3692 359.4493   0.024855
  62  C3'   CT    M  11   8   7   1.5272 116.3826  36.3615   0.16450
```

```
  63  H3'   H1    E  62  11   8   1.0901 111.1425  33.6094   0.06230
  64  C2'   CT    B  62  11   8   1.5202 102.8645 273.6140  -0.05610
  65  H2'1  HC    E  64  62  11   1.0903 111.3657 205.0801   0.04860
  66  H2'2  HC    E  64  62  11   1.0894 111.3650  85.0665   0.04860
  67  O3'   OS    M  62  11   8   1.4255 112.2122 156.4009  -0.49980

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 C2   C1   N2   H22
 N7   N9   C8   H8
 N1   N3   C2   N2
 C41  C61  N5   C18
 C1   C16  C18  N5
 C7   C15  C61  N5
 C9   C13  C11  N12
 C21  C18  C1   N2
 C13  C16  C15  C61

LOOP
 C1'  C2'
 C4   C5
 N5   C18
 C61  C15
 C18  C1
 C4   N9

DONE
STOP
```

```
0    0    2

Second guanine covalently bonded to MMC
GMO.data
GMO INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU    M   1   0  -1   1.5502   0.0000   0.0000   0.00000
   3  DUMM  DU    M   2   1   0   1.4595 109.5041 -90.0000   0.00000
   4  P     P     M   3   2   1   1.6098 109.5229 171.7957   1.07350
   5  O1P   O2    E   4   3   2   1.4760 107.8942 128.9165  -0.75450
   6  O2P   O2    E   4   3   2   1.4829 107.0576   3.4646  -0.75450
   7  O5'   OS    M   4   3   2   1.5945 105.9806 246.7548  -0.41280
   8  C5'   CI    M   7   4   3   1.4498 118.7822 312.9643  -0.01110
   9  H5'1  H1    E   8   7   4   1.0900 109.7613  93.9934   0.06620
  10  H5'2  H1    E   8   7   4   1.0895 109.7582 333.9654   0.06620
  11  C4'   CT    M   8   7   4   1.5109 109.7677 213.9796   0.07210
  12  H4'   H1    E  11   8   7   1.0901 107.7608 156.3397   0.08870
  13  O4'   OS    S  11   8   7   1.4632 108.6489 279.1130  -0.31240
  14  C1'   CT    B  13  11   8   1.4145 109.7846 106.4011   0.15570
  15  H1'   H2    E  14  13  11   1.0901 107.9285 115.7842   0.08860
  16  N9    N*    S  14  13  11   1.4877 107.7679 232.4945  -0.038777
  17  C8    CK    B  16  14  13   1.3792 129.2830  82.1486   0.123243
  18  H8    H5    E  17  16  14   1.0899 122.9320 359.9051   0.154475
  19  N7    NB    S  17  16  14   1.3081 114.1026 179.8800  -0.559339
  20  C5    CB    S  19  17  16   1.3965 103.9975   0.1633   0.294664
  21  C6    C     B  20  19  17   1.4133 130.2191 179.8644   0.388537
  22  O6    O     E  21  20  19   1.2309 129.0731   0.0000  -0.530049
  23  N1    NA    B  21  20  19   1.4059 111.4324 179.8800  -0.216712
  24  H1    H     E  23  21  20   1.0303 117.4688 179.6679   0.276784
  25  C2    CA    B  23  21  20   1.3798 125.0383 359.6854   0.105208
  26  N2    N2    S  25  23  21   1.3366 116.1488 180.2365  -0.309644
  27  H22   H     E  26  25  23   1.0305 120.0123   0.0000   0.274286
  28  N3    NC    S  25  23  21   1.3304 123.5537   0.7501  -0.211589
  29  C4    CB    E  28  25  23   1.3591 112.3692 359.4493   0.024855
  30  C3'   CT    M  11   8   7   1.5272 116.3826  36.3615   0.16450
  31  H3'   H1    E  30  11   8   1.0901 111.1425  33.6094   0.06230
  32  C2'   CT    B  30  11   8   1.5202 102.8645 273.6140  -0.05610
  33  H2'1  HC    E  32  30  11   1.0903 111.3657 205.0801   0.04860
  34  H2'2  HC    E  32  30  11   1.0894 111.3650  85.0665   0.04860
  35  O3'   OS    M  30  11   8   1.4255 112.2122 156.4009  -0.49980

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 N7   N9   C8   H8
 N1   N3   C2   N2

LOOP
 C1'  C2'
 C4   C5
 C4   N9

DONE
STOP
```

Modifications to the parmbsc0 force field for *YONDELIS®*, *ZALYPSIS®* and PM01183®

MASS

```
OM 16.00                              Yondelis
C2 12.01                              Zalypsis
```

BOND

```
NH-H   434.0    1.010       YONDELIS
CT-NH  337.0    1.450       YONDELIS
CA-OH  450.0    1.380       YONDELIS
CA-OS  320.0    1.380       YONDELIS
C -OS  450.0    1.380       YONDELIS
CA-OM  450.0    1.364       YONDELIS
OM-CT  320.0    1.410       YONDELIS
C -C2  469.0    1.490       ZALYPSIS
C2-CA  469.0    1.470       ZALYPSIS
C2-C2  469.0    1.320       ZALYPSIS
C2-HA  367.0    1.080       ZALYPSIS
CB-CT  480.0    1.500       PM01183
NB-H   434.0    1.010       PM01183
```

ANGLE

```
CT-NH-CT    50.0      114.33    YONDELIS
C -CT-NH    80.0      108.28    YONDELIS
CT-NH-H     50.0      110.00    YONDELIS
CT-CT-NH    80.0      114.96    YONDELIS
H1-CT-NH    50.0      107.50    YONDELIS
CA-CA-OM    70.0      120.00    YONDELIS
CA-OM-CT    70.0      115.00    YONDELIS
H1-CT-OM    50.0      109.50    YONDELIS
OS-CT-OS    80.0      109.50    YONDELIS
C -CA-OM    70.0      120.00    YONDELIS
CA-CT-S     50.0      114.70    YONDELIS
CA-CT-H1    50.0      109.50    YONDELIS
C -CA-CT    70.0      120.00    YONDELIS
CA-CT-C     63.0      109.50    YONDELIS
O -C -OS    70.0      119.00    YONDELIS
CT-C -OS    70.0      112.00    YONDELIS
CA-CA-OS    70.0      120.00    YONDELIS
C -OS-CA    70.0      120.00    YONDELIS
CA-OS-CT    60.0      106.50    YONDELIS
C -OS-CT    70.0      117.00    YONDELIS
HC-CT-N3    35.0      109.50    YONDELIS
H1-CT-N3    35.0      109.50    YONDELIS
CA-CT-N3    80.0      111.00    YONDELIS
N3-CT-OH    70.0      109.50    YONDELIS
CT-C -CT    40.0      109.50    YONDELIS
N2-CT-HC    35.0      109.50    YONDELIS
N2-CT-N3    70.0      120.00    YONDELIS
CA-C -OS    70.0      120.00    YONDELIS
OS-CT-C     70.0      112.00    YONDELIS
C -CT-C     63.0      109.50    YONDELIS
N -C -C2    70.0      114.93    ZALYPSIS
C -C2-HA    35.0      118.06    ZALYPSIS
C -C2-C2    63.0      119.92    ZALYPSIS
O -C -C2    70.0      122.79    ZALYPSIS
C2-C2-CA    63.0      127.31    ZALYPSIS
C2-C2-HA    35.0      122.01    ZALYPSIS
CA-CA-C2    63.0      123.13    ZALYPSIS
CA-C2-HA    35.0      116.12    ZALYPSIS
CA-CT-F     50.0      111.73    ZALYPSIS
HC-CT-NT    50.0      110.00    PM01183
CB-CT-HC    50.0      109.00    PM01183
CB-CT-CT    63.0      105.00    PM01183
H -NB-CB    50.0      124.00    PM01183
NB-CB-CT    70.0      123.82    PM01183
CB-NB-CB    70.0      110.12    PM01183
OS-CT-HC    50.0      109.50    PM01183
CB-CB-CB    63.0      106.00    PM01183
```

```
CB-CB-CT    63.0      125.53    PM01183
HC-CT-S     50.0      108.50    PM01183
C -CT-CB    63.0      127.41    PM01183
C -CT-NT    70.0      109.67    PM01183
CB-CT-NT    80.0      103.47    PM01183

DIHE
X -CA-OM-X    2    1.8         180.0            2.         YONDELIS
X -CT-OM-X    3    1.15          0.0            3.         YONDELIS
X -C -OS-X    1    2.25        180.0            2.         YONDELIS
CT-OS-C -O    1    3.85        180.0           -2.         YONDELIS
CT-OS-C -O    1    0.965       180.0            1.         YONDELIS
CT-OS-C -CT   2    4.10        180.0            2.         YONDELIS
CA-CA-OS-CT   2    5.10        180.0            2.         YONDELIS
X -C -C2-X    4   14.50        180.0            2.         ZALYPSIS
X -C2-CA-X    4   14.50        180.0            2.         ZALYPSIS
X -C2-C2-X    4   14.50        180.0            2.         ZALYPSIS
CA-CA-OS-C    2    0.00          0.0            2.         ZALYPSIS
CB-CB-CT-NT   1    0.00          0.0            2.         PM1183
NB-CB-CT-HC   1    0.00          0.0            2.         PM1183
NB-CB-CT-CT   1    0.00          0.0            2.         PM1183
CB-CB-CT-HC   1    0.00          0.0            2.         PM1183
CB-CB-CT-CT   1    0.00          0.0            2.         PM1183
C -CT-CB-CB   1    0.00          0.0            2.         PM1183

IMPR
X -X -CA-OM         1.8          180.          2.          YONDELIS
X -C -CA-X          1.1          180.          2.          YONDELIS
X -CA-C -X          1.1          180.          2.          YONDELIS
CA-C -CA-CT         1.1          180.          2.          YONDELIS
X -X -C2-H          1.0          180.          2.          ZALYPSIS
X -X -CA-C2         1.1          180.          2.          ZALYPSIS

NONBON
C2          1.9080  0.1094            copied from CT Spellmeyer (parmbsc0)
OM          1.6837  0.1700            YONDELIS
```

Modifications to the parmbsc0 force field for MMC

BOND

CC–CC	480.0	1.360	MMC
CB–CC	480.0	1.430	MMC
CM–N2	445.0	1.340	MMC
N2–CC	340.0	1.380	MMC
N2–CB	340.0	1.350	MMC
C –N2	425.0	1.340	MMC
CC–HC	340.0	1.080	MMC
N3–CB	436.0	1.410	MMC
N3–CC	436.0	1.370	MMC
CC–HA	340.0	1.070	MMC

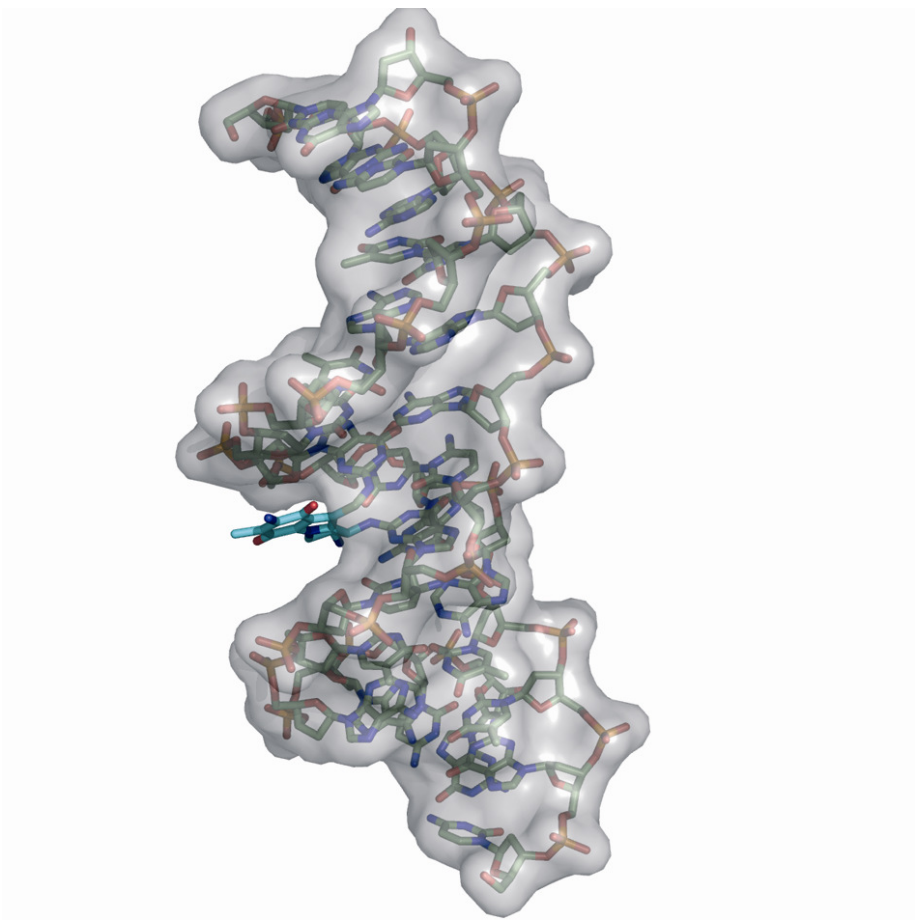
ANGLE

N2–CT–CC	70.0	114.35	MMC
CC–CC–CT	70.0	135.00	MMC
CB–CC–CC	70.0	105.40	MMC
C –CB–CC	65.0	131.80	MMC
N2–CM–C	70.0	112.00	MMC
CM–N2–H	30.0	112.30	MMC
CM–C –CB	70.0	114.96	MMC
CM–CM–N2	70.0	124.97	MMC
CB–N2–CC	70.0	109.30	MMC
CB–C –CM	70.0	114.65	MMC
CB–CB–CC	65.0	108.65	MMC
N2–CB–C	70.0	126.06	MMC
N2–CB–CB	70.0	107.70	MMC
N2–CC–CC	70.0	108.90	MMC
CT–N2–CB	60.0	134.65	MMC
CT–N2–CC	60.0	114.00	MMC
HC–CT–N3	50.0	113.25	MMC
CT–CC–N2	70.0	109.47	MMC
CB–CC–CT	70.0	127,56	MMC
OS–C –N	80.0	110.93	MMC
CC–CT–OS	50.0	106.94	MMC
O –C –N2	80.0	126.20	MMC
C –N2–H	33.0	120.40	MMC
OS–C–N2	80.0	110.40	MMC
HC–CC–CT	50.0	108.00	MMC
CT–CC–CT	40.0	114.0	MMC
CB–CC–HC	50.0	112.0	MMC
N2–CT–OS	50.0	113.0	MMC
CT–N2–CT	50.0	59.80	MMC
CB–N3–CC	70.0	110.70	MMC
N3–CB–C	70.0	127.50	MMC
N3–CB–CB	70.0	107.70	MMC
N3–CC–CC	70.0	108.50	MMC
CT–N3–CB	60.0	134.30	MMC
CT–N3–CC	60.0	115.03	MMC
OH–CA–CB	70.0	121.80	MMC
CA–CB–CC	63.0	134.00	MMC
N2–CA–CA	70.0	118.70	MMC
N2–CB–CA	70.0	130.25	MMC
HA–CC–CT	35.0	126.10	MMC
HA–CC–CC	35.0	123.75	MMC
CC–CT–N3	60.0	109.35	MMC
CC–CC–CC	60.0	144.03	MMC
OS–CT–HC	50.0	111.00	MMC
N2–CT–HC	50.0	107.50	MMC

```
DIHE
X -CM-N2-X    4    9.60        180.0          2.         MMC
X -CB-N2-X    4    9.60        180.0          2.         MMC
X -CB-CC-X    4   14.50        180.0          2.         MMC
X -N2-CC-X    4    9.60        180.0          2.         MMC
X -CC-CC-X    4   14.50        180.0          2.         MMC
X -C -N2-X    4    5.80        180.0          2.         MMC
```

ARTÍCULO V.

"Rationale for the opposite stereochemistry of the major monoadducts and interstrand crosslinks formed by mitomycin C and its decarbamoylated analogue at CpG steps in DNA and the effect of cytosine modification on reactivity"



La mitomicina C (MMC) es un potente agente antitumoral que forma un enlace covalente con el grupo 2-amino de determinadas guaninas en el surco menor del ADN de doble cadena después de una reducción intracelular de su anillo de quinona y la apertura del grupo aziridinio. En algunos pasos 5'-CG-3' (CpG) el aducto monofuncional resultante puede evolucionar hacia un aducto bifuncional que se conoce como entrecruzamiento intercatenario o *"interstrand crosslink"* (ICL). La reactividad de MMC se incrementa cuando las bases de citosina se encuentran metiladas (5MC) y disminuye cuando se sustituyen con 5-F-citosina (5FC). Hemos estudiado tres oligodesoxinucleótidos de doble cadena de secuencia general d(CGATAAXGCTAACG) en el cual X representa C, 5MC o 5FC. Usando una combinación de simulaciones de MD en solución acuosa, mecánica cuántica y cálculos electrostáticos, hemos sido capaces de (i) caracterizar los efectos de la modificación de la citosina en la estructura y en las propiedades de la molécula de ADN, (ii) obtener una serie de complejos estructurales que facilitan la comprensión de la formación de tanto del monoaducto como del ICL por MMC con un nivel de detalle atómico, (iii) mostrar cómo el ADN es capaz de adaptarse a ambas modificaciones covalentes en el surco menor experimentando sólo una mínima distorsión, y (iv) proporcionar una explicación para la reactividad alterada de MMC hacia las moléculas de ADN modificadas en las bases de citosina.

Rationale for the opposite stereochemistry of the major monoadducts and interstrand crosslinks formed by mitomycin C and its decarbamoylated analogue at CpG steps in DNA and the effect of cytosine modification on reactivity

Juan A. Bueren-Calabuig, Ana Negri,[1] Antonio Morreale[2] and Federico Gago[*]

Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain

ABSTRACT (203 words): Mitomycin C (MMC) is a potent antitumour agent that forms a covalent bond with the 2-amino group of selected guanines in the minor groove of double-stranded DNA following intracellular reduction of its quinone ring and opening of its aziridine moiety. At some 5'-CG-3' (CpG) steps the resulting monofunctional adduct can evolve towards a more deleterious bifunctional lesion, which is known as an interstrand crosslink (ICL). MMC reactivity is enhanced when the cytosine bases are methylated (5MC) and decreased when they are replaced with 5-F-cytosine (5FC) whereas the stereochemical preference of alkylation changes upon decarbamoylation. We have studied three duplex oligonucleotides of general formula d(CGATAAXGCTAACG) in which X stands for C, 5MC or 5FC. Using a combination of molecular dynamics simulations in aqueous solution, quantum mechanics and continuum electrostatics, we have been able to (i) obtain a large series of snapshots that facilitate an understanding in atomic detail of the distinct stereochemistry of monoadduct and ICL formation by MMC and its decarbamoylated analogue, (ii) provide an explanation for the altered reactivity of MMC towards DNA molecules containing 5MC or 5FC, and (iii) show the distinct accommodation in the DNA minor groove of the different covalent modifications, particularly the most cytotoxic C1α and C1β ICLs.

Keywords: DNA-binding antitumour agents; molecular dynamics simulations; quantum mechanics; QM/MM methods; methylated cytosine.

Running title: Monoadduct and interstrand crosslink formation by mitomycin C.

[1]*Present address:* Department of Structural and Chemical Biology, Mount Sinai School of Medicine, New York, NY 10029-657, USA.
[2]Unidad de Bioinformática, Centro de Biología Molecular Severo Ochoa, 28049 Madrid, Spain

Corresponding author: Prof. Federico Gago, Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain. Telephone: +34 - 918 854 514; Fax: +34 - 918 854 591; E-mail: federico.gago@uah.es.

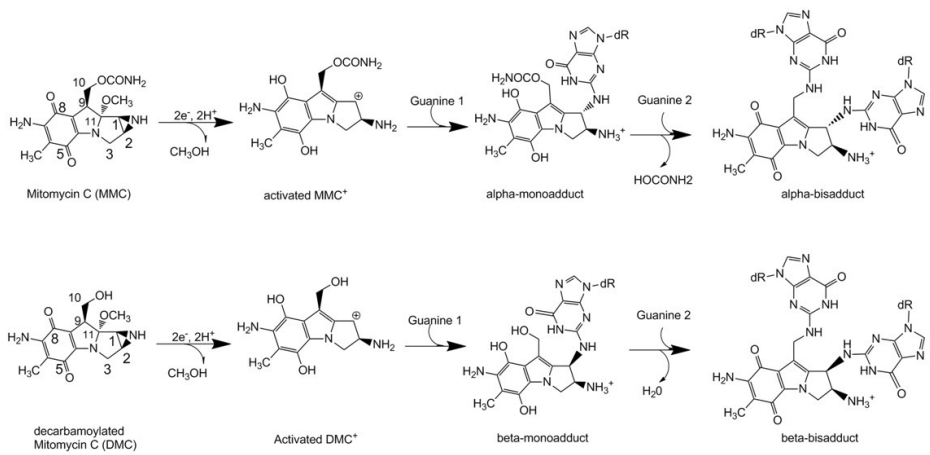
INTRODUCTION

Mitomycin C (MMC) was discovered as a potent anticancer antibiotic produced by *Streptomyces caespitosus* (1) and is now biosynthetically obtained from fermentation cultures of *S. lavendulae* because of the challenge that its total synthesis still represents (2). MMC's broad spectrum of activity against solid tumours was shown to be primarily due to its potent inhibition of DNA replication mostly because of its ability to crosslink the complementary strands of the double helix (3) specifically at 5'-CG-3' (CpG) steps (4,5) in hypoxic cells. Nonetheless, the major product upon reaction of MMC with DNA is actually the monofunctional adduct N^2-(2"β,7"-diaminomitosen-l"α-yl)-2'-deoxyguanosine (6) that is formed upon covalent bond formation with the exocyclic amino group of guanine at position 2 of the purine ring in just one strand (7).

In clinical practice since the 1970's, MMC is intravesically instilled in the chemotherapy of superficial bladder tumours (8) and used as topical adjunctive therapy in various ocular surgeries to inhibit the wound healing response and reduce scarring (9). MMC is also used parenterally, in combination with other approved drugs, in the therapy of disseminated adenocarcinomas of the stomach or pancreas and as palliative treatment of these malignancies when other modalities have failed (10). It is noteworthy that hypersensitivity to MMC is a hallmark of cells deficient in one or more of the proteins involved in homologous recombination repair such as Mus81 (11), XPG (12), Pir51 (13), Brca1 (14), or the Partner and Localizer of Brca2 (PalB2) (15). Interestingly, biallelic inactivating mutations in the gene encoding this latter protein were recently shown to underlie the robust activity of MMC observed in nude mice xenografted with a gemcitabine-resistant human pancreatic cancer as well as the dramatically favourable response attained in the patient from whom the tumour had been resected upon treatment with this drug (16). Moreover, sensitivity to MMC has been used to classify Fanconi anemia (FA) patients into different complementation groups (17), can still be used for diagnostic purposes *in vitro* (18) and has been instrumental to delineate the so-called FA/BRCA pathway of the DNA damage response (19).

The deceptively "simple" structure of MMC (Scheme I) possesses four contiguous stereogenic carbons and economically packs in a constrained architecture (i) a quinone-containing pyrroloindole skeleton, (ii) a methoxy group, (iii) an aziridine ring, and (iv) a carbamate moiety (2). The aziridine at C1 and the carbamate at C10 are critical to its biological activity because they constitute two masked alkylating functionalities. Indeed, MMC itself is remarkably unreactive toward DNA at pH 7–8 (3) but efficient alkylation of this macromolecule is observed following enzymatic, electrochemical or chemical reduction (20), hence its greater cytotoxicity under hypoxic conditions and the recognition of MMC as the first bioreductive (and bifunctional) alkylating agent (3,21). Two-electron reduction of the quinone ring of MMC facilitates methoxide elimination from position 9a, formation of a leuco-aziridinomitosene and opening of the aziridine ring to provide the extended quinone methide that is the initial alkylating entity leading to a DNA monoadduct with a guanine (G) in the minor groove (7,20). At a CpG step, the second alkylating centre of MMC is the iminium ion that forms upon the reverse Michael elimination of carbamic acid at C10 from the monoadduct. It is at this position then that MMC can undergo a second nucleophilic attack either by a water molecule (to yield a C10-decarbamoylated monoadduct) (7) or, exclusively at CpG steps, by the 2-amino group of the guanine in the opposite strand (i.e. the G pairing with the cytosine in CpG) thus giving rise to the ICL (Scheme I). This DNA lesion represents a major threat to cell viability because it is highly effective in blocking replication and transcription forks (22). The fact that, under standard conditions, alkylation at C10 always follows reaction at C1 explains why ICLs in duplex DNA are observed only at CpG sites. Intrastrand crosslinks at GpG steps (23), as well as other non-cytotoxic monoadducts that arise from the MMC metabolite 2,7-diaminomitosene (24), have also been characterized but will not be dealt with in the present work.

The nature of the base on the 3'-side of the guanine that is alkylated by MMC (G) plays a relatively minor role in site selectivity, with reactivity decreasing in the order CGC > CGT > CGG > CGA (5) reportedly due to the possibility of the 3'-pyrimidine assisting in the removal of the proton from the 2-amino group of guanine during the nucleophilic attack (25). On the other hand, the preference of MMC for different dinucleotide



Scheme I. Reaction pathways leading to the formation of α-MMC–DNA and β-DMC–DNA monoadducts and their corresponding bisadducts or interstrand crosslinks (ICL) of opposite stereochemistry. The DNA-alkylating functionalities of MMC and DMC are unmasked by reductive activation.

steps, quantified in terms of yields and reaction rates with guanine, follows the order CpG > GpG >> TpG ≈ ApG (5). The presumed involvement of the 2-amino group of a second guanine in the clear enhancement observed at the former two steps, and particularly at the CpG step, through a putative hydrogen bond with the carbamate oxygen attached to C10 (OM) in activated MMC was indirectly supported by the demonstration that replacement of the guanine in the complementary strand with hypoxanthine (as found in deoxyinosine) resulted in much lower reactivity (5) and also by the finding that MMC analogues lacking this OM did not show any selectivity towards CpG steps (26). Conversely, 5-methylation of the cytosines making up this site was seen to result in a ~2-fold increase in the reactivity of the synthetic oligodeoxynucleotides thus modified. This effect was attributed to either electronic or steric factors, namely, either enhanced nucleophilicity of the 2-amino group of guanine or greater accessibility of the amino group due to a local conformational change (27). Later work using a 162-bp DNA fragment in which all cytosines had been replaced with either 5-methyl-cytosine (5MC) or 5-F-cytosine (5FC) confirmed this increased reactivity for CpG sites containing 5MC and further showed that the opposite was true when the CpG site contained 5FC in lieu of C. The respective stimulatory and inhibitory effects of 5MC and 5FC were shown to be selective to the G that is paired with the modified pyrimidine. In this case, the explanation put forward to account for these reactivities, which were said to correlate with the Hammett σ constants, was that the electron-donating effect of the 5-methyl substituent of the cytosine was transmitted to guanine through H-bonding of the 5MC:G base pair (28). On the basis of density functional calculations on simplified model systems it was later proposed that one of the H-bonding hydrogens from the 2-amino position of the attacking G would be temporarily transferred to the cytosine oxygen with which it pairs to facilitate the reaction (29) but no further support for this hypothesis has been obtained.

It is then clear that MMC is a prodrug unable to bind to DNA unless it has been reduced. But precisely because its reductively activated forms are too reactive and short-lived for binding studies (7,21), examination of the noncovalent association of this drug or its active metabolite(s) with different double helices has not been feasible. Furthermore, the structural work has been very limited despite the early use of CPK (6) and computer-generated models (4,30,31), which only in a few instances were built making use of NMR-derived information on the monoalkylated product (32,33). Thus, to date and to the best of our knowledge, only the solution NMR structure of the d(ICACGTCIT)·d(ACGACGTGC) duplex containing the major MMC monoadduct at the underlined guanine (32) is deposited in the public domain (Nucleic Acid Database (34) entry 199D) and models of an MMC ICL have been reported only for the self-complementary $d(GCATCGATGC)_2$ decamer (4) and $d(TACGTA)_2$ hexamer (33), which do not represent optimal bonding sites. A common feature of all of these models is the proposal that the N^2-guanine-bonded MMC molecule fits snugly into the minor groove without appreciable perturbation of the DNA structure but none of them provides information about the precovalent complex(es) or the rationale for the stereochemistry of the drug-G covalent bond. In fact, the current paradigm holds that the preference for the CpG site is a consequence of the hydrogen bond between the carbamate oxygen of the MMC activated species and the amino group on the guanine of the nonbonding strand, a proposal that appears to be corroborated by the NMR structural work on the monoadduct (32). A role for the carbonyl oxygen in this recognition was ruled out in view of the apparently similar bonding patterns obtained for MMC and 10-decarbamoylmitomycin C (DMC, Scheme I) in early λ exonuclease stop assays (31) and also by the largely decreased sequence selectivity displayed by MMC derivatives bearing a halogen atom in place of the carbamoyl group (26). Nonetheless, although DMC was initially regarded as a less active monofunctional derivative of MMC (35) later work showed it to produce more adducts and be more cytotoxic than MMC in certain cells, including FA fibroblasts, and to give rise to bisadducts (36,37). Strikingly, however, the chirality of the mitosene linkage between DMC and the amino group of guanine was shown to be largely C1-β rather than the C1-α that is overwhelmingly observed for MMC, both in the more readily formed monoadducts and in the less frequent ICLs. Therefore the current paradigm cannot account for the opposite stereochemistry of the major DNA adducts produced in mammalian cells by MMC and DMC and for the differential sensitivity of FA cells to these drugs (37).

To obtain information in atomic detail about the mechanism of CpG recognition and bonding by MMC and DMC, we have focused on a DNA double helix of sequence d(CGATAACGCTAACG) in one strand and the complementary sequence in the opposite strand. Using this 14mer, which embeds a high-affinity CGC site for mono- and bis-alkylation in the middle region, we have simulated in a continuous fashion all the steps of the reaction leading from the initial precovalent complexes to the final (and distinct) ICLs through prior formation of the respective monoadducts. Bond making and breaking during the molecular dynamics (MD) simulations was made possible by coupling a quantum mechanical (QM) function to a classical molecular mechanics (MM) potential, as recently implemented (38) in the popular AMBER suite of programmes (http://ambermd.org/). The main advantage of a hybrid approach involving MD and QM for the solvated drug-DNA complexes over studies that focus on simplified model systems is that changes in atom connectivities and charge redistributions are performed in a dynamic context that realistically allows reorganization of both solutes and water molecules through the whole procedure. Besides, in order to find a rationale to the observation that MMC-DNA monoadduct formation (and therefore subsequent interstrand crosslinking) is altered at CpG sequences bearing cytosine modifications (27,28), simulations were also performed on similar DNA 14mers in which the cytosines making up the central CpG step contained either a methyl group or a fluorine atom at position 5. For comparison purposes and completeness, the three oligodeoxynucleotides studied as targets for MMC and DMC were also simulated in their free states. The trajectories were analyzed in terms of structural parameters, suitable geometries for nucleophilic attack, molecular electrostatic potentials and solvent-corrected binding energies.

MATERIALS AND METHODS

Geometries, AMBER parameters and charges for the non-standard residues

The X-ray crystal structure of MMC (39) was the starting point for model building this prodrug, as well as DMC and their corresponding reactive intermediates. The geometries of MMC, DMC, activated MMC (both as a hemiquinone methide and as a carbocation, MMC^+),

activated DMC (as a carbocation, DMC^+), reduced MMC-G monoadduct, reduced DMC-G monoadduct, activated MMC-G monoadduct, activated DMC-G monoadduct, G-MMC-G and G-DMC-G bisadducts, 5FC, and 5MC were first refined by means of the semiempirical QM program MOPAC2009 (40) using the AM1 Hamiltonian (41) and PRECISE stopping criteria, and further optimized using the restricted Hartree–Fock (RHF) method and a 6–31G(d) basis set, as implemented in the ab initio quantum chemistry program Gaussian03 (42). The modified nucleic bases incorporated a methyl group on either N^9 (G) or N^1 (C) in place of the deoxyribose ring. For each molecular system the calculated wave function was then used to derive electrostatic potential-derived (ESP) charges employing the RESP methodology (43). Point charges for deoxyribose and phosphate atoms in the modified nucleotides were restrained to the values these atoms have in the *parmbsc0* AMBER force field (44), which includes corrections for an improved description of DNA conformations on a multinanosecond time scale (45). Additional bonded and nonbonded parameters for drugs and modified bases are reported in the Supplementary Data.

Construction of the oligodeoxynucleotides

An initial model for the free 14mer of sequence d(CGATAACGCTAACG)·d(CGTTAGCGTTATCG) was built using the *nucgen* module in AMBER and optimized parameters for B-DNA (46). Terminal G:C base-pairs on both sides of the A,T-rich regions were used to avoid fraying of the double helix (47). Replacement of the cytosine base with either 5MC or 5FC in the central CpG step provided initial models for the modified oligonucleotides. The MMC-N^2(G) monoadduct found in NDB entry 199D (32) was used as a template for model building the precovalent intermediates by placing the activated MMC molecule in a similar location and avoiding the steric clash upon bond removal by steepest descent energy minimization.

Molecular dynamics simulations

Each molecular system was immersed in a truncated octahedron of TIP3P water molecules (48) and neutralized by addition of the appropriate number of Na^+ ions (49) at random locations. The standard MD simulations were run using the *pmemd* module in the AMBER 11 suite of programs. Periodic boundary conditions were applied and electrostatic interactions were treated using the smooth particle mesh Ewald method (50) with a grid spacing of 1 Å. The cutoff distance for the non–bonded interactions was 9 Å. The SHAKE algorithm (51) was applied to all bonds involving hydrogens and an integration step of 2.0 fs was used throughout. Solvent molecules and counterions were relaxed by energy minimization and allowed to redistribute around the positionally restrained solute (25 $kcal \cdot mol^{-1} \cdot Å^{-2}$) during 50 ps of MD at constant temperature (300 K) and pressure (1 atm), essentially as described previously (52). These initial harmonic restraints were gradually reduced in a series of progressive energy minimizations until they were completely removed. The resulting systems were heated from 100 to 300 K during 20 ps, equilibrated at 300 K for 1 ns in the absence of any restraints and further simulated under the same conditions up to a total time of 10–20 ns during which system coordinates were collected every 20 ps for further analysis.

Hybrid QM/MM calculations

The system was partitioned into two distinct parts: a QM region defined by the *iqmatoms* keyword that consisted of the atoms relevant for the specific bonding reaction being studied and their immediate surroundings, and an MM region with all the remaining atoms, including the solvent molecules and the counterions. The AM1 Hamiltonian (41) and full electrostatic interactions between the QM charge density (expanded in a STO-6G minimal basis set) and the standard RESP/6-31G(d) point charges of the MM atoms were used (38), as implemented in the *sander* module of AMBER 11. In the case of the precovalent complexes, the QM region included the reactive MMC and DMC derivatives plus the whole CpG step (to ascertain any possible differences involving the cytosines) whereas in the study of ICL formation, the QM region included the drug-G8 monoadduct plus the second guanine (G22) in the opposite strand. Reaction coordinates were specified for covalent bond formation, first between N^2 of G8 and C1 of MMC^+ or DMC^+ and then between N^2 of G22 and C10 in the monoadduct, using the steered MD procedure implemented in *sander* (53). In brief, the N–C distance between the two atoms to be bonded was shortened linearly from its current value in the precovalent complex to a target value of 1.35 Å over 5 ps using a harmonic restraint of 1000 kcal $(mol \cdot Å)^{-1}$ while all other variables were free to change. During the QM/MM part of the MD simulation SHAKE (keyword *qmshake*) was turned off and the integration step was reduced to 0.5 fs. Upon achievement of the final state, the AMBER topology file was updated with the new bond and atom type definitions, and a short energy minimization using the redefined connectivities and point charges allowed this geometry to be optimized in the MM force field prior to continuing the standard MD simulation. A scheme depicting the different MD and QM/MM phases of the whole study is given in the Supplementary Data.

Analysis of the molecular dynamics trajectories and electrostatic energy calculations

Three-dimensional structures and trajectories were visually inspected using the computer graphics program PyMOL (54). Interatomic distances and angles, as well as root–mean–square deviations (rmsd) from a given structure, were monitored using the *ptraj* module in AMBER. The DNA conformational and helical parameters were analyzed by means of program Curves+ (55). Molecular electrostatic potentials (MEP) around the low-dielectric solutes ($\varepsilon = 1$) were calculated by treating the surrounding solvent as a continuous high-dielectric medium ($\varepsilon = 78.5$) and solving the non-linear Poisson-Boltzmann (PB) equation by means of a finite difference method as implemented in program DelPhi version 4 (56) and described in more detail elsewhere (52,57). To highlight the most dissimilar MEP regions between any two DNA molecules, their central CGC triplets were best-fit superimposed and identical grids were calculated for both using the VSDMIP plugin (58). The energy value at each grid point for one molecule was then subtracted from the value calculated at the same grid point for the other molecule, as reported previously (59), and the difference energy map was visualized in PyMOL.

MM–GBSA binding energy calculations

The binding free energy ($\Delta G_{binding}$) between a reactive intermediate of MMC (MMC^+) and a given DNA

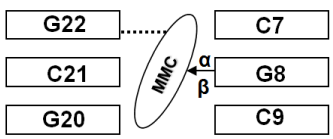
oligonucleotide (standard, 5MC-DNA or 5FC-DNA) was calculated as:

$$\Delta G_{binding} = \left\langle G^{DNA-MMC^+}(i) - G^{DNA}(i) - G^{MMC^+}(i) \right\rangle_i$$

where $\langle \ \rangle_i$ denotes an ensemble average over i snapshots taken from the MD trajectories of the precovalent complexes. For comparison purposes we took 150 snapshots from 3.0 ns of each simulation during which MMC^+ was similarly oriented in the minor groove of the three complexes studied. The G_i values for each species were calculated using the following energy decomposition scheme, which is implemented in the MM–GBSA method within AMBER 11 (script mm_pbsa.pl) (60):

$$G = E_{gas} + G_{sol} - TS$$
$$E_{gas} = E_{bond} + E_{angle} + E_{torsion} + E_{ele} + E_{vdw}$$
$$G_{sol} = G_{GB} + G_{nonpolar}$$
$$G_{nonpolar} = \gamma\ SASA$$

E_{gas} is the gas phase energy, calculated using the AMBER force field, which encompasses internal energies from bond lengths (E_{bond}), valence angles (E_{angle}) and torsional angles ($E_{torsion}$), as well as coulombic (E_{ele}) and van der Waals (E_{vdw}) non-bonded contributions calculated with no cutoff. In the case of individual trajectories, as in the present case, the contribution of the internal energies to $\Delta G_{binding}$ amounts to zero. The solvation free energy, G_{sol}, was decomposed into polar and nonpolar contributions; the former (G_{GB}) was calculated by solving the generalized Born (GB) equation (61) using dielectric constants of 1 and 78.5 for solute and solvent, respectively, whereas the latter, $G_{nonpolar}$, which is due to due to cavity formation and van der Waals interactions between the solute and the solvent, was estimated from the nonpolar solvent accessible surface area (SASA) (62) using 1.4 Å as the water probe radius and a value of 0.005 kcal $(mol \cdot Å^2)^{-1}$ for the surface tension constant γ. T and S are the temperature and the total solute entropy but this term, which is cumbersome to compute and most likely similar for the three DNA-MMC complexes under study, was obviated in our approximation to the binding free energy differences (63).



Scheme II. Simplified diagram representing the alignment of MMC^+ in the minor groove of the DNA duplex central region showing the numbering of the base pairs involved in adduct formation.

RESULTS AND DISCUSSION

Drug conformation and orientation in the precovalent complexes

The MD simulation of the initial complex containing the standard oligodeoxynucleotide and an activated MMC molecule (MMC^+) in the minor groove of the central region corresponding to the C7:G22/G8:C21 base-pair step (Scheme II) revealed that the hydrogen bond between the carbamate OM atom and the exocyclic

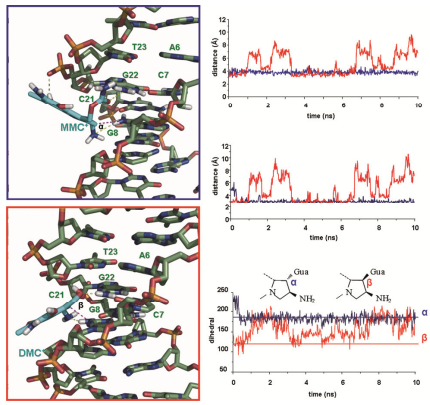


Figure 1. *Left*: Details of representative MD snapshots from the MD simulations showing the distinct orientations of MMC^+ (top) and DMC^+ (bottom) in the DNA minor groove prior to nucleophilic attack by the 2-amino group of G8. Drug and DNA carbon atoms are coloured in cyan and green, respectively. *Right*, from top to bottom: Time evolution of N^2(G8)–C1(drug) and N^2(G8)–N2(drug) distances for MMC^+ (blue) and DMC^+ (red), and the angle for attack.

amino group of G22 was not particularly stable in the face of competition with water molecules. This hydrogen bond was reported for the monoadduct structure solved by NMR spectroscopy (32) and presumed to dominate sequence recognition in the precovalent complex as well. However, in addition to the weak character and short life of this interaction (Figure 1), we also noted that this orientation of the drug in the minor groove was incompatible with a nucleophilic attack by N^2(G8) on the α face of the MMC tetrahydropyrrole ring.

This unexpected result led us to reconsider the conformation of MMC^+ inside the minor groove and explore alternative hydrogen bonding modes. In our lab, previous conformational studies on MMC^+ in the gas phase (data not shown) had revealed a preferred geometry that involved formation of an internal hydrogen bond between the phenol group at C8 and oxygen OM, and this conformer was also detected in aqueous solution when the dynamic behaviour of MMC^+ was simulated (Supplementary Figure S1). Besides, additional QM work on a CpG dinucleotide model system had shown that, using this conformation, a reaction coordinate defined to create the covalent bond leading to the α monoadduct could be accomplished. Interestingly, when this particular MMC^+ conformer was docked into the oligo's minor groove region comprising C7:G22/G8:C21, we realized that a good hydrogen bond with the amino group of G22 could actually be established by the carbonyl oxygen (OC) instead of OM, as in the previous case. Therefore OC, rather than facing the solvent as in the former simulation, was directed towards the strand opposite to that being alkylated (Figure 1). The MD trajectory of this new complex allowed us to see that this binding pose remained stable and not only provided a distance between C1 of MMC^+ and N^2(G8) that was suitable for nucleophilic attack, i.e. ~4 Å (Figure 1), but also allowed an approach on the ring face opposite to that of the primary amino group at C2 (Scheme I). This particular orientation was additionally stabilized by two hydrogen bonds: a strong one between the phenolic O5 and phosphate OP1 oxygens, and another between the sp^2 N^2(G8) and the sp^3 nitrogen of the primary amino group at position 2 of the activated drug. This hydrogen-bonding anchoring, which

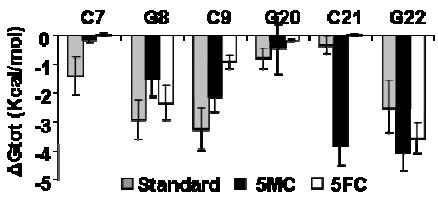


Figure 2. Calculated binding energies between MMC^+ and individual DNA bases in the precovalent complexes formed with standard DNA, 5MC-DNA and 5FC-DNA.

is maintained for virtually the whole simulation (Supplementary Figure S2), contributes to placing the mitosene ring system in an asymmetric position within the minor groove so that only one face is in van der Waals contact with the the sugar-phosphate backbone of the unmodified strand whereas the other one remains exposed to the aqueous milieu.

In DMC, on the other hand, a hydroxymethyl substituent replaces the methyl carbamate moiety of MMC. Consequently, only one oxygen (OM) is present in the drug's side chain available for hydrogen-bonding to the DNA. When DMC^+ was docked in the minor groove region of the C7:G22/G8:C21 base-pair step similarly to MMC^+ and the resulting complex was simulated using MD, we noticed that OM could indeed accept a hydrogen bond from the amino group of G22 (Figure 1) but the half-life of this interaction was shorter because of exchange with solvent molecules in the nanosecond timescale (data not shown). Nonetheless, this hydrogen bond largely contributed to placing DMC^+ in an orientation such that attack by N^2(G8) would occur preferentially, although not exclusively, on the β face of the drug's tetrahydropyrrole ring (Figure 1). In this complex, the activated drug is roughly equidistant from both DNA strands and no direct hydrogen bond with any phosphate group is detected.

Taken together, these results strongly suggest that the very distinct relative occurrence of α and β DNA monoadducts formed in the presence of MMC and DMC (36,36) is largely dictated by the nature of the hydrogen bond involving either the carbamate or the hydroxyl oxygen in the activated drugs and the amino group of the guanine in the unmodified strand that is complementary to the cytosine in the CpG step undergoing G-alkylation. The fact that DMC and MMC show opposite stereochemical preferences supports the roles that we assign to the carbonyl oxygen, OC, which is only present in the former compound, and to OM, which is present in both MMC (as part of the carbamate) and DMC (as a hydroxyl). Altogether, the drug-DNA hydrogen bonds involving these oxygens crucially determine the proposed near-attack conformations and binding modes for MMC and DMC, which are distinct and can be distinguished by monitoring an improper dihedral angle for the incoming 2-amino group relative to the mitosene tetrahydropyrrole ring (Figure 1).

Precovalent complexes between MMC^+ and DNA containing modified cytosines

Once the putative binding mode for MMC^+ in the minor groove of the central C7:G22/G8:C21 region was identified, we carried out similar MD simulations with the oligodeoxynucleotides in which C7 and C21 had been replaced with either 5MC or 5FC. When we calculated the binding energy ($\Delta G_{binding}$) between MMC^+ and the central CGC triplet using the MM-GBSA method, we found that, relative to the standard oligo-

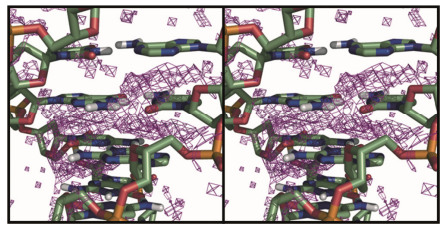


Figure 3. Close-up view of the CpG region where MMC binds and display of the difference in molecular electrostatic potential (MEP) between 5MC-DNA and standard DNA (contoured at -5 kcal/mol) . Note the more negative character of the MEP around the O^2 of both methylated cytosines.

nucleotide, it was more favourable for the duplex containing 5MC whereas it was significantly decreased in the duplex containing 5FC. Furthermore, an energy decomposition analysis (Figure 2) showed that, in both cases, the base providing the largest differences is precisely C21, i.e. the cytosine complementary to the guanine undergoing alkylation, which displays an improved interaction with the drug when it is methylated. Our theoretical result that C21, rather than C9, is largely responsible for these differences in $\Delta G_{binding}$ is in very good agreement with experimental evidence showing the prevalence of the 5MC that is paired to the alkylated guanine over that placed 5' to it in the same strand (64).

Since the major energy component appeared to be electrostatic we decided to explore the MEP generated in the minor groove of this central region in the three oligonucleotides studied. To facilitate comparisons we also calculated and displayed the MEP differences relative to the duplex containing standard cytosines (Figure 3). It can be seen that significant differences do arise in the minor groove (and also between the stacked bases), despite the fact that the structural changes are located in the major groove. This type of observation has been used earlier to account for differences in the binding preferences of several bisintercalators (e.g. references 52,59) and can be used here to provide a rationale to the experimental finding that cytosine methylation enhances the rate of DNA alkylation by MMC whereas the opposite is true if C is replaced with 5FC (27,28). Our interpretation is that the more negative MEP found in the drug-binding region when the cytosines in the CpG step are methylated facilitates and strengthens the electrostatic interaction with the positively charged alkylating species.

DNA conformational parameters

To complete the study of the precovalent complexes, we examined their geometrical properties and compared them to those of free DNA, 5MC-DNA and 5FC-DNA. The very similar DNA conformations were well maintained during the course of the MD simulations (data not shown) and only subtle differences involving the critical G8pC9 step were observed. In particular, analysis of the stacking geometry between neighboring G8:C21 and C9:G20 base-pairs revealed a significant displacement towards negative tilt values relative to free DNA (Figure 4) when MMC^+ is located in the DNA minor groove in a suitable position for nucleophilic attack by N^2(G8). This means that the angle formed between these base-pairs needs to be opened towards the unmodified DNA strand in order for the reaction to proceed. Remarkably, this slightly distorted geometry is

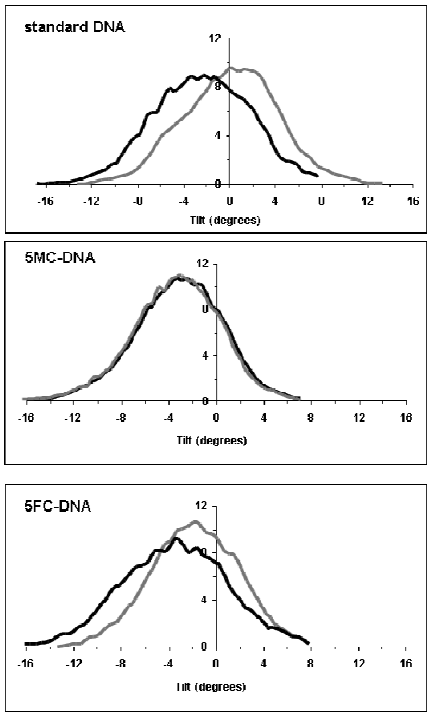


Figure 4. Distribution of tilt values at the G8pC9 step of the central CGC triplet over the MD simulations of free (grey) standard DNA (top), 5MC-DNA (middle) and 5FC-DNA (bottom) and their respective precovalent complexes with MMC^+ (black). No significant differences were observed at the C7pG8 step.

naturally present, in the absence of bound drug, when C9 and C21 are methylated and also, to a lesser extent, when C is replaced with 5FC. It is therefore likely that this structural preorganization for MMC^+ binding in the DNA containing the modified cytosines acts in concert with the more negative MEP in the minor groove to enhance alkylation by MMC when the 5 substituent is an electron-donating methyl group. On the other hand, in the case of the DNA containing 5FC in the central region, it seems that the less favourable electrostatic interaction with MMC^+ prevails over the more preorganized geometry to decrease drug binding to this site.

Formation and description of the MMC and DMC monoadducts

We used different snapshots from the MD trajectories of the MMC^+–DNA and DMC^+–DNA precovalent complexes reported above to start simulating the QM reaction coordinate that leads to covalent bond formation between N^2 of the attacking guanine (G8) and C1 of the drugs by systematically reducing this distance using the hybrid QM/MM method. Interestingly, a consistent pattern was found that revealed that the opposite α or β stereochemistry at C1 of the resulting adduct (Schemes I and II) was entirely dependent on the conformation-driven positioning of the activated drug in the DNA minor groove. Thus, in the simulation involving MMC^+, hydrogen-bonding of the drug's OC to N^2(G22) determined formation of the C1α-monoadduct whereas

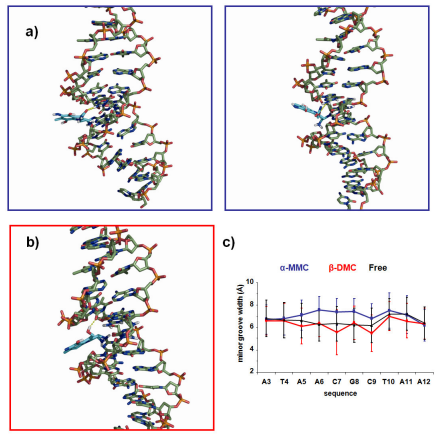


Figure 5. a) Representative MMC-DNA α-monoadducts showing the two most populated orientations of the drug in the minor groove depending on whether the carbonyl oxygen of the drug's carbamate group is hydrogen bonded to N^2(G22) (left) or not (right). b) Representative DMC-DNA β-monoadduct showing the hydrogen bond between the drug's hydroxyl and O^2(T23). c) Widths of the minor groove in free DNA and in the two types of adducts.

the C1β-stereoisomer was systematically obtained when this hydrogen bond was lost. An identical result was obtained when the same procedure was used on the precovalent (5MC)DNA–MMC^+ and (5FC)DNA–MMC^+ complexes. The free energy profiles for this step of the reaction, on the other hand, showed no significant differences among the three complexes (Supplementary Figure 2A).

In the resulting MMC-DNA covalent intermediate N^2(G8) still bears two hydrogens, one of which (H^{21}) can still be used to pair with the complementary O^2(C21) following relaxation of the complex. But the extra H^{22} must be transferred to the amino group of the bonded drug (Scheme I) and, given the geometry of the resulting adduct, this proton shuttle could only be effected with the concourse of the nucleobase 3' to the alkylated guanine, in our case O^2(C9). Nonetheless, simulation of these reactions using the QM/MM method turned out to be unfeasible. We then realized that in the precovalent complexes reported above for both MMC^+ and DMC^+ the sp^2–hybridized N^2(G8) was close not only to C1 of the drugs but even more so to the sp^3–hybridized amino group at C2, with which it could actually establish a good hydrogen bond. In fact, both sets of interatomic distances were found to be strictly correlated during the MD simulations (Figure 1), which suggests that transfer of H^{22} to N2(MMC^+) or N2(DMC^+) could be concomitant with the formation of the C1–N^2(G8) bond, but this alternative reaction pathway was not explored further.

The structures of the resulting MMC-DNA and DMC-DNA complexes containing, respectively, the α- and β-monoadducts were relaxed upon relocation of the extra H^{22} on N^2(G8) to the amino group of the bonded drug and equilibrated for 10 ns of unrestrained MD simulations. The MMC-DNA complex faithfully reproduced the major features of the models originally refined on the basis of NOE intensities and deposited in the NDB with code 199D (32), despite differences in base composition and sequence on both sides of the central CpG step and the fact that the NMR solution structure considered a benzoquinone ring in MMC rather than the hydroquinone present in our model and

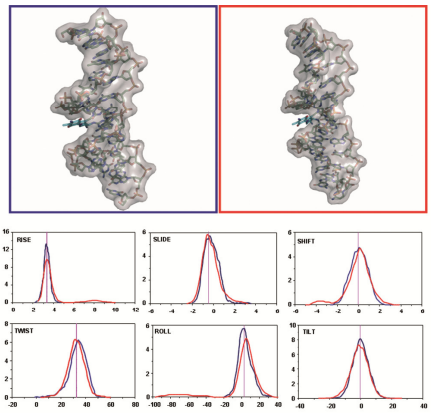


Figure 6. *Top*: representative snapshots of the duplexes containing the C1α (left, blue frame) and C1β (right, red frame) bis-adducts that crosslink the two DNA strands. The DNA atoms are surrounded by a semi-transparent van der Waals surface. *Bottom*: Distributions of base-pair helical parameters (excluding both ends) in the duplexes containing the C1α (blue) and C1β(red) ICL. Twist, tilt and roll are in degrees. Shift, slide and rise are in Å. The vertical line represents the average for the simulated drug-free DNA.

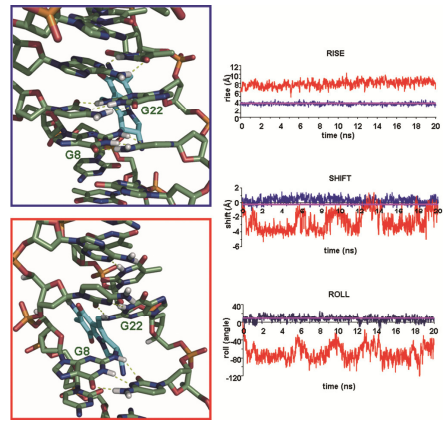


Figure 7. *Left*: detail of the central region of the duplexes containing the C1α (top, blue frame) and C1β (bottom, red frame) bis-adducts that crosslink the two DNA strands, as seen from the major groove. Note the loss of intrastrand stacking interactions at the central CpG step in the C1β ICL. *Right*: Time evolution of rise (top), shift (middle) and roll (bottom) parameters at the C7pG8 step during the MD simulations of the bis-adducts. The pink horizontal line in each plot represents the average value for drug-free DNA.

necessary for evolution towards a bisadduct. Most noteworthy are (i) the absence of significant deviations in helical parameters with respect to free DNA (data not shown), (ii) a slight widening of the minor groove (Figure 5) where the covalently bonded drug remains in an off-centre location despite the loss of the O5-OP1 hydrogen bond, and (iii) alternative orientations for the carbamate side chain of MMC.

In the case of the DMC β-adduct, the drug is lodged in a slightly compressed minor groove (Figure 5), where it interacts simultaneously with the sugar-phosphate backbone of both strands, and the hydroxyl group at C10 keeps its hydrogen-bonding interaction with O^2(T23).

From the above, it must be noted that a DMC-DNA complex resulting from water-mediated decarbamoylation of DNA-bonded MMC (36), which appears to be more abundant in cells than the original adduct (37), is stereochemically identical (C1α) to only one of the two monoadducts that readily originate from direct attack of DMC to DNA (Scheme I), although in this case the C1β is obtained in a significantly larger proportion (37).

Interstrand crosslinking by DNA-bonded MMC and DMC at the CpG step

Once the structures of the duplexes containing the α- and β-monoadducts were equilibrated, we proceeded to remove the carbamate (in the case of MMC) and the hydroxyl (in the case of DMC) so as to generate the second alkylating species. After 10 ns of equilibration of these new complexes using standard MD, we could observe that the separation between the exocyclic N^2(G22) and C10 of MMC or DMC fluctuated less in the case of DMC and remained close to 4 Å (Supplementary Figure 3), a suitable distance for nucleophilic attack and similar to that reported above for monoadduct formation. By switching to the QM/MM method we shortened the N^2–C10 distance so as to create the second covalent bond and give rise to the corresponding bis-adducts, which were equilibrated thereafter for 20 ns of unrestrained MD simulation.

Our results show that the DNA is able to accommodate this dual covalent modification (on G8 in one strand and G22 in the opposite strand) undergoing minimal distortion only in the case of the C1α-ICL, in agreement with early work that employed MM exclusively (4). In this bis-adduct, the crosslinking mitosene unit is settled in the minor groove of the CpG step in close contact with the strand containing the second alkylated guanine; the global structural parameters of the DNA molecule hardly change with respect to the drug-free duplex (Figure 6). On the contrary, in the C1β-ICL, whose structure has not been reported yet as far as we know, the mitosene unit is more equidistant to both DNA strands but the rise, shift and roll parameters at the C7pG8 step significantly deviate from the values in free DNA and this causes a marked change in the distribution of these values (Figure 6). A more detailed view from the major groove is shown in Figure 7 where the loss of stacking interactions at this step can be visually appreciated and also deduced from the increased rise values that are monitored along the MD trajectory. The negative shift is a consequence of the displacement of the drug-bonded guanines into the minor groove whereas the correlated negative roll indicates bending of the helix toward the minor groove, a situation contrary to what is observed in DNA duplexes containing an adduct with the monofunctional antitumour tetrahydroisoquinolines trabectedin (Yondelis®) or PM01183, which have been shown to bend the helix toward the major groove by introducing positive roll (65,66) and can functionally mimic an ICL (67).

CONCLUSIONS

The oligodeoxynucleotide used in the present study contains a highly preferred CGC triplet site in the middle region for MMC and DMC binding. Furthermore the initial monoadducts formed with G8 can evolve towards an ICL by forming a second covalent bond with G22 in the opposite strand. By simulating each step of the reaction using a combination of MD and QM/MM calculations, we have been able to obtain information in atomic detail about the recognition events previous to monoadduct formation as well as about the structural

and electronic requirements that possibly account for the increased reactivity of MMC towards CpG steps containing 5MC in place of cytosine. Moreover, we provide a possible explanation to the origin of the formation of either C1α or C1β monoadducts depending on the drug-DNA hydrogen bonds that are established in the precovalent DNA-MMC$^+$ and DNA-DMC$^+$ complexes. According to our findings, formation of a hydrogen bond between the carbamate oxygen of activated MMC and the guanine in the strand opposite to that containing the G that undergoes the first alkylation reaction appears to favour formation of the C1α-monoadduct. A change in the positioning of its decarbamoylated analogue DMC due to a distinct hydrogen-bonding pattern leads to a preferred attack on the opposite face of the pyrrolidine ring and creation of the C1β-monoadduct.

The DNA helical structure at the target CpG step appears to be more preorganized for binding the positively charged activated MMC, and also to possess a more negative MEP in the minor groove, when the cytosines are methylated. Both factors acting in concert may well explain the enhanced reactivity of MMC towards this 5MC-DNA that has been detected experimentally.

Our unrestrained MD simulations in aqueous solution faithfully reproduce the drug-DNA interactions previously described for the NMR solution structure of the major MMC-DNA C1α-monoadduct and provide further insight into (i) the earlier precovalent complex, (ii) the reaction of the activated monoadduct with a second guanine in the opposite strand at a CpG step, and (iii) the final ICL, which is embedded in a minimally distorted DNA helix that does not show any appreciable curvature, in good agreement with the experimental evidence (33). We then studied in a similar way the formation of the major DMC-DNA monoadduct with opposite stereochemistry at C1 and its stepwise evolution to an ICL. The three-dimensional structures of these C1β-adducts have not been reported before, to the best of our knowledge, and they were found to display some distinct features. Most notably, the drug-bonded CpG step is severely distorted in the β-ICL and the helix is bent towards the minor groove.

The differences we have detected in the conformations of these adducts, which appear in different proportions in the DNA of cells that have been exposed to either MMC or DMC, are likely to pose dissimilar challenges to the DNA repair machineries. This may account for the experimental findings that FA cells are hypersensitive to MMC but have normal sensitivity to DMC whereas in other cell types DMC is more cytotoxic than MMC (37).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

This work was partially supported by Comisión Interministerial de Ciencia y Tecnología (SAF2006-12713-C02-02 and SAF2009-13914-C02-02) and Comunidad de Madrid (S-BIO/0214/2006). J.A.B.-C. enjoys a research fellowship from PharmaMar S.A.U. A.M. acknowledges Comunidad Autónoma de Madrid for financial support through the AMAROUTO program to the Fundación Severo Ochoa. Funding for open access charge: Comisión Interministerial de Ciencia y Tecnología (SAF2009-13914-C02-02).

ACKNOWLEDGEMENTS

We thankfully acknowledge the generous allowance of computer resources, technical expertise and assistance provided by the Red Española de Supercomputación at the Barcelona Supercomputing Center (MareNostrum). We are grateful to Prof. María José Camarasa (IQM, CSIC), Dr. Sonsoles Velázquez (IQM, CSIC), Dr. Carolina Burgos (UAH) and Prof. Antonio Jiménez-Ruiz (UAH) for many fruitful discussions.

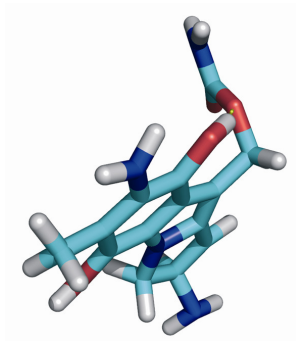
References

1. Hata, T., Hoshi, T., Kanamori, K., Matsumae, A., Sano, Y., Shima, T. and Sugawara, R. (1956) Mitomycin, a new antibiotic from *Streptomyces*. I. *J. Antibiot. (Tokyo)*, **9**, 141–146.
2. Andrez,J.C. (2009) Mitomycins syntheses: a recent update. *Beilstein J. Org. Chem*., 5, 33.
3. Iyer VN, Szybalski W. (1963). A molecular mechanism of mitomycin action: linking of complementary DNA strands. *Proc Natl Acad Sci U S A,* **50**, 355–362.
4. Tomasz,M., Lipman,R., Chowdary,D., Pawlak,J., Verdine,G.L. and Nakanishi,K. (1987) Isolation and structure of a covalent cross-link adduct between mitomycin C and DNA. *Science*, **235**, 1204–1208.
5. Kumar,S., Lipman,R. and Tomasz,M. (1992) Recognition of specific DNA sequences by mitomycin C for alkylation. *Biochemistry*, **31**, 1399–1407.
6. Tomasz,M., Chowdary,D., Lipman,R., Shimotakahara,S., Veiro,D., Walker,V. and Verdine,G.L. (1986) Reaction of DNA with chemically or enzymatically activated mitomycin C: isolation and structure of the major covalent adduct. *Proc. Natl. Acad. Sci. USA,* **83**, 6702–6706.
7. Tomasz,M. (1995) Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.,* **2**, 575–579.
8. Shelley,M.D., Mason,M.D. and Kynaston,H. (2010) Intravesical therapy for superficial bladder cancer: a systematic review of randomised trials and meta-analyses. *Cancer Treat Rev.,* **36**(3):195–205.
9. Abraham,L.M., Selva,D., Casson,R. and Leibovitch,I. (2006) Mitomycin: clinical applications in ophthalmic practice. *Drugs*, **66**, 321–340.
10. Hofheinz,R.D., Beyer,U., Al-Batran,S.E. and Hartmann,J.T. (2008) Mitomycin C in the treatment of gastrointestinal tumours: recent data and perspectives. *Onkologie,* **31**, 271–281.
11. McPherson,J.P., JP, Lemmers,B., Chahwan,R., Pamidi,A., Migon,E., Matysiak-Zablocki,E., Moynahan,M.E., Essers,J., Hanada,K., Poonepalli,A. et al. (2004). Involvement of mammalian Mus81 in genome integrity and tumor suppression. *Science,* **304**, 1822–1826.
12. Lee,Y.J., Park,S.J., Ciccone,S.L., Kim,C.R. and Lee,S.H. (2006) An *in vivo* analysis of MMC-induced DNA damage and its repair. *Carcinogenesis,* **27**, 3, 446–453.
13. Henson,S.E., Tsai,S.C., Malone,C.S., Soghomonian,S.V., Ouyang,Y., Wall,R., Marahrens,Y. and Teitell,M.A. (2006) Pir51, a Rad51-interacting protein with high expression in aggressive lymphoma, controls mitomycin C sensitivity and prevents chromosomal breaks. *Mut. Res.,* **601**, 113–124.
14. Moynahan,M.E., Cui,T.Y. and Jasin,M. (2001) Homology-directed DNA repair, mitomycin-C resistance, and chromosome stability is restored with correction of a Brca1 mutation. *Cancer Res*., **61**, 4842–4850.
15. Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B, and Yu,X. (2009) PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Curr. Biol*., **19**, 524–529.
16. Villarroel,M.C., Rajeshkumar,N.V., Garrido-Laguna,I., De Jesús-Acosta,A., Jones,S., Maitra,A., Hruban,R.H., Eshleman,J.R., Klein,A., Laheru,D., Donehower,R., Hidalgo,M. (2011) Personalizing cancer treatment in the age of global genomic analyses: *PALB2* gene mutations and the response to DNA damaging agents in pancreatic cancer. *Mol. Cancer Ther*., **10**, 3–8.

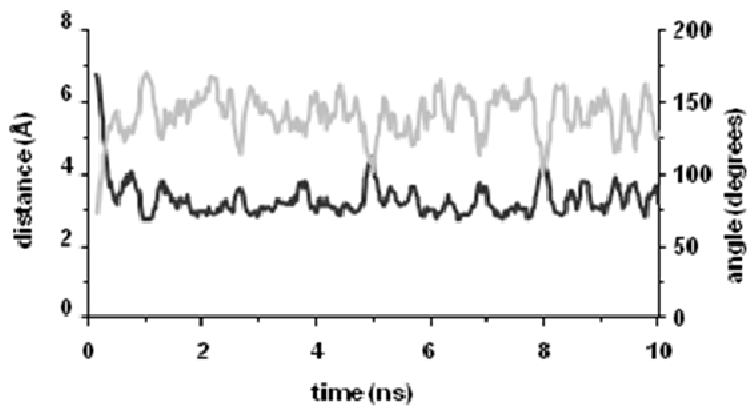
17. Strathdee, C. A., Duncan, A. M. and Buchwald, M. (1992) Evidence for at least four Fanconi anaemia genes including FACC on chromosome 9. *Nature Genet.,* **1**, 196–198.
18. Cervenka,J., Arthur,D. and Yasis,C. (1981) Mitomycin C test for diagnostic differentiation of idiopathic aplastic anemia and Fanconi anemia, *Pediatrics,* **67**, 119–127.
19. Mirchandani,K.D. and D'Andrea,A.D. (2006) The Fanconi anemia/BRCA pathway: a coordinator of cross-link repair. *Exp Cell Res*., **312**, 2647–2653.
20. Wolkenberg,S.E. and Boger,D.L. (2002) Mechanisms of in situ activation for DNA-targeting antitumor agents. *Chem. Rev.,* **102**, 2477–2495.
21 Tomasz,M. and Palom,Y. (1997) The mitomycin bioreductive antitumor agents: cross-linking and alkylation of DNA as the molecular basis of their activity. *Pharmacol Ther*., **76**, 73–87.
22. Noll,D.M., Mason,T.M. and Miller,P.S. (2006) Formation and repair of interstrand cross-links in DNA. *Pharmacol Ther.,* **106**, 277–301.
23. Bizanek,R., McGuinness,B.F., Nakanishi,K. and Tomasz,M. (1992) Isolation and structure of an intrastrand cross-link adduct of mitomycin C and DNA. *Biochemistry*, **31**, 3084–3091.
24. Palom,Y., Belcourt,M.F., Musser,S.M., Sartorelli,A.C., Rockwell,S. and Tomasz,M. (2000) Structure of adduct X, the last unknown of the six major DNA adducts of mitomycin C formed in EMT6 mouse mammary tumor cells. *Chem. Res. Toxicol*. **13**, 479–488.
25. Borowy-Borowski,H., Lipman,R. and Tomasz,M. (1990) Recognition between mitomycin C and specific DNA sequences for cross-link formation. *Biochemistry,* **29**, 2999–3006.
26. Li,V.-S., Choi,D.,Wang,Z., Jimenez,L.S., Tang,M. and Kohn,H. (1996) Role of the C-10 substituent in mitomycin C-1–DNA bonding. *J. Am. Chem. Soc*., **118**, 2326–2331.
27. Millard,J.T. and Beachy,T.M. (1993) Cytosine methylation enhances mitomycin C cross-linking. *Biochemistry*, **32**, 12850–12856.
28. Das,A., Tang,K.S., Gopalakrishnan,S., Waring,M.J. and Tomasz,M. (1999) Reactivity of guanine at m^5CpG steps in DNA: evidence for electronic effects transmitted through the base pairs. *Chem. Biol.,* **6**, 461–471.
29. Dannenberg,J.J. and Tomasz,M. (2000) Hydrogen-bond acid/base catalysis: a density functional theory study of protonated guanine-(substituted) cytosine base pairs as models for nucleophilic attack on mitomycin in DNA. *J. Am. Chem. Soc.,* **122**, 2062–2068.
30. Rao,S.N., Singh,U.C. and Kollman,P.A. (1986) Conformations of the noncovalent and covalent complexes between mitomycins A and C and d(GCGCGCGCGC). *J. Am. Chem. Soc*., **108**, 2058–2068.
31. Li,V.-S. and Kohn,H. (1991) Studies on the bonding specificity for mitomycin C-DNA monoalkylation processes.). *J. Am. Chem. Soc*., **113**, 275–283.
32. Sastry,M., Fiala,R., Lipman,R., Tomasz,M. and Patel,D.J. (1995) Solution structure of the monoalkylated mitomycin C-DNA complex. *J. Mol. Biol*., **247**, 338–359.
33. Norman,D., Live,D., Sastry,M., Lipman,R., Hingerty,B.E., Tomasz,M., Broyde,S. and Patel,D.J. (1990) NMR and computational characterization of mitomycin cross-linked to adjacent deoxyguanosines in the minor groove of the d(T-A-C-G-T-A)· d(T-A-C-G-T-A) duplex. *Biochemistry*, **29**, 2861–2875.
34. Berman,H.M., Olson,W.K., Beveridge,D.L., Westbrook,J., Gelbin,A., Demeny,T., Hsieh,S.-H., Srinivasan,A.R. and Schneider,B. (1992) The Nucleic Acid Database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J*., **63**, 751–759.
35. Kim,S.Y. and Rockwell,S. (1995) Cytotoxic potential of monoalkylation products between mitomycins and DNA: studies of decarbamoyl mitomycin C in wild-type and repair-deficient cell lines. *Oncol Res*. 7, 39–47.
36. Palom,Y., Suresh Kumar,G., Tang,L.Q., Paz, M.M., Musser,S.M., Rockwell,S. and Tomasz,M. (2002) Relative toxicities of DNA crosslinks and monoadducts: New insights from studies of decarbamoyl mitomycin C and mitomycin C. *Chem. Res. Toxicol*. **15**, 1398–1406.
37. Paz,M.M., Ladwa,S., Champeil,E., Liu,Y., Rockwell,S., Boamah,E.K., Bargonetti,J., Callahan,J., Roach,J. and Tomasz,M. (2008) Mapping DNA adducts of mitomycin C and decarbamoyl mitomycin C in cell lines using liquid chromatography/ electrospray tandem mass spectrometry. *Chem Res Toxicol*. 21, 2370–2378.
38. Walker,R.C., Crowley,M.F. and Case,D.A. (2008) The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comp. Chem*. **29**, 1019–1031.
39. Arora, S.K. (1979) Structural investigations of mode of action of drugs. I. Molecular structure of mitomycin C. *Life Sci*., **24**, 1519–1526.
40. Steward, J.J.P. (2009) MOPAC2009. Steward Computational Chemistry. URL: http://openmopac.net/MOPAC2009.html.
41. Dewar,M.J.S., Zoebisch,E.G., Healy,E.F. and Stewart,J.J.P. (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*., **107**, 3902–3909.
42. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Zakrzewski,V.G., Montgomery,J.A.,Jr., Stratmann,R.E. and Burant,J.C. et al. (2003) *Gaussian 03*, revision B. 04, Gaussian, Inc.: Pittsburgh, PA.
43. Bayly,C.I., Cieplak,P., Cornell,W.D. and Kollman,P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: the RESP model, *J. Phys. Chem*., **97**, 10269–10280.
44. 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. and Kollman,P.A. (1995) A second-generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
45. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E.,3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J*., **92**, 3817–3829.
46. Arnott,S. and Hukins,D.W. (1972) Optimised parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Comm*., **47**, 1504–1509.
47. Bueren-Calabuig,J.A., Giraudon,C., Galmarini,C.M., Egly,J.M. and Gago,F. (2011) Temperature-induced melting of double-stranded DNA in the absence and presence of covalently bonded antitumour drugs: insight from molecular dynamics simulations. *Nucleic Acids Res*. doi: 10.1093/nar/gkr512
48. Jorgensen,W.L.,Chandrasekhar,J. and Madura,J.D. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*., **79**, 926–935.
49. Åqvist,J. (1990) Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem*., **94**, 8021–8024.
50. Darden,T.A., York,D. and Pedersen,L.G. (1993) Particle mesh Ewald: an N·log(N) method for computing Ewald sums. *J. Chem. Phys*., **98**, 10089–10092.
51. Ryckaert,J.P., Ciccoti,G. and Berendsen,H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys*., **23**, 327–341.
52. Marco,E., Negri,A., Luque,F.J. and Gago,F. (2005) Role of stacking interactions in the binding sequence preferences of DNA bis-intercalators: insight from thermodynamic integration free energy simulations. *Nucleic Acids Res*. **33**, 6214–6224.

53. Babin,V., Roland,C. and Sagui,C. (2008) Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys*., **128**, 134101.
54. Lano, W.D. (2006) PyMOL version 0.99. DeLano Scientific LLC, URL: http://www.pymol.org/.
55. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res*., **37**, 5917–5929.
56. Gilson,M.K., Sharp,K. and Honig,B. (1987) Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem*. 9, 327-335.
57. Marco,E., Laine,W., Tardy,C., Lansiaux,A., Iwao,M., Ishibashi,F., Bailly,C. and Gago,F. (2005) Molecular determinants of topoisomerase I poisoning by lamellarins: comparison with camptothecin and structure-activity relationships. *J. Med. Chem*., **48**, 3796–3807.
58. Cabrera,A.C., Gil-Redondo,R., Perona,A., Gago,F. and Morreale,A. (2011) VSDMIP 1.5: an automated structure- and ligand-based virtual screening platform with a PyMOL graphical user interface. *J. Comput. Aided Mol. Des*. Aug 9. [DOI: 10.1007/s10822-011-9465-6].
59. Bailly,C., Echepare,S., Gago,F. and Waring,M.J. (1999) Recognition elements that determine affinity and sequence-specific binding to DNA of 2QN, a biosynthetic bis-quinoline analogue of echinomycin. *Anticancer Drug Des*. **14**, 291–303.
60. Kollman,P.A., Massova,I., Reyes,C., Kuhn,B., Huo,S., Chong,L., Lee,M., Lee,T., Duan,Y., Wang,W. et al. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res*., **33**, 889–897.
61. Onufriev,A., Bashford,D. and Case,D.A. (2000) A modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem.*, **104**, 3712–3720.
62. Sitkoff,D., Sharp,K. and Honig,B. (1994) Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**:1978–1988.
63. Negri,A., Rodríguez-Larrea,D., Marco,E., Jiménez-Ruiz,A., Sánchez-Ruiz,J.M. and Gago,F. (2010) Protein-protein interactions at an enzyme-substrate interface: characterization of transient reaction intermediates throughout a full catalytic cycle of *Escherichia coli* thioredoxin reductase. *Proteins*, **78**, 36–51.
64. Johnson,W.S., He,Q.Y. and Tomasz,M. (1995) Selective recognition of the m5CpG dinucleotide sequence in DNA by mitomycin C for alkylation and cross-linking. *Bioorg. Med. Chem.* **3**, 851–860.
65. García-Nieto,R., Manzanares,I., Cuevas,C. and Gago,F. (2000) Bending of DNA upon binding of ecteinascidin 743 and phthalascidin 650 studied by unrestrained molecular dynamics simulations. *J. Am. Chem. Soc.* **122:**7172–7182.
66. Leal,J., Martínez-Díez,M., García-Hernández,V., Moneo,V., Domingo,A., Bueren-Calabuig,J., Negri, A., Gago,F., Guillén-Navarro,M., Avilés,P., Cuevas,C., García-Fernández,L. and Galmarini, C. (2010) PM01183, a new DNA minor groove covalent binder with potent in vitro and in vivo anti-tumour activity. *Brit. J. Pharmacol*. **161**:1099–1110.
67. Feuerhahn,S., Giraudon,C., Martínez-Díez,M., Bueren-Calabuig,J.A., Galmarini,C.M., Gago,F. and Egly,J.M. (2011) XPF-dependent DNA breaks and RNA polymerase II arrest induced by antitumor DNA interstrand crosslinking-mimetic alkaloids. *Chem Biol.* **18**:988–999.

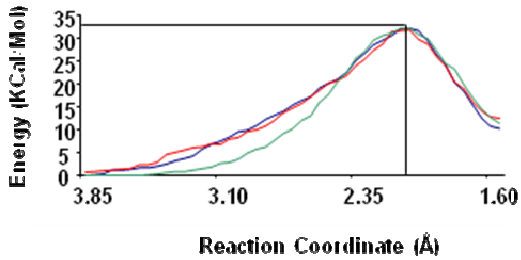
Supporting Information



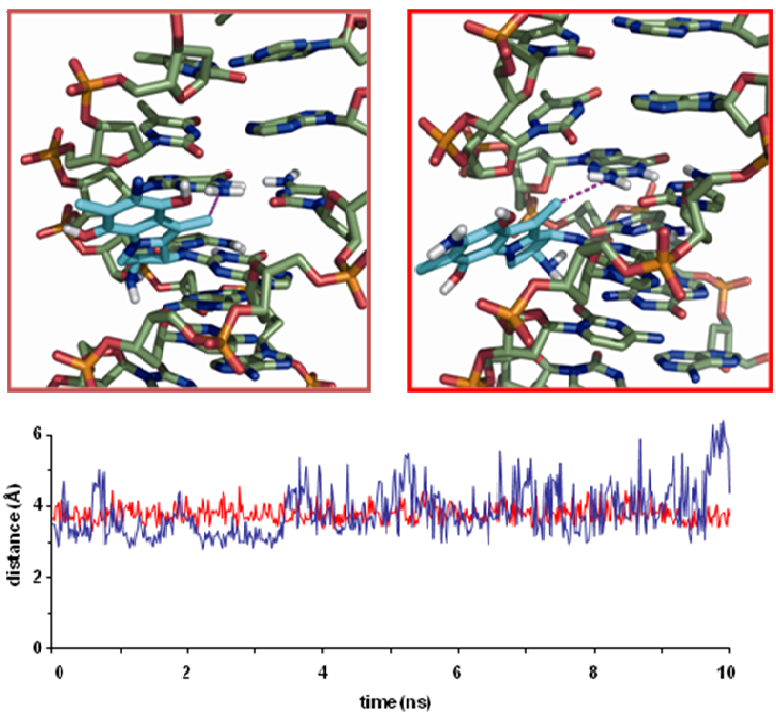
Supplementary Figure S1: representative structure of MMC^{+} obtained during the MD simulation in aqueous solution. An internal hydrogen bond between the phenol group at C8 and OM is detected.



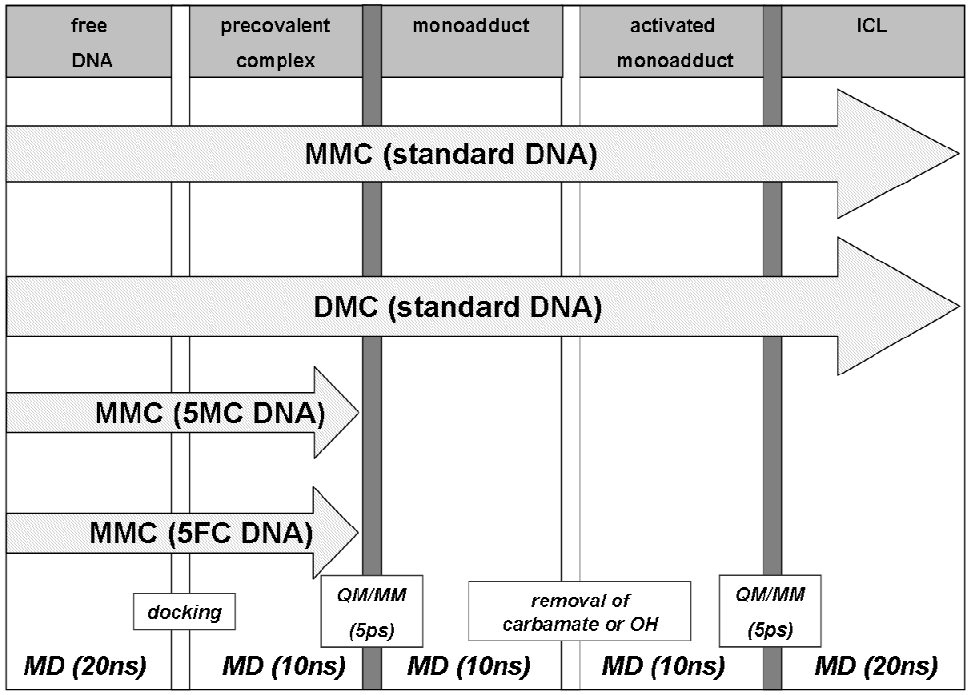
Supplementary Figure S2: O5(MMC^{+})–OP^{1}(T23) distance (black, left axis) and O5-H5-OP^{1} angle (grey, right axis) showing the feasibility of a good hydrogen bond.



Supplementary Figure S3: Free energy profiles during the formation of the MMC-monoadducts in standard (blue), 5MC- (red) and 5FC-containing (green) DNA.



Supplementary Figure S4: *Top*: Representative structures of the α- and β-monoadducts containing the unmasked second reactive group that originates upon removal of either the carbamate (MMC, left) or the hydroxyl group (DMC, right) . *Bottom*: Time evolution of the distance between N^2(G22) and C10 of the drug showing the feasibility of the second nucleophilic attack that will lead to the α-(blue) and β-(red) ICLs.



Supplementary Figure S5: Schematic representation of the different molecular complexes and techniques employed during the study of the reaction between MMC and DMC with the DNA oligonucleotides.

AMBER PREP FILES FOR MMC^+, DMC^+, 5-methyl-cytosine (5MC), 5-fluorocytosine (5FC), MMC-DNA monoadduct, DMC-DNA monoadduct, activated monoadduct and ICLs.

```
0    0    2

MMC+
MNA.data
MNA INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU    M   1   0  -1   1.7142   0.0000   0.0000   0.00000
   3  DUMM  DU    M   2   1   0   1.7139  60.0035 -90.0000   0.00000
   4  C1    CC    M   3   1   2   1.8081 125.2992 138.5085  -0.107758
   5  H1    HA    E   4   3   1   1.0673  75.2649  22.0074   0.196170
   6  C2    CT    M   4   3   1   1.5294 141.3323 258.9744   0.461678
   7  H2    HC    E   6   4   3   1.0832 107.8133  38.2064   0.030597
   8  N3    NT    B   6   4   3   1.4385 110.9918 156.4472  -1.097330
   9  H31   H     E   8   6   4   0.9981 118.1344 214.6330   0.433418
  10  H32   H     E   8   6   4   0.9993 117.9280  67.7379   0.433418
  11  C4    CT    M   6   4   3   1.5835 103.7579 282.3043  -0.088190
  12  H41   HC    E  11   6   4   1.0770 113.1085 125.2981   0.094765
  13  H42   HC    E  11   6   4   1.0778 110.5053 249.4844   0.094765
  14  N5    N2    M  11   6   4   1.4738 101.8811   6.5806   0.034748
  15  C6    CB    M  14  11   6   1.3079 137.7514 178.7114   0.093363
  16  C7    CA    M  15  14  11   1.4053 130.7339 357.9256   0.122615
  17  O8    OH    S  16  15  14   1.3873 116.3064 359.8220  -0.525990
  18  H8    HO    E  17  16  15   0.9681 113.3918 237.8675   0.406319
  19  C9    CA    M  16  15  14   1.3533 118.4193 176.3481   0.090496
  20  C10   CT    3  19  16  15   1.5063 119.9125 184.6373  -0.077929
  21  H101  HC    E  20  19  16   1.0787 110.7062  11.1528   0.066707
  22  H102  HC    E  20  19  16   1.0833 109.6068 130.4821   0.066707
  23  H103  HC    E  20  19  16   1.0861 110.8739 250.1324   0.066707
  24  C11   CA    M  19  16  15   1.4604 121.8982   4.2911   0.059848
  25  N12   N2    B  24  19  16   1.3649 119.6230 177.9162  -0.676411
  26  H121  H     E  25  24  19   0.9981 117.3720 174.9334   0.350286
  27  H122  H     E  25  24  19   0.9932 123.1237 345.1746   0.350286
  28  C13   CA    M  24  19  16   1.3551 120.4758 356.5455   0.142966
  29  O14   OH    S  28  24  19   1.3748 119.5271 180.9536  -0.487506
  30  H14   HO    E  29  28  24   0.9859 111.7590 237.6658   0.373540
  31  C15   CB    M  28  24  19   1.4211 119.2704 359.2240  -0.049679
  32  C16   CC    M  31  28  24   1.3653 132.8160 184.4347   0.017982
  33  C17   CT    3  32  31  28   1.4999 128.1634   1.2898   0.046768
  34  H171  HC    E  33  32  31   1.0773 111.0659 136.2747   0.134279
  35  H172  HC    E  33  32  31   1.0741 109.6581  13.9121   0.134279
  36  O18   OS    S  33  32  31   1.4538 110.3775 252.0078  -0.432400
  37  C19   C     B  36  33  32   1.3560 118.0497  82.5269   0.922287
  38  O20   O     E  37  36  33   1.2213 122.5661 357.3986  -0.594454
  39  N21   N     B  37  36  33   1.3270 111.3914 177.4351  -0.951931
  40  H211  H     E  39  37  36   0.9969 119.5128 180.9609   0.442576
  41  H212  H     E  39  37  36   0.9963 120.6669 357.9132   0.442576
  42  C22   CC    M   4   3   1   1.3267  62.8081 168.6436  -0.020565

IMPROPER
 C7   C15  C6   N5
 C6   C13  C15  C16
 C9   C13  C11  N12
 C7   C10  C9   C11

LOOP
 C6  C15
 C22 N5
 C1  C22

DONE
STOP
```

```
0    0    2

DMC+
DMC.data
DMC INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU     M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU     M   1   0  -1   1.7152   0.0000   0.0000   0.00000
   3  DUMM  DU     M   2   1   0   1.7144  59.9997 -90.0000   0.00000
   4  C1    CC     M   2   1   3   2.7788 134.9457 288.5103  -0.130366
   5  H1    HA     E   4   2   1   1.0627  17.5794  19.8035   0.190681
   6  C2    CT     M   4   2   1   1.4858 105.3775 213.5106   0.474506
   7  H2    HC     E   6   4   2   1.0895 112.4630 317.0778   0.052349
   8  N3    NT     B   6   4   2   1.4637 112.3300  84.8862  -1.151601
   9  H31   H      E   8   6   4   0.9970 118.8616  57.0949   0.448147
  10  H32   H      E   8   6   4   1.0231 111.6254 290.7907   0.448147
  11  C4    CT     M   6   4   2   1.5213  98.0429 201.6351  -0.086180
  12  H41   HC     E  11   6   4   1.0900 109.4919 115.2468   0.106372
  13  H42   HC     E  11   6   4   1.0895 111.1349 238.8411   0.106372
  14  N5    N2     M  11   6   4   1.4733 106.3634 357.8692   0.122252
  15  C6    CB     M  14  11   6   1.3419 137.1472 177.6360   0.013039
  16  C7    CA     M  15  14  11   1.4127 130.0937   4.4450   0.115683
  17  O8    OH     S  16  15  14   1.3662 119.3267 358.5538  -0.544045
  18  H8    HO     E  17  16  15   0.9632 114.6483 351.6282   0.420926
  19  C9    CA     M  16  15  14   1.4106 120.0487 185.7963   0.107912
  20  C10   CT     3  19  16  15   1.5272 120.7658 177.4117   0.044303
  21  H101  HC     E  20  19  16   1.0913 109.0736 264.6300   0.032262
  22  H102  HC     E  20  19  16   1.0911 110.7812  24.3618   0.032262
  23  H103  HC     E  20  19  16   1.0884 110.4762 143.6953   0.032262
  24  C11   CA     M  19  16  15   1.4116 118.3723   2.6175   0.051405
  25  N12   N2     B  24  19  16   1.3470 120.5414 173.3221  -0.729382
  26  H121  H      E  25  24  19   1.0146 117.3743 185.4478   0.350621
  27  H122  H      E  25  24  19   1.0068 121.7954 355.1018   0.350621
  28  C13   CA     M  24  19  16   1.4146 120.6326 358.3038   0.170825
  29  O14   OH     S  28  24  19   1.3702 118.6877 174.9865  -0.455208
  30  H14   HO     E  29  28  24   0.9641 113.3923 220.2506   0.377895
  31  C15   CB     M  15  14  11   1.3823 107.4654 190.8827  -0.077235
  32  C16   CC     M  31  15  14   1.4424 108.5450 356.0500   0.143998
  33  C17   CT     3  32  31  15   1.5140 125.1109 178.1091   0.072333
  34  H171  HC     E  33  32  31   1.0551 102.9783 175.7546   0.133537
  35  H172  HC     E  33  32  31   1.0925 113.7161  50.6016   0.133537
  36  O18   OH     S  33  32  31   1.4219 109.6285 280.9019  -0.629121
  37  H18   HO     E  36  33  32   0.9598 109.4817 359.8184   0.452791
  38  C22   CC     M   4   2   1   1.3586 130.4235  59.8028  -0.181899

IMPROPER
 C7   C15  C6   N5
 C6   C13  C15  C16
 C9   C13  C11  N12
 C7   C10  C9   C11

LOOP
 C6  C15
 C22 N5
 C1  C22

DONE
STOP
```

```
0    0    2

5MC
5MC.data
5MC INT 1
CORRECT OMIT DU BEG
0.00000
   1   DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2   DUMM  DU    M   1   0  -1   3.0000   0.0000   0.0000   0.00000
   3   DUMM  DU    M   2   1   0   1.0000 180.0000 -90.0000   0.00000
   4   P     P     M   1   2   3   4.3139  90.0000   0.0000   1.0735
   5   O1P   O2    E   4   1   2   1.4895  77.2736 354.8445  -0.7545
   6   O2P   O2    E   4   1   2   1.4907 148.5323 122.8410  -0.7545
   7   O5'   OS    M   4   1   2   1.5823  42.2503 132.7391  -0.4128
   8   C5'   CI    M   7   4   1   1.4096 119.0125 109.6841  -0.0111
   9   H5'1  H1    E   8   7   4   1.0902 109.5307 122.6459   0.0662
  10   H5'2  H1    E   8   7   4   1.0901 109.5057   2.6113   0.0662
  11   C4'   CT    M   8   7   4   1.5027 109.5079 242.6301   0.0721
  12   H4'   H1    E  11   8   7   1.0903 108.7974  17.8616   0.0887
  13   O4'   OS    S  11   8   7   1.4429 108.7818 137.8474  -0.3124
  14   C1'   CT    B  13  11   8   1.4162 110.8906 115.2571   0.1421
  15   H1'   H2    E  14  13  11   1.0899 106.5238 107.4191   0.0851
  16   N1    N*    S  14  13  11   1.4678 107.2760 226.7210  -0.054283
  17   C6    CM    B  16  14  13   1.3631 120.7320  60.0161  -0.168892
  18   H6    H4    E  17  16  14   1.0898 119.9275 359.7695   0.196567
  19   C5    CM    B  17  16  14   1.3283 120.2181 179.7685  -0.039121
  20   C51   CT    3  19  17  16   1.4939 119.6981 178.5209  -0.283416
  21   H511  HC    E  20  19  17   1.0906 109.5042 325.5594   0.086524
  22   H512  HC    E  20  19  17   1.0901 109.4897  85.5859   0.086524
  23   H513  HC    E  20  19  17   1.0901 109.4801 160.6337   0.086524
  24   C4    CA    B  19  17  16   1.4328 119.0361 358.8897   0.608709
  25   N4    N2    B  24  19  17   1.3188 121.6766 183.0450  -0.866133
  26   H41   H     E  25  24  19   1.0302 119.9979 358.1684   0.401555
  27   H42   H     E  25  24  19   1.0303 119.9906 178.1519   0.401555
  28   N3    NC    S  24  19  17   1.3248 120.1954   1.1970  -0.715006
  29   C2    C     S  28  24  19   1.3503 120.4517 359.7750   0.730138
  30   O2    O     E  29  28  24   1.2318 121.5643 178.9773  -0.590049
  31   C3'   CT    M  11   8   7   1.5224 111.7199 255.6649   0.1645
  32   H3'   H1    E  31  11   8   1.0904 111.0672  21.9562   0.0623
  33   C2'   CT    B  31  11   8   1.5187 103.3798 262.1936  -0.0561
  34   H2'1  HC    E  33  31  11   1.0899 104.2081 212.1208   0.0486
  35   H2'2  HC    E  33  31  11   1.0898 104.2226  92.0959   0.0486
  36   O3'   OS    M  31  11   8   1.4254 111.0644 141.9800  -0.4998

IMPROPER
 C6   C2   N1   C1'
 N1   N3   C2   O2
 C4   H41  N4   H42
 N1   C5   C6   H6
 C6   C4   C5   C5
 N3   C5   C4   N4

LOOP
 C1'  C2'
 C2   N1

DONE
STOP
```

```
0    0    2

5FC
5FC.data
5FC INT 1
CORRECT OMIT DU BEG
0.00000
   1   DUMM DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2   DUMM DU    M   1   0  -1   1.7152   0.0000   0.0000   0.00000
   3   DUMM DU    M   2   1   0   1.7141  59.9872 -90.0000   0.00000
   4   P    P     M   2   1   3   2.3519 102.8903  82.6663   0.985430
   5   O1P  O2    E   4   2   1   1.4774  64.9890 197.2207  -0.718327
   6   O2P  O2    E   4   2   1   1.4809  59.0901 348.1034  -0.718327
   7   O5'  OS    M   4   2   1   1.5941 103.5344  96.2536  -0.241912
   8   C5'  CI    M   7   4   2   1.4486 118.7635 129.0080  -0.080361
   9   H5'1 H1    E   8   7   4   1.0871 113.2007  89.9906   0.067076
  10   H5'2 H1    E   8   7   4   1.0912 108.7754 330.6867   0.067076
  11   C4'  CT    M   8   7   4   1.5121 109.7885 213.9196   0.253120
  12   H4'  H1    E  11   8   7   1.0895 108.6202 160.4756   0.083045
  13   O4'  OS    S  11   8   7   1.4620 108.7076 279.2058  -0.369232
  14   C1'  CT    B  13  11   8   1.4147 109.8083 106.3629   0.039775
  15   H1'  H2    E  14  13  11   1.0828 109.9510 114.1402   0.166181
  16   N1   N*    S  14  13  11   1.4885 107.8118 232.5314  -0.049889
  17   C6   CM    B  16  14  13   1.3565 121.2723  81.8255  -0.044172
  18   H6   H4    E  17  16  14   1.0647 120.8250   1.4511   0.172042
  19   C5   CM    B  17  16  14   1.3621 121.3395 179.7777  -0.005665
  20   F5   F     E  19  17  16   1.0899 120.2368 180.9219  -0.215007
  21   C4   CA    B  19  17  16   1.4304 116.6922   0.6635   0.695439
  22   N4   N2    B  21  19  17   1.3238 120.1478 179.6551  -0.926795
  23   H41  H     E  22  21  19   1.0148 118.7408 175.3518   0.410614
  24   H42  H     E  22  21  19   0.9994 119.6149 354.5054   0.410614
  25   N3   NC    S  21  19  17   1.3382 121.4500 359.2389  -0.766650
  26   C2   C     S  25  21  19   1.3554 120.8857   0.6418   0.780061
  27   O2   O     E  26  25  21   1.2299 117.3287 180.3243  -0.662521
  28   C3'  CT    M  11   8   7   1.5277 116.2964  36.4618   0.181949
  29   H3'  H1    E  28  11   8   1.0921 111.6861  32.1967   0.078359
  30   C2'  CT    B  28  11   8   1.5204 102.8626 273.5618  -0.057999
  31   H2'1 HC    E  30  28  11   1.0775 107.8695  87.3817   0.053334
  32   H2'2 HC    E  30  28  11   1.0755 122.9561 204.4826   0.053334
  33   O3'  OS    M  28  11   8   1.4243 112.1882 156.3259  -0.641033

IMPROPER
 C6   C2   N1   C1'
 N1   N3   C2   O2
 C4   H41  N4   H42
 N1   C5   C6   H6
 C6   C4   C5   C5
 N3   C5   C4   N4

LOOP
 C1'  C2'
 C2   N1

DONE
STOP
```

```
0    0    2

Modified first guanine covalently bonded to MMC
GMA.data
GMA INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU    M   1   0  -1   1.5507   0.0000   0.0000   0.00000
   3  DUMM  DU    M   2   1   0   1.4598 109.5046 -90.0000   0.00000
   4  P     P     M   3   2   1   1.6098 109.4815 171.8037   1.07350
   5  O1P   O2    E   4   3   2   1.4757 107.8660 128.9552  -0.75450
   6  O2P   O2    E   4   3   2   1.4826 107.0798   3.4750  -0.75450
   7  O5'   OS    M   4   3   2   1.5946 105.9670 246.7517  -0.41280
   8  C5'   CI    M   7   4   3   1.4491 118.8038 313.0108  -0.01110
   9  H5'1  H1    E   8   7   4   1.0895 109.7875  93.9948   0.06620
  10  H5'2  H1    E   8   7   4   1.0896 109.7503 333.8969   0.06620
  11  C4'   CT    M   8   7   4   1.5112 109.7559 213.9956   0.07210
  12  H4'   H1    E  11   8   7   1.0906 107.7490 156.3048   0.08870
  13  O4'   OS    S  11   8   7   1.4633 108.6274 279.0603  -0.31240
  14  C1'   CT    B  13  11   8   1.3949 111.5744 107.8753  -0.090196
  15  H1'   H2    E  14  13  11   1.0093 119.8636 113.4915   0.073923
  16  N9    N*    S  14  13  11   1.4628 109.0420 232.3016  -0.001596
  17  C8    CK    B  16  14  13   1.3930 126.5881  79.3409   0.150746
  18  H8    H5    E  17  16  14   1.0629 121.6263   0.4003   0.169314
  19  N7    NB    S  17  16  14   1.2926 111.8010 180.7062  -0.557905
  20  C5    CB    S  19  17  16   1.3844 105.7664   0.1348   0.238612
  21  C6    C     B  20  19  17   1.4266 130.1384 179.8183   0.529446
  22  O6    O     E  21  20  19   1.2056 131.9219 359.6613  -0.523593
  23  N1    NA    B  21  20  19   1.4378 109.7013 179.9680  -0.479606
  24  H1    H     E  23  21  20   1.0007 113.1710 180.1423   0.341750
  25  C2    CA    B  23  21  20   1.3541 125.6099   0.0676   0.336728
  26  N2    N2    B  25  23  21   1.3690 116.8536 180.7562  -0.379035
  27  H22   H     E  26  25  23   0.9967 119.2255 354.2878   0.319880
  28  C1    CT    B  26  25  23   1.4424 120.4474 173.5182   0.050474
  29  H11   HC    E  28  26  25   1.0835 103.2665 333.4524  -0.002830
  30  C21   CT    3  28  26  25   1.5580 112.0006 214.8592   0.210357
  31  H2    HC    E  30  28  26   1.0782 113.8990  18.3638   0.061006
  32  N31   N3    3  30  28  26   1.5360 109.9070 145.7215  -0.565438
  33  H31   H     E  32  30  28   1.0178 110.8497 223.2633   0.381371
  34  H32   H     E  32  30  28   1.0190 110.2273 101.4153   0.381371
  35  H33   H     E  32  30  28   1.0182 110.5451 340.2503   0.381371
  36  C41   CT    3  30  28  26   1.5512 103.9528 261.8929  -0.061161
  37  H41   HC    E  36  30  28   1.0773 116.2187 203.3156   0.097063
  38  H42   HC    E  36  30  28   1.0817 109.6605  72.9852   0.097063
  39  N5    N2    S  36  30  28   1.4538  98.3564 319.4666   0.076165
  40  C61   CB    S  39  36  30   1.3735 138.3439 210.9405  -0.097532
  41  C7    CA    B  40  39  36   1.3820 130.2455   0.5304   0.204247
  42  O8    OH    S  41  40  39   1.3909 115.6979   1.3519  -0.602760
  43  H88   HO    E  42  41  40   0.9658 113.4836 220.2069   0.448677
  43  C9    CA    B  41  40  39   1.3716 119.0190 180.8515  -0.027554
  44  C10   CT    3  44  41  40   1.5128 121.2229 176.9890  -0.196565
  45  H101  HC    E  45  44  41   1.0858 110.6428 130.6879   0.085476
  46  H102  HC    E  45  44  41   1.0857 110.2625 248.3155   0.085476
  47  H103  HC    E  45  44  41   1.0796 112.3737  10.0772   0.085476
  48  C11   CA    B  44  41  40   1.4191 119.7581 359.6987   0.037467
  49  N12   N2    B  49  44  41   1.3711 120.9228 179.7409  -0.695826
  50  H121  H     E  50  49  44   0.9955 118.3684 179.6910   0.344397
  51  H122  H     E  50  49  44   0.9918 122.7541 359.7868   0.344397
  52  C13   CA    B  49  44  41   1.3834 120.4037   0.1286   0.321327
  53  O14   OH    S  53  49  44   1.3966 118.2325 180.9384  -0.497586
  55  H14   HO    E  54  53  49   0.9701 112.1294  91.6167   0.381704
  54  C15   CB    S  53  49  44   1.3884 119.9383   0.1877  -0.278328
  55  C16   CC    B  56  53  49   1.4583 133.9797 179.8322   0.022496
  56  C17   CT    3  57  56  53   1.4955 128.5548 358.7011   0.034859
  57  H171  HC    E  58  57  56   1.0739 110.7849 191.8127   0.117799
  58  H172  HC    E  58  57  56   1.0781 113.0571  62.8287   0.117799
  59  O18   OS    S  58  57  56   1.4643 107.7873 307.3385  -0.353773
```

```
  60   C19   C      B  61   58   57   1.3621 117.0801 134.8037   0.894227
  61   O20   O      E  62   61   58   1.2158 121.9035  26.6266  -0.613284
  62   N21   N      B  62   61   58   1.3344 111.2884 204.8474  -1.031107
  63   H211  H      E  64   62   61   0.9956 119.5194 184.2416   0.548170
  64   H212  H      E  64   62   61   0.9936 120.0944   1.1124   0.548170
  65   C22   CC     E  57   56   53   1.3512 105.8634 179.2058  -0.090689
  66   N3    NC     S  25   23   21   1.2931 123.4868   0.0000  -0.231293
  67   C4    CB     E  16   14   13   1.3521 127.3597 260.3045  -0.031994
  68   C3'   CT     M  11    8    7   1.5272 116.4108  36.3269   0.16450
  69   H3'   H1     E  70   11    8   1.0901 111.0785  33.6484   0.06230
  70   C2'   CT     B  70   11    8   1.5198 102.8652 273.6188  -0.05610
  71   H2'1  HC     E  72   70   11   1.0898 111.3656 205.0787   0.04860
  72   H2'2  HC     E  72   70   11   1.0898 111.3543  85.0500   0.04860
  73   O3'   OS     M  70   11    8   1.4254 112.1999 156.3781  -0.49980

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 C2   C1   N2   H22
 N7   N9   C8   H8
 N1   N3   C2   N2
 C41  C61  N5   C22
 C1   C16  C22  N5
 C7   C15  C61  N5
 C9   C13  C11  N12
 C21  C22  C1   N2
 C13  C16  C15  C61

LOOP
 C1'  C2'
 C4   C5
 N5   C22
 C61  C15
 C22  C1
 C4   N9

DONE
STOP
```

```
0    0    2

Modified First Guanine covalently bonded to DMC
GDC.data
GDC INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU     M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU     M   1   0  -1   1.7146   0.0000   0.0000   0.00000
   3  DUMM  DU     M   2   1   0   1.7139  59.9969 -90.0000   0.00000
   4  P     P      M   1   2   3   5.7201 111.9836 291.0711   1.07350
   5  O1P   O2     E   4   1   2   1.4749 119.7162  50.0352  -0.75450
   6  O2P   O2     E   4   1   2   1.4785 118.8360 219.2088  -0.75450
   7  O5'   OS     M   4   1   2   1.6021  37.5749 132.0207  -0.41280
   8  C5'   CI     M   7   4   1   1.4248 118.9456 299.7463  -0.01110
   9  H5'1  H1     E   8   7   4   1.0907 110.0932  71.9653   0.06620
  10  H5'2  H1     E   8   7   4   1.0909 108.7006 313.8732   0.06620
  11  C4'   CT     M   8   7   4   1.5248 113.2894 194.0929   0.07210
  12  H4'   H1     E  11   8   7   1.0925 108.4905 170.2909   0.08870
  13  O4'   OS     S  11   8   7   1.4215 111.8989 290.5628  -0.31240
  14  C1'   CT     B  13  11   8   1.4141 109.3653 117.5153  -0.090196
  15  H1'   H2     E  14  13  11   1.0897 109.4893  98.5528   0.073923
  16  N9    N*     S  14  13  11   1.4774 107.4530 219.3298  -0.001596
  17  C8    CK     B  16  14  13   1.3738 126.7138  69.2664   0.150746
  18  H8    H5     E  17  16  14   1.0839 122.5168   5.4627   0.169314
  19  N7    NB     S  17  16  14   1.3029 114.3423 187.1852  -0.557905
  20  C5    CB     S  19  17  16   1.3993 103.4494   0.4421   0.238612
  21  C6    C      B  20  19  17   1.4350 130.3430 177.5570   0.529446
  22  O6    O      E  21  20  19   1.2286 129.3617 359.9178  -0.523593
  23  N1    NA     B  21  20  19   1.3837 113.3227 179.4668  -0.479606
  24  H1    H      E  23  21  20   1.0153 114.3394 177.9732   0.341750
  25  C2    CA     B  23  21  20   1.3769 125.0621   0.1635   0.336728
  26  N2    N2     B  25  23  21   1.3390 110.7241 179.5442  -0.379035
  27  H22   H      E  26  25  23   1.0183 111.5860 354.6159   0.319880
  28  C1    CT     B  26  25  23   1.4770 140.6415 181.0435  -0.082618
  29  H11   HC     E  28  26  25   1.0883  97.2937 131.6880   0.208403
  30  C21   CT     3  28  26  25   1.5385 112.0054 240.5704  -0.157904
  31  H2    HC     E  30  28  26   1.0925 111.5894 301.0929   0.198411
  32  N31   N3     3  30  28  26   1.4711 111.2296  70.2531  -0.151075
  33  H31   H      E  32  30  28   1.0157 109.9915 293.5859   0.287377
  34  H32   H      E  32  30  28   1.0186 112.5078 173.6498   0.287377
  35  H33   H      E  32  30  28   1.0102 109.3482  49.4897   0.287377
  36  C41   CT     3  30  28  26   1.5196  98.2353 185.3961  -0.186922
  37  H41   HC     E  36  30  28   1.0870 111.0060  85.9143   0.255941
  38  H42   HC     E  36  30  28   1.0896 113.2337 209.1665   0.255941
  39  N5    N2     S  36  30  28   1.4590 105.9094 329.6579   0.181172
  40  C61   CB     S  39  36  30   1.3246 136.6508 191.2051  -0.359320
  41  C7    CA     B  40  39  36   1.4287 126.5875 358.5318   0.374278
  42  O8    OH     S  41  40  39   1.2327 124.2490   4.8172  -0.648299
  43  H88   HO     E  42  41  40   0.9603 109.5058  56.4033   0.441862
  44  C9    CA     B  41  40  39   1.4570 114.0605 185.6119  -0.018231
  45  C10   CT     3  44  41  40   1.5321 117.6909 174.2241  -0.032504
  46  H101  HC     E  45  44  41   1.0875 113.3109 176.1748   0.039666
  47  H102  HC     E  45  44  41   1.0923 108.7377 297.6023   0.039666
  48  H103  HC     E  45  44  41   1.0894 109.0203  54.9970   0.039666
  49  C11   CA     B  44  41  40   1.3623 121.0613 354.6366   0.015773
  50  N12   N2     B  49  44  41   1.3524 123.6832 177.1728  -0.908107
  51  H121  H      E  50  49  44   1.0199 114.8446 187.8083   0.421200
  52  H122  H      E  50  49  44   1.0074 118.5094 350.2279   0.421200
  53  C13   CA     B  49  44  41   1.4811 123.2355 359.4428   0.457905
  54  O14   OH     S  53  49  44   1.2292 120.2422 185.2234  -0.618378
  55  H14   HO     E  54  53  49   0.9597 109.5223 132.2934   0.440889
  56  C15   CB     S  40  39  36   1.3753 107.1986 178.4885  -0.288534
  57  C16   CC     B  56  40  39   1.4513 109.4729   0.4971  -0.172464
  58  C17   CT     3  57  56  40   1.5180 121.8560 185.6183   0.846720
  59  H171  HC     E  58  57  56   1.0822 110.3874 152.5726  -0.150574
  60  H172  HC     E  58  57  56   1.0849 105.8355  37.9614  -0.150574
  61  O18   OH     S  58  57  56   1.4208 114.1077 278.0869  -0.785208
```

```
  62  H18   HO     E  61  58  57   0.9606 109.2408 125.1983   0.433409
  63  C22   CC     E  57  56  40   1.3743 102.8672   4.0532   0.020879
  64  N3    NC     S  25  23  21   1.3465 122.4953 359.2169  -0.231293
  65  C4    CB     E  64  25  23   1.3777 113.8776   1.6441  -0.031994
  66  C3'   CT     M  11   8   7   1.5247 109.3298  50.8067   0.16450
  67  H3'   H1     E  66  11   8   1.0914 112.4336  20.1678   0.06230
  68  C2'   CT     B  66  11   8   1.5213 101.1117 261.3994  -0.05610
  69  H2'1  HC     E  68  66  11   1.0878 109.5841  85.4433   0.04860
  70  H2'2  HC     E  68  66  11   1.0780 111.1855 206.0222   0.04860
  71  O3'   OS     M  66  11   8   1.4046 106.6317 141.4279  -0.49980

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 C2   C1   N2   H22
 N7   N9   C8   H8
 N1   N3   C2   N2
 C41  C61  N5   C22
 C1   C16  C22  N5
 C7   C15  C61  N5
 C9   C13  C11  N12
 C21  C22  C1   N2
 C13  C16  C15  C61

LOOP
 C1'  C2'
 C4   C5
 N5   C22
 C61  C15
 C22  C1
 C4   N9

DONE
STOP
```

```
0    0    2

Activated Monoadduct
GMB.data
GMB INT 1
CORRECT OMIT DU BEG
0.00000
   1   DU   DU     M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2   DU   DU     M   1   0  -1   2.0000   0.0000   0.0000   0.00000
   3   DU   DU     M   2   1   0   1.0000   0.0000 -90.0000   0.00000
   4   P    P      M   2   1   3   2.0000 180.0000   0.0000   1.07350
   5   O1P  O2     E   4   2   1   1.5172  64.2983   0.0000  -0.75450
   6   O2P  O2     E   4   2   1   1.4599  56.7706   0.0000  -0.75450
   7   O5'  OS     M   4   2   1   1.6211 141.0667   0.0000  -0.41280
   8   C5'  CI     M   7   4   2   1.4277 117.9994  87.8020  -0.01110
   9   H5'1 H1     E   8   7   4   1.0905 117.3148  38.4066   0.06620
  10   H5'2 H1     E   8   7   4   1.0902 104.0114 281.7556   0.06620
  11   C4'  CT     M   8   7   4   1.5289 112.5281 162.2313   0.07210
  12   H4'  H1     E  11   8   7   1.0900 111.5559 178.3906   0.08870
  13   O4'  OS     S  11   8   7   1.4786 106.8498 295.1284  -0.31240
  14   C1'  CT     B  13  11   8   1.4475 109.3292 116.2392   0.15570
  15   H1'  H2     E  14  13  11   1.0906 111.0174 102.2480   0.08860
  16   N9   N*     S  14  13  11   1.4947 115.8354 227.8053  -0.03790
  17   C8   CK     B  16  14  13   1.3486 126.3958  66.8633   0.13530
  18   H8   H5     E  17  16  14   1.0799 118.2109  11.9717   0.15080
  19   N7   NB     S  17  16  14   1.3339 116.8009 189.4697  -0.51390
  20   C5   CB     S  19  17  16   1.4226 100.5797   0.3012   0.05850
  21   C6   C      B  20  19  17   1.4882 130.3444 180.3825   0.50710
  22   O6   O      E  21  20  19   1.2600 127.3849 348.5150  -0.53010
  23   N1   NA     B  21  20  19   1.3635 112.6229 177.3431  -0.54050
  24   H1   H      E  23  21  20   1.0093 107.8548 183.0836   0.37500
  25   C2   CA     B  23  21  20   1.3437 123.0149 359.0884   0.68340
  26   N2   N2     B  25  23  21   1.3371 113.7386 178.3364  -0.490295
  27   H22  H      E  26  25  23   1.0104 118.9170 350.7094   0.324924
  28   C1   CT     B  26  25  23   1.4941 123.7662 157.9892   0.038317
  29   H11  HC     E  28  26  25   1.0894 104.5648 327.9915   0.126164
  30   C21  CT     3  28  26  25   1.4978 109.6269 206.8399   0.306071
  31   H2   HC     E  30  28  26   1.0903 104.9903 303.1545   0.052448
  32   N31  NT     B  30  28  26   1.4726 114.1531  72.8611  -0.991732
  33   H31  H      E  32  30  28   1.0103 109.5211 100.4852   0.389895
  34   H32  H      E  32  30  28   1.0099 109.5264 340.3997   0.389895
  35   C41  CT     3  30  28  26   1.5346  99.8383 195.2789   0.066318
  36   H41  HC     E  35  30  28   1.0900 111.3261 152.2275   0.062603
  37   H42  HC     E  35  30  28   1.0899 109.3475 275.1616   0.062603
  38   N5   N2     S  35  30  28   1.4703 106.4019  28.3593   0.030090
  39   C61  CB     S  38  35  30   1.3481 137.7121 159.3059  -0.132023
  40   C7   CA     B  39  38  35   1.4130 132.0898   8.3843   0.186079
  41   O8   OH     S  40  39  38   1.3766 123.2105 344.4828  -0.519782
  42   H88  HO     E  41  40  39   0.9596 108.9218  35.6161   0.418230
  43   C9   CA     B  40  39  38   1.4361 115.8401 177.1587   0.109653
  44   C10  CT     3  43  40  39   1.5222 120.6029 169.7373  -0.235035
  45   H101 HC     E  44  43  40   1.0908 110.6403 315.3434   0.097113
  46   H102 HC     E  44  43  40   1.0896 104.5428  76.7885   0.097113
  47   H103 HC     E  44  43  40   1.0894 110.7521 198.3206   0.097113
  48   C11  CA     B  43  40  39   1.3702 120.6771 357.0419   0.068702
  49   N12  N2     B  48  43  40   1.3626 116.8595 174.3588  -0.710275
  50   H121 H      E  49  48  43   1.0102 125.0917 197.3846   0.363299
  51   H122 H      E  49  48  43   1.0101 119.4106  21.6517   0.363299
  52   C13  CA     B  48  43  40   1.3849 123.4720   0.1537   0.201287
  53   O14  OH     S  52  48  43   1.3872 118.2145 176.3497  -0.450982
  54   H14  HO     E  53  52  48   0.9598 110.0702 137.6605   0.374216
  55   C15  CB     S  39  38  35   1.3991 105.5175 186.7479  -0.152831
  56   C16  CC     B  55  39  38   1.4817 107.3280   5.4654   0.122655
  57   C17  CT     B  56  55  39   1.5151 129.0685 178.5067  -0.037432
  58   H171 HC     E  57  56  55   1.0894 109.5184 199.4987   0.151105
  59   H172 HC     E  57  56  55   1.0905 107.4538  80.3294   0.151105
  60   C18  CC     E  38  35  30   1.3314 109.2262 348.3882   0.007632
  61   N3   NC     S  25  23  21   1.3377 126.8670   2.4559  -0.65020
```

```
  62   C4   CB     E  16  14  13   1.3534 128.0421 253.9133   0.29090
  63   C3'  CT     M  11   8   7   1.5329 117.1726  57.1110   0.16450
  64   H3'  H1     E  63  11   8   1.0896 118.4319  17.0873   0.06230
  65   C2'  CT     B  63  11   8   1.4801  97.0368 266.8781  -0.05610
  66   H2'1 HC     E  65  63  11   1.0894 106.9236  81.7710   0.04860
  67   H2'2 HC     E  65  63  11   1.0894 111.0148 204.1459   0.04860
  68   O3'  OS     M  63  11   8   1.4123 111.8090 146.8356  -0.49980

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 C2   C1   N2   H22
 N7   N9   C8   H8
 N1   N3   C2   N2
 C41  C61  N5   C18
 C1   C16  C18  N5
 C7   C15  C61  N5
 C9   C13  C11  N12
 C21  C18  C1   N2
 C13  C16  C15  C61

LOOP
 C1'  C2'
 C4   C5
 N5   C18
 C61  C15
 C18  C1
 C4   N9

DONE
STOP
```

```
0    0    2

First modified guanine in the ICL
GMO.data
GMO INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU    M   1   0  -1   1.5502   0.0000   0.0000   0.00000
   3  DUMM  DU    M   2   1   0   1.4595 109.5041 -90.0000   0.00000
   4  P     P     M   3   2   1   1.6098 109.5229 171.7957   1.07350
   5  O1P   O2    E   4   3   2   1.4760 107.8942 128.9165  -0.75450
   6  O2P   O2    E   4   3   2   1.4829 107.0576   3.4646  -0.75450
   7  O5'   OS    M   4   3   2   1.5945 105.9806 246.7548  -0.41280
   8  C5'   CI    M   7   4   3   1.4498 118.7822 312.9643  -0.01110
   9  H5'1  H1    E   8   7   4   1.0900 109.7613  93.9934   0.06620
  10  H5'2  H1    E   8   7   4   1.0895 109.7582 333.9654   0.06620
  11  C4'   CT    M   8   7   4   1.5109 109.7677 213.9796   0.07210
  12  H4'   H1    E  11   8   7   1.0901 107.7608 156.3397   0.08870
  13  O4'   OS    S  11   8   7   1.4632 108.6489 279.1130  -0.31240
  14  C1'   CT    B  13  11   8   1.4145 109.7846 106.4011   0.15570
  15  H1'   H2    E  14  13  11   1.0901 107.9285 115.7842   0.08860
  16  N9    N*    S  14  13  11   1.4877 107.7679 232.4945  -0.038777
  17  C8    CK    B  16  14  13   1.3792 129.2830  82.1486   0.123243
  18  H8    H5    E  17  16  14   1.0899 122.9320 359.9051   0.154475
  19  N7    NB    S  17  16  14   1.3081 114.1026 179.8800  -0.559339
  20  C5    CB    S  19  17  16   1.3965 103.9975   0.1633   0.294664
  21  C6    C     B  20  19  17   1.4133 130.2191 179.8644   0.388537
  22  O6    O     E  21  20  19   1.2309 129.0731   0.0000  -0.530049
  23  N1    NA    B  21  20  19   1.4059 111.4324 179.8800  -0.216712
  24  H1    H     E  23  21  20   1.0303 117.4688 179.6679   0.276784
  25  C2    CA    B  23  21  20   1.3798 125.0383 359.6854   0.105208
  26  N2    N2    S  25  23  21   1.3366 116.1488 180.2365  -0.309644
  27  H22   H     E  26  25  23   1.0305 120.0123   0.0000   0.274286
  28  N3    NC    S  25  23  21   1.3304 123.5537   0.7501  -0.211589
  29  C4    CB    E  28  25  23   1.3591 112.3692 359.4493   0.024855
  30  C3'   CT    M  11   8   7   1.5272 116.3826  36.3615   0.16450
  31  H3'   H1    E  30  11   8   1.0901 111.1425  33.6094   0.06230
  32  C2'   CT    B  30  11   8   1.5202 102.8645 273.6140  -0.05610
  33  H2'1  HC    E  32  30  11   1.0903 111.3657 205.0801   0.04860
  34  H2'2  HC    E  32  30  11   1.0894 111.3650  85.0665   0.04860
  35  O3'   OS    M  30  11   8   1.4255 112.2122 156.4009  -0.49980

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 N7   N9   C8   H8
 N1   N3   C2   N2

LOOP
 C1'  C2'
 C4   C5
 C4   N9

DONE
STOP
```

```
0    0    2

Second modified guanine in the ICL
GMM.data
GMM INT 1
CORRECT OMIT DU BEG
0.00000
   1  DUMM  DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2  DUMM  DU    M   1   0  -1   1.5502   0.0000   0.0000   0.00000
   3  DUMM  DU    M   2   1   0   1.4595 109.5041 -90.0000   0.00000
   4  P     P     M   3   2   1   1.6098 109.5229 171.7957   1.07350
   5  O1P   O2    E   4   3   2   1.4760 107.8942 128.9165  -0.75450
   6  O2P   O2    E   4   3   2   1.4829 107.0576   3.4646  -0.75450
   7  O5'   OS    M   4   3   2   1.5945 105.9806 246.7548  -0.41280
   8  C5'   CI    M   7   4   3   1.4498 118.7822 312.9643  -0.01110
   9  H5'1  H1    E   8   7   4   1.0900 109.7613  93.9934   0.06620
  10  H5'2  H1    E   8   7   4   1.0895 109.7582 333.9654   0.06620
  11  C4'   CT    M   8   7   4   1.5109 109.7677 213.9796   0.07210
  12  H4'   H1    E  11   8   7   1.0901 107.7608 156.3397   0.08870
  13  O4'   OS    S  11   8   7   1.4632 108.6489 279.1130  -0.31240
  14  C1'   CT    B  13  11   8   1.4145 109.7846 106.4011   0.15570
  15  H1'   H2    E  14  13  11   1.0901 107.9285 115.7842   0.08860
  16  N9    N*    S  14  13  11   1.4877 107.7679 232.4945  -0.038777
  17  C8    CK    B  16  14  13   1.3792 129.2830  82.1486   0.123243
  18  H8    H5    E  17  16  14   1.0899 122.9320 359.9051   0.154475
  19  N7    NB    S  17  16  14   1.3081 114.1026 179.8800  -0.559339
  20  C5    CB    S  19  17  16   1.3965 103.9975   0.1633   0.294664
  21  C6    C     B  20  19  17   1.4133 130.2191 179.8644   0.388537
  22  O6    O     E  21  20  19   1.2309 129.0731   0.0000  -0.530049
  23  N1    NA    B  21  20  19   1.4059 111.4324 179.8800  -0.216712
  24  H1    H     E  23  21  20   1.0303 117.4688 179.6679   0.276784
  25  C2    CA    B  23  21  20   1.3798 125.0383 359.6854   0.105208
  26  N2    N2    B  25  23  21   1.3366 116.1488 180.2365  -0.309644
  27  H22   H     E  26  25  23   1.0305 120.0123   0.0000   0.274286
  28  C1    CT    B  26  25  23   1.5127 117.6538 173.7370   0.050471
  29  H11   HC    E  28  26  25   1.0809 104.2807 342.1348   0.038004
  30  C21   CT    3  28  26  25   1.5724 112.3552 223.5178   0.203713
  31  H2    HC    E  30  28  26   1.0762 115.3011  18.2558   0.087449
  32  N31   N3    3  30  28  26   1.5410 107.4922 145.5955  -0.463291
  33  H31   H     E  32  30  28   1.0190 111.3881 177.5966   0.351639
  34  H32   H     E  32  30  28   1.0219 107.6278  58.7999   0.351639
  35  H33   H     E  32  30  28   1.0178 112.0462 299.1398   0.351639
  36  C41   CT    3  30  28  26   1.5552 105.7249 256.8189   0.002681
  37  H41   HC    E  36  30  28   1.0775 115.2474 210.6396   0.082141
  38  H42   HC    E  36  30  28   1.0818 109.9666  87.2135   0.082141
  39  N5    N2    S  36  30  28   1.4549 100.0325 330.6351  -0.001023
  40  C61   CB    S  39  36  30   1.3479 134.6292 187.9487  -0.119607
  41  C7    C     B  40  39  36   1.4678 126.0789  16.9804   0.401268
  42  O8    O     E  41  40  39   1.2278 119.0215 359.8077  -0.474053
  43  C9    CM    B  41  40  39   1.4451 114.6522 180.0688  -0.063269
  44  C10   CT    3  43  41  40   1.5103 117.5083 179.3860  -0.156271
  45  H101  HC    E  44  43  41   1.0870 111.3755 119.0958   0.068978
  46  H102  HC    E  44  43  41   1.0853 111.4504 240.3782   0.068978
  47  H103  HC    E  44  43  41   1.0803 109.5025 359.8728   0.068978
  48  C11   CM    B  43  41  40   1.3533 121.5368 359.2821   0.014983
  49  N12   N2    B  48  43  41   1.3384 124.9855 180.3719  -0.673605
  50  H121  H     E  49  48  43   0.9994 118.7702 180.1443   0.372075
  51  H122  H     E  49  48  43   0.9966 122.2374 359.5462   0.372075
  52  C13   C     B  48  43  41   1.5272 123.0227 359.7784   0.625178
  53  O14   O     E  52  48  43   1.2130 119.7557 181.9962  -0.450562
  54  C15   CB    S  40  39  36   1.3576 107.7085 196.8374  -0.150074
  55  C16   CC    B  54  40  39   1.4293 108.6265   2.0357  -0.036816
  56  C17   CT    B  55  54  40   1.4987 127.5583 175.6767   0.029926
  57  H171  HC    E  57  55  54   1.0774 108.9511  23.8400   0.086068
  58  H172  HC    E  57  55  54   1.0821 109.0602 143.5287   0.086068
  59  C18   CC    E  55  54  40   1.3565 105.4075 357.8175  -0.110424
  60  N3    NC    S  25  23  21   1.3304 123.5537   0.7501  -0.211589
  61  C4    CB    E  60  25  23   1.3591 112.3692 359.4493   0.024855
```

```
  62  C3'   CT    M  11   8   7   1.5272 116.3826  36.3615   0.16450
  63  H3'   H1    E  62  11   8   1.0901 111.1425  33.6094   0.06230
  64  C2'   CT    B  62  11   8   1.5202 102.8645 273.6140  -0.05610
  65  H2'1  HC    E  64  62  11   1.0903 111.3657 205.0801   0.08860
  66  H2'2  HC    E  64  62  11   1.0894 111.3650  85.0665   0.08860
  67  O3'   OS    M  62  11   8   1.4255 112.2122 156.4009  -0.49980

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 C2   C1   N2   H22
 N7   N9   C8   H8
 N1   N3   C2   N2
 C41  C61  N5   C18
 C1   C16  C18  N5
 C7   C15  C61  N5
 C9   C13  C11  N12
 C21  C18  C1   N2
 C13  C16  C15  C61

LOOP
 C1'  C2'
 C4   C5
 N5   C18
 C61  C15
 C18  C1
 C4   N9

DONE
STOP
```

Additions to the *parmbsc0* AMBER force field for non-standard residues MMC^+, DMC^+, 5MC, 5FC, GDC, GMA, GMB, GMO and GMM.

```
#  MMC+

BOND
CC-CC   480.0     1.360         (MMC+)
CB-CC   480.0     1.430         (MMC+)
CM-N2   445.0     1.340         (MMC+)
N2-CC   340.0     1.380         (MMC+)
N2-CB   340.0     1.350         (MMC+)
C -N2   425.0     1.340         (MMC+)
CC-HC   340.0     1.080         (MMC+)
N3-CB   436.0     1.410         (MMC+)
N3-CC   436.0     1.370         (MMC+)
CC-HA   340.0     1.070         (MMC+)

ANGLE
N2-CT-CC     70.0       114.35      (MMC+)
CC-CC-CT     70.0       135.00      (MMC+)
CC-CT-NT     70.0       111.00      (MMC+)
CB-CC-CC     70.0       105.40      (MMC+)
C -CB-CC     65.0       131.80      (MMC+)
N2-CM-C      70.0       112.00      (MMC+)
CM-N2-H      30.0       112.30      (MMC+)
CM-C -CB     70.0       114.96      (MMC+)
CM-CM-N2     70.0       124.97      (MMC+)
CB-N2-CC     70.0       109.30      (MMC+)
CB-C -CM     70.0       114.65      (MMC+)
CB-CB-CC     65.0       108.65      (MMC+)
N2-CB-C      70.0       126.06      (MMC+)
N2-CB-CB     70.0       107.70      (MMC+)
N2-CC-CC     70.0       108.90      (MMC+)
CT-N2-CB     60.0       134.65      (MMC+)
CT-N2-CC     60.0       114.00      (MMC+)
HC-CT-N3     50.0       113.25      (MMC+)
CT-CC-N2     70.0       109.47      (MMC+)
CB-CC-CT     70.0       127,56      (MMC+)
OS-C -N      80.0       110.93      (MMC+)
CC-CT-OS     50.0       106.94      (MMC+)
O -C -N2     80.0       126.20      (MMC+)
C -N2-H      33.0       120.40      (MMC+)
OS-C-N2      80.0       110.40      (MMC+)
HC-CC-CT     50.0       108.00      (MMC+)
HC-CT-NT     50.0       110.10      (MMC+)
CT-CC-CT     40.0       114.0       (MMC+)
CB-CC-HC     50.0       112.0       (MMC+)
N2-CT-OS     50.0       113.0       (MMC+)
CT-N2-CT     50.0        59.80      (MMC+)
CB-N3-CC     70.0       110.70      (MMC+)
N3-CB-C      70.0       127.50      (MMC+)
N3-CB-CB     70.0       107.70      (MMC+)
N3-CC-CC     70.0       108.50      (MMC+)
CT-N3-CB     60.0       134.30      (MMC+)
CT-N3-CC     60.0       115.03      (MMC+)
OH-CA-CB     70.0       121.80      (MMC+)
CA-CB-CC     63.0       134.00      (MMC+)
N2-CA-CA     70.0       118.70      (MMC+)
N2-CB-CA     70.0       130.25      (MMC+)
HA-CC-CT     35.0       126.10      (MMC+)
HA-CC-CC     35.0       123.75      (MMC+)
CC-CT-N3     60.0       109.35      (MMC+)
CC-CC-CC     60.0       144.03      (MMC+)
CT-CM-CA     60.0       121.26      (MMC+)
OS-CT-HC     50.0       111.00      (MMC+)
N2-CT-HC     50.0       107.50      (MMC+)
```

```
DIHE
X -CM-N2-X    4    9.60        180.0            2.          (MMC+)
X -CB-N2-X    4    9.60        180.0            2.          (MMC+)
X -CB-CC-X    4   14.50        180.0            2.          (MMC+)
X -N2-CC-X    4    9.60        180.0            2.          (MMC+)
X -CC-CC-X    4   14.50        180.0            2.          (MMC+)
X -C -N2-X    4    5.80        180.0            2.          (MMC+)


#  DMC+

BOND
CC-CC  480.0    1.360        (DMC+)
CB-CC  480.0    1.430        (DMC+)
CM-N2  445.0    1.340        (DMC+)
N2-CC  340.0    1.380        (DMC+)
N2-CB  340.0    1.350        (DMC+)
C -N2  425.0    1.340        (DMC+)
CC-HC  340.0    1.080        (DMC+)
N3-CB  436.0    1.410        (DMC+)
N3-CC  436.0    1.370        (DMC+)
CC-HA  340.0    1.070        (DMC+)

ANGLE
N2-CT-CC    70.0      114.35     (DMC+)
CC-CC-CT    70.0      135.00     (DMC+)
CC-CT-NT    70.0      111.00     (DMC+)
CB-CC-CC    70.0      105.40     (DMC+)
C -CB-CC    65.0      131.80     (DMC+)
N2-CM-C     70.0      112.00     (DMC+)
CM-N2-H     30.0      112.30     (DMC+)
CM-C -CB    70.0      114.96     (DMC+)
CM-CM-N2    70.0      124.97     (DMC+)
CB-N2-CC    70.0      109.30     (DMC+)
CB-C -CM    70.0      114.65     (DMC+)
CB-CB-CC    65.0      108.65     (DMC+)
N2-CB-C     70.0      126.06     (DMC+)
N2-CB-CB    70.0      107.70     (DMC+)
N2-CC-CC    70.0      108.90     (DMC+)
CT-N2-CB    60.0      134.65     (DMC+)
CT-N2-CC    60.0      114.00     (DMC+)
HC-CT-N3    50.0      113.25     (DMC+)
CT-CC-N2    70.0      109.47     (DMC+)
CB-CC-CT    70.0      127,56     (DMC+)
OS-C -N     80.0      110.93     (DMC+)
CC-CT-OS    50.0      106.94     (DMC+)
CC-CT-OH    50.0      106.94     (DMC+)
O -C -N2    80.0      126.20     (DMC+)
C -N2-H     33.0      120.40     (DMC+)
OS-C-N2     80.0      110.40     (DMC+)
HC-CC-CT    50.0      108.00     (DMC+)
HC-CT-NT    50.0      110.10     (DMC+)
CT-CC-CT    40.0      114.0      (DMC+)
CB-CC-HC    50.0      112.0      (DMC+)
N2-CT-OS    50.0      113.0      (DMC+)
CT-N2-CT    50.0       59.80     (DMC+)
CB-N3-CC    70.0      110.70     (DMC+)
N3-CB-C     70.0      127.50     (DMC+)
N3-CB-CB    70.0      107.70     (DMC+)
N3-CC-CC    70.0      108.50     (DMC+)
CT-N3-CB    60.0      134.30     (DMC+)
CT-N3-CC    60.0      115.03     (DMC+)
OH-CA-CB    70.0      121.80     (DMC+)
CA-CB-CC    63.0      134.00     (DMC+)
N2-CA-CA    70.0      118.70     (DMC+)
N2-CB-CA    70.0      130.25     (DMC+)
HA-CC-CT    35.0      126.10     (DMC+)
```

```
HA-CC-CC    35.0      123.75     (DMC+)
CC-CT-N3    60.0      109.35     (DMC+)
CC-CC-CC    60.0      144.03     (DMC+)
CT-CM-CA    60.0      121.26     (DMC+)
OS-CT-HC    50.0      111.00     (DMC+)
OH-CT-HC    50.0      111.00     (DMC+)
N2-CT-HC    50.0      107.50     (DMC+)

DIHE
X -CM-N2-X    4    9.60        180.0            2.          (DMC+)
X -CB-N2-X    4    9.60        180.0            2.          (DMC+)
X -CB-CC-X    4   14.50        180.0            2.          (DMC+)
X -N2-CC-X    4    9.60        180.0            2.          (DMC+)
X -CC-CC-X    4   14.50        180.0            2.          (DMC+)
X -C -N2-X    4    5.80        180.0            2.          (DMC+)



#  5MC
ANGLE
CT-CM-CA    60.0      121.26    5MC


#  5FC

BOND
CM-F  367.0    1.090   5FC

ANGLE
F -CM-CA    70.0      123.07     5FC
CM-CM-F     70.0      120.24     5FC



#  MMC monoadduct

BOND
CC-CC  480.0    1.360       MMC_adduct
CB-CC  480.0    1.430       MMC_adduct
CM-N2  445.0    1.340       MMC_adduct
N2-CC  340.0    1.380       MMC_adduct
N2-CB  340.0    1.350       MMC_adduct
C -N2  425.0    1.340       MMC_adduct
CC-HC  340.0    1.080       MMC_adduct
N3-CB  436.0    1.410       MMC_adduct
N3-CC  436.0    1.370       MMC_adduct
CC-HA  340.0    1.070       MMC_adduct

ANGLE
N2-CT-CC    70.0      114.35    MMC_adduct
CC-CC-CT    70.0      135.00    MMC_adduct
CB-CC-CC    70.0      105.40    MMC_adduct
C -CB-CC    65.0      131.80    MMC_adduct
N2-CM-C     70.0      112.00    MMC_adduct
CM-N2-H     30.0      112.30    MMC_adduct
CM-C -CB    70.0      114.96    MMC_adduct
CM-CM-N2    70.0      124.97    MMC_adduct
CB-N2-CC    70.0      109.30    MMC_adduct
CB-C -CM    70.0      114.65    MMC_adduct
CB-CB-CC    65.0      108.65    MMC_adduct
N2-CB-C     70.0      126.06    MMC_adduct
N2-CB-CB    70.0      107.70    MMC_adduct
N2-CC-CC    70.0      108.90    MMC_adduct
CT-N2-CB    60.0      134.65    MMC_adduct
CT-N2-CC    60.0      114.00    MMC_adduct
HC-CT-N3    50.0      113.25    MMC_adduct
CT-CC-N2    70.0      109.47    MMC_adduct
```

```
CB-CC-CT    70.0      127,56    MMC_adduct
OS-C -N     80.0      110.93    MMC_adduct
CC-CT-OS    50.0      106.94    MMC_adduct
CC-CT-OH    50.0      106.94    MMC_adduct
O -C -N2    80.0      126.20    MMC_adduct
C -N2-H     33.0      120.40    MMC_adduct
OS-C-N2     80.0      110.40    MMC_adduct
HC-CC-CT    50.0      108.00    MMC_adduct
CT-CC-CT    40.0      114.0     MMC_adduct
CB-CC-HC    50.0      112.0     MMC_adduct
N2-CT-OS    50.0      113.0     MMC_adduct
CT-N2-CT    50.0       59.80    MMC_adduct
CB-N3-CC    70.0      110.70    MMC_adduct
N3-CB-C     70.0      127.50    MMC_adduct
N3-CB-CB    70.0      107.70    MMC_adduct
N3-CC-CC    70.0      108.50    MMC_adduct
CT-N3-CB    60.0      134.30    MMC_adduct
CT-N3-CC    60.0      115.03    MMC_adduct
OH-CA-CB    70.0      121.80    MMC_adduct
CA-CB-CC    63.0      134.00    MMC_adduct
N2-CA-CA    70.0      118.70    MMC_adduct
N2-CB-CA    70.0      130.25    MMC_adduct
HA-CC-CT    35.0      126.10    MMC_adduct
HA-CC-CC    35.0      123.75    MMC_adduct
CC-CT-N3    60.0      109.35    MMC_adduct
CC-CC-CC    60.0      144.03    MMC_adduct
CT-CM-CA    60.0      121.26    MMC_adduct
OS-CT-HC    50.0      111.00    MMC_adduct
OH-CT-HC    50.0      111.00    MMC_adduct
N2-CT-HC    50.0      107.50    MMC_adduct

DIHE
X -CM-N2-X    4    9.60        180.0            2.         MMC_adduct
X -CB-N2-X    4    9.60        180.0            2.         MMC_adduct
X -CB-CC-X    4   14.50        180.0            2.         MMC_adduct
X -N2-CC-X    4    9.60        180.0            2.         MMC_adduct
X -CC-CC-X    4   14.50        180.0            2.         MMC_adduct
X -C -N2-X    4    5.80        180.0            2.         MMC_adduct


#  activated monoadducts

BOND
CC-CC  480.0    1.360       activated monoadduct
CB-CC  480.0    1.430       activated monoadduct
CM-N2  445.0    1.340       activated monoadduct
N2-CC  340.0    1.380       activated monoadduct
N2-CB  340.0    1.350       activated monoadduct
C -N2  425.0    1.340       activated monoadduct
CC-HC  340.0    1.080       activated monoadduct
N3-CB  436.0    1.410       activated monoadduct
N3-CC  436.0    1.370       activated monoadduct
CC-HA  340.0    1.070       activated monoadduct

ANGLE
N2-CT-CC    70.0      114.35    activated monoadduct
CC-CC-CT    70.0      135.00    activated monoadduct
CB-CC-CC    70.0      105.40    activated monoadduct
C -CB-CC    65.0      131.80    activated monoadduct
N2-CM-C     70.0      112.00    activated monoadduct
CM-N2-H     30.0      112.30    activated monoadduct
CM-C -CB    70.0      114.96    activated monoadduct
CM-CM-N2    70.0      124.97    activated monoadduct
CB-N2-CC    70.0      109.30    activated monoadduct
CB-C -CM    70.0      114.65    activated monoadduct
CB-CB-CC    65.0      108.65    activated monoadduct
N2-CB-C     70.0      126.06    activated monoadduct
N2-CB-CB    70.0      107.70    activated monoadduct
```

```
N2-CC-CC    70.0      108.90    activated monoadduct
CT-N2-CB    60.0      134.65    activated monoadduct
CT-N2-CC    60.0      114.00    activated monoadduct
HC-CT-NT    50.0      113.25    activated monoadduct
CT-CC-N2    70.0      109.47    activated monoadduct
CB-CC-CT    70.0      127,56    activated monoadduct
OS-C -N     80.0      110.93    activated monoadduct
CC-CT-OS    50.0      106.94    activated monoadduct
O -C -N2    80.0      126.20    activated monoadduct
C -N2-H     33.0      120.40    activated monoadduct
OS-C-N2     80.0      110.40    activated monoadduct
HC-CC-CT    50.0      108.00    activated monoadduct
CT-CC-CT    40.0      114.0     activated monoadduct
CB-CC-HC    50.0      112.0     activated monoadduct
N2-CT-OS    50.0      113.0     activated monoadduct
CT-N2-CT    50.0       59.80    activated monoadduct
CB-N3-CC    70.0      110.70    activated monoadduct
N3-CB-C     70.0      127.50    activated monoadduct
N3-CB-CB    70.0      107.70    activated monoadduct
N3-CC-CC    70.0      108.50    activated monoadduct
CT-N3-CB    60.0      134.30    activated monoadduct
CT-N3-CC    60.0      115.03    activated monoadduct
OH-CA-CB    70.0      121.80    activated monoadduct
CA-CB-CC    63.0      134.00    activated monoadduct
N2-CA-CA    70.0      118.70    activated monoadduct
N2-CB-CA    70.0      130.25    activated monoadduct
HA-CC-CT    35.0      126.10    activated monoadduct
HA-CC-CC    35.0      123.75    activated monoadduct
CC-CT-N3    60.0      109.35    activated monoadduct
CC-CC-CC    60.0      144.03    activated monoadduct
OS-CT-HC    50.0      111.00    activated monoadduct
N2-CT-HC    50.0      107.50    activated monoadduct

DIHE
X -CM-N2-X    4    9.60        180.0            2.         activated monoadduct
X -CB-N2-X    4    9.60        180.0            2.         activated monoadduct
X -CB-CC-X    4   14.50        180.0            2.         activated monoadduct
X -N2-CC-X    4    9.60        180.0            2.         activated monoadduct
X -CC-CC-X    4   14.50        180.0            2.         activated monoadduct
X -C -N2-X    4    5.80        180.0            2.         activated monoadduct



#  ICLs

BOND
CC-CC  480.0    1.360       ICL
CB-CC  480.0    1.430       ICL
CM-N2  445.0    1.340       ICL
N2-CC  340.0    1.380       ICL
N2-CB  340.0    1.350       ICL
C -N2  425.0    1.340       ICL
CC-HC  340.0    1.080       ICL
N3-CB  436.0    1.410       ICL
N3-CC  436.0    1.370       ICL
CC-HA  340.0    1.070       ICL

ANGLE
N2-CT-CC    70.0      114.35    ICL
CC-CC-CT    70.0      135.00    ICL
CB-CC-CC    70.0      105.40    ICL
C -CB-CC    65.0      131.80    ICL
N2-CM-C     70.0      112.00    ICL
CM-N2-H     30.0      112.30    ICL
CM-C -CB    70.0      114.96    ICL
CM-CM-N2    70.0      124.97    ICL
CB-N2-CC    70.0      109.30    ICL
```

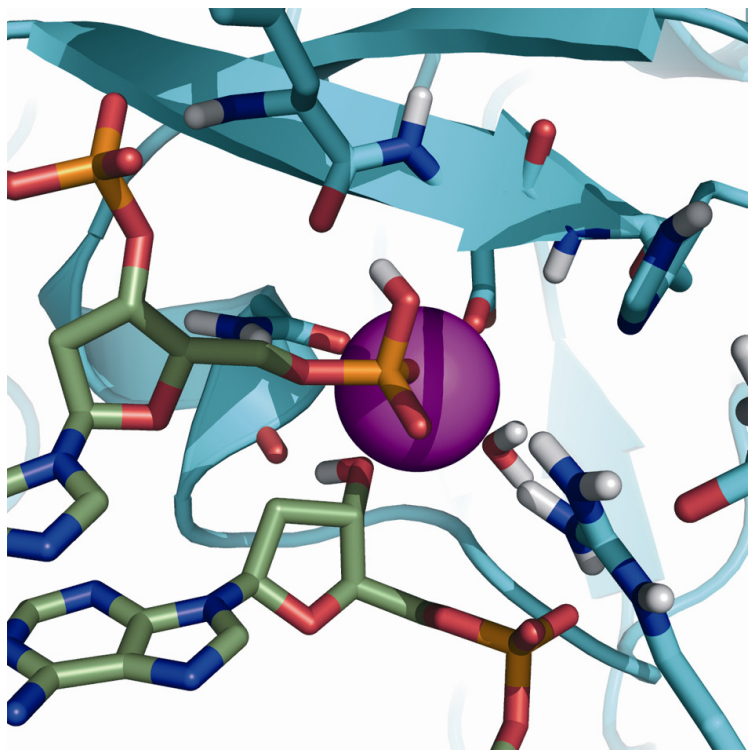
```
CB-C -CM    70.0      114.65    ICL
CB-CB-CC    65.0      108.65    ICL
N2-CB-C     70.0      126.06    ICL
N2-CB-CB    70.0      107.70    ICL
N2-CC-CC    70.0      108.90    ICL
CT-N2-CB    60.0      134.65    ICL
CT-N2-CC    60.0      114.00    ICL
HC-CT-N3    50.0      113.25    ICL
CT-CC-N2    70.0      109.47    ICL
CB-CC-CT    70.0      127,56    ICL
OS-C -N     80.0      110.93    ICL
CC-CT-OS    50.0      106.94    ICL
O -C -N2    80.0      126.20    ICL
C -N2-H     33.0      120.40    ICL
OS-C-N2     80.0      110.40    ICL
HC-CC-CT    50.0      108.00    ICL
CT-CC-CT    40.0      114.0     ICL
CB-CC-HC    50.0      112.0     ICL
N2-CT-OS    50.0      113.0     ICL
CT-N2-CT    50.0       59.80    ICL
CB-N3-CC    70.0      110.70    ICL
N3-CB-C     70.0      127.50    ICL
N3-CB-CB    70.0      107.70    ICL
N3-CC-CC    70.0      108.50    ICL
CT-N3-CB    60.0      134.30    ICL
CT-N3-CC    60.0      115.03    ICL
OH-CA-CB    70.0      121.80    ICL
CA-CB-CC    63.0      134.00    ICL
N2-CA-CA    70.0      118.70    ICL
N2-CB-CA    70.0      130.25    ICL
HA-CC-CT    35.0      126.10    ICL
HA-CC-CC    35.0      123.75    ICL
CC-CT-N3    60.0      109.35    ICL
CC-CC-CC    60.0      144.03    ICL
OS-CT-HC    50.0      111.00    ICL)
N2-CT-HC    50.0      107.50    ICL

DIHE
X -CM-N2-X    4    9.60        180.0            2. ICL
X -CB-N2-X    4    9.60        180.0            2. ICL
X -CB-CC-X    4   14.50        180.0            2. ICL
X -N2-CC-X    4    9.60        180.0            2. ICL
X -CC-CC-X    4   14.50        180.0            2. ICL
X -C -N2-X    4    5.80        180.0            2. ICL
```

ARTÍCULO VI.

"Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: Vibrio vulnificus periplasmic nuclease as a test case"

ChemBioChem DOI: 10.1002/cbic.201100485



Utilizando información de *Vibrio vulnificus* (Vvn) de tipo silvestre y varios mutantes así como de la *homing* endonucleasa I-PpoI co-cristalizadas con distintos oligodesoxinucleótidos, hemos construido el complejo de Vvn junto con un octámero de ADN y llevado a cabo una serie de simulaciones para analizar el mecanismo catalítico de esta enzima paso por paso. Las distintas funciones desempeñadas en la reacción por los residuos que forman el sitio activo, el catión metálico y las moléculas de agua se han aclarado utilizando una combinación de simulaciones de MD clásica y cálculos de mecánica cuántica. Nuestros resultados apoyan fuertemente como mecanismo catalítico más probable aquél en el que una única molécula de agua procedente del entorno acuoso se utiliza para hidrolizar el enlace fosfodiéster y protonar el grupo 3'-hidroxilo saliente.

DOI: 10.1002/cbic.201100485

Computational study of the catalytic mechanism of Vvn

Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: *Vibrio vulnificus* periplasmic nuclease as a test case

Juan A. Bueren-Calabuig,[a] Claire Coderch,[a] Eva Rico,[b] Antonio Jiménez-Ruiz,[b] and Federico Gago*[a]

Using information from wild-type and mutant Vibrio vulnificus nuclease (Vvn) and I-PpoI homing endonuclease co-crystallized with different oligodeoxynucleotides, we have built the complex of Vvn with a DNA octamer and carried out a series of simulations to dissect the catalytic mechanism of this metallonuclease in a stepwise fashion. The distinct roles played in the reaction by individual active site residues, the metal cation and water molecules have been clarified using a combination of classical molecular dynamics simulations and quantum mechanical calculations. Our results strongly support the most parsimonious catalytic mechanism, namely one in which a single water molecule from bulk solvent is used to cleave the phosphodiester bond and protonate the 3'-hydroxylate leaving group.

Introduction

Sugar-nonspecific endonucleases belong to a family of enzymes that are able to cleave single- and double-stranded (ds) DNA and/or RNA molecules with little or no sequence specificity.[1] They have been identified in several organisms and the three-dimensional structures of some of them have been solved by X-ray crystallography and deposited in the Protein Data Bank (PDB).[2] The best studied representatives are *Serratia marcescens* nuclease (Smn)[3,4], nuclease A from *Anabaena sp.* (NucA),[5] and the periplasmic nuclease from *Vibrio vulnificus* (Vvn).[6,7] From a structural viewpoint, all of these enzymes belong to the ββα-metal superfamily of nucleases that use just one divalent metal ion[8] to bring the reactant species into close proximity and catalyze the reaction.[9] In this respect, the cation, usually but not exclusively magnesium, has been proposed to act as a Lewis acid to activate a bound nucleophile (e.g. promoting ionization of water to a more reactive hydroxide) and/or to stabilize the negative charge on the pentacoordinate trigonal bipyramidal transition state (TS) and the 3'-hydroxylate leaving group.

The common structural core of these metallonucleases is made up of two antiparallel β-sheets and one α-helix that contains an essential and strictly conserved asparagine (Asn) residue whose carboxamide oxygen is used to coordinate the metal ion, hence the family name (Supporting Figure S1). A signature motif in the first β-strand is a DRGH sequence, where the H is the active site histidine,[1] although in Vvn it is replaced by EWEH,[6] with the peculiarity that the side-chain carboxylate of the second glutamate (E) residue is also involved in metal binding (Mg^{2+} or Ca^{2+}). The finding that site-directed mutation of the common histidine (H) in this motif leads to a dramatic decrease in endonuclease activity has demonstrated the critical role of this residue in catalysis, most likely as a general base that activates the water molecule responsible for internucleotide phosphodiester cleavage.[6] Thus, following abstraction of a proton from the water molecule, the resulting hydroxide anion would initiate an in-line nucleophilic attack on the phosphorus atom leading to cleavage of the O3'–P bond. The two final products of the reaction are a nucleic acid strand with a 5'-phosphorylated end and another one with a terminal 3' hydroxyl group. Nonetheless, there appears to be no consensus yet as to whether the attacking water is in direct coordination with Mg^{2+} or is a free solvent molecule.[3,10-12] Equally debated is the origin of the proton that needs to be donated to the 3'-hydroxylate leaving group. In this respect, either a general acid (e.g. Glu127 in Smn[4,13] or Glu163 in NucA[14]) or a second water molecule from the inner coordination sphere of the metal ion have been advocated.[6,13,15] The large number and diversity of the proposed mechanisms attest to the fact that the atomic details of some steps of the reaction are still unclear despite the relative wealth of crystallographic structures, albeit in the absence of DNA.[1] An exception is Vvn, for which two different crystals were

[a] *Juan A. Bueren-Calabuig, Claire Coderch, Prof. Federico Gago*
Department of Pharmacology, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid (Spain)
Fax: (+34)918 854 514
E-mail: federico.gago@uah.es
[b] *Dr. Eva Rico, Prof. Antonio Jiménez-Ruiz*
Department of Biochemistry and Molecular Biology, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid (Spain)

Supporting information for this article is available on the WWW under http://www.chembiochem.org.

solved some years ago, one containing the active form of the metallonuclease in the absence of DNA and another containing two His80Ala mutant enzymes with Ca^{2+} in the active site forming a complex with two DNA octanucleotides, one intact and one cleaved.[6,15] More recently, the structure of another complex of Vvn in complex with a 16-bp DNA in the absence of metal ions was solved.[7] However, none of these enzyme–substrate (E–S) or enzyme–product (E–P) complexes provides an answer to the above-mentioned questions principally because no water molecule is properly positioned for attack and because the DNA phosphate backbone adopts a standard conformation that is unsuitable for the reaction to proceed.

The active site of Smn, NucA and Vvn is also very similar to those of apoptotic endonuclease G from *Drosophila melanogaster* (Dm-endoG),[16] the His-Cys box homing endonuclease I-*Ppo*I from the slime mould *Physarum polycephalum*,[17-19] the DNA-entry nuclease EndA from *Streptococcus pneumoniae*,[20] and a recently identified endonuclease from the protozoan parasite *Leishmania infantum* (Li-endoG)[21] (Supporting Figure S1). Moreover, cleavage experiments with deoxythymidine 3',5'-bis-(*p*-nitrophenylphosphate) as an artificial minimal substrate have indeed shown that Smn, NucA and I-*Ppo*I follow the same mechanism of phosphodiester bond hydrolysis[11,13,22] Of particular interest for our purposes is I-*Ppo*I because its DNA-bound structure has been determined as three distinct species on the reaction pathway, including two trapped E–S complexes that clearly display the hydrolytic water in the vicinity of the strictly conserved catalytic histidine (His98) and one E–P complex with the same ds-DNA 21-mer cleaved in one strand.[17,18] Of note, in the complex containing an inactive His98Ala mutant, the Mg^{2+} ion is coordinated by the side-chain carboxamide oxygen of the conserved asparagine (Asn119), three water oxygens and two phosphate oxygens from the DNA substrate: the bridging O3' and the non-bridging OP1. This information is particularly worthy as it clearly indicates that the phosphate backbone has to undergo a well-defined conformational change at the site of cleavage.

Given that advances in computational power and maturity of force fields for nucleic acids and proteins allow us to fill in some of the missing pieces that experimental techniques cannot always provide,[23] it is currently possible to simulate the course of an enzymatic reaction stepwise, starting with formation of the Michaelis-Menten protein-substrate complex and ending up with product release after resolution of the TS.[24,25] In order to shed light into the catalytic mechanism of Vvn, we decided to use a combination of molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) calculations on the active form of Vvn in complex with a ds-DNA octamer. The main advantage of a hybrid approach involving MD and QM for the solvated DNA-protein complex over studies that focus on simplified model systems is that bond making and breaking are possible in a dynamic context and also that reorganization of active site residues and water molecules is realistically allowed through the whole procedure. Our results strongly support the most parsimonious catalytic mechanism, namely one in which there is no need to invoke more than a single water molecule to accomplish the complete reaction, that is, cleavage of the phosphodiester bond and protonation of the leaving hydroxylate.

Results and Discussion

Initial Vvn-DNA complex

The initial model used in our simulations represents a Michaelis-Menten complex of the active form of Vvn, that is, a fully competent enzyme containing the catalytic His80 and Mg^{2+} as the metal cation, in complex with a ds-DNA octamer in which the phosphodiester bond to be cleaved adopts a suitable conformation for the nucleophilic attack, as found in the crystal structure of a I-*Ppo*I-DNA complex (PDB id. 1CYQ).[17] Consequently, this structure, which still lacks a properly positioned hydrolytic water, differs from those obtained for Vvn using X-ray crystallography, which show either an active enzyme in the absence of DNA (PDB id. 1OUO) or a mutant His80Ala enzyme in complex with either an octanucleotide and bound Ca^{2+} (PDB id. 1OUP) or a 15-mer in the absence of any cation (PDB id. 2IVK). In the latter two cases, the DNA phosphate backbone adopts a standard B-type conformation.[6,7] In our model, on the other hand, the non-bridging OP1(Gua_{15}) and the bridging O3'(Ade_{14}) complete the octahedral coordination of Mg^{2+} in addition to Oε1(Glu79), Oδ(Asn127), and the oxygens of two water molecules that are present in the apo-form crystal of Vvn. According to this geometrical disposition, the O3'–P phosphodiester bond to be cleaved belongs to the phosphate linking Ade_{14} to Gua_{15} in the second DNA strand, i.e. the underlined step in d(TCTTAAGA)·d(TCTTAAGA). Importantly, by adopting this backbone conformation, binding of the DNA strand to Vvn is further stabilized by two additional hydrogen bonds: one between O5'(Gua_{15}) and Nδ(Asn127), and another between OP2(Gua_{15}) and the guanidinium of Arg99, which is fixed in position by ion pair formation with the carboxylate of Glu113 and appears bound to the terminal 5'-phosphate in the E–P complex.[6]

The side chain of the catalytically essential Asn127 is thus remarkably fixed for binding to O5' through Nδ and for bringing the negatively charged phosphate into a proper conformation for attack because, in addition to its backbone carbonyl oxygen being coordinated to Mg^{2+}, Asn127 simultaneously engages its Nδ in a hydrogen bond to the carboxylate of Glu77, which is in turn stabilized by the guanidinium of Arg72 (Figure 1). On the other hand, the imidazole ring of His80, the residue responsible for water molecule activation (i.e. proton abstraction by Nδ), is also relatively fixed and made more basic because of a hydrogen bond between its Nε and the backbone carbonyl oxygen of Glu113[6] (positionally equivalent to Cys105 in I-*Ppo*I[17] and Asn106 in Smn[3,4]).

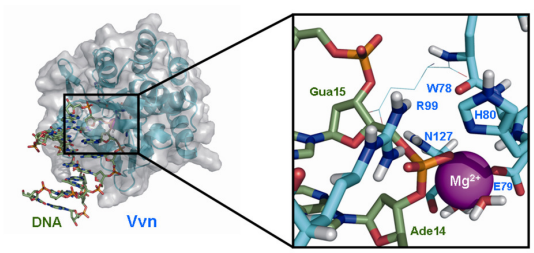


Figure 1. Structural model and close-up view of Vvn (cyan) in complex with DNA (green) in a near-attack reactive conformation. The active site Mg^{2+} is coordinated by two water molecules, O=C(Asn127), a carboxylate oxygen from Glu79, the non-bridging oxygen OP1(Gua_{15}) and O3'(Ade_{14}). The Vvn-DNA complex is stabilized in this region by an electrostatic interaction between the guanidinium of Arg99 and the scissile phosphate and also by a hydrogen bond that is established between Nδ(Asn127) and O5'(Gua_{15}).

Incidentally, another stable hydrogen bond that was observed in our simulation, as a consequence of the sugar-phosphate rearrangement and mutual DNA-protein adaptation, was that established between NH1(Arg72) and N3(Ade_{14}). This interaction could account for the slight abundance of adenines at sites cleaved by Vvn despite its little sequence preference for DNA.[7] The guanidinium of Arg72 is held in position for facing the DNA minor groove by means of a buttressing interaction with the carboxylate of Glu77 (the first E in the EWEH motif), as also seen in the Vvn–DNA crystal structures. However, the closer interaction between Arg72 and the DNA base is not so apparent possibly because the side chain of this residue faces a G:C base pair in PDB entries 2IVK and 1OUP and, consequently, the presence of an exocyclic amino group in the minor groove decreases the goodness of this interaction with either N3 of t nor any need to invoke the **unlikely deprotonation ADDIN**

Activation of a water molecule by His80

After 3 ns of MD equilibration of the Vvn-DNA complex in aqueous solution, a water molecule from the bulk solvent entered the active site and remained fixed at a location suitable for the nucleophilic attack to proceed. In fact, an average OW(water)–P(Gua_{15}) distance of ≤ 4.0 Å and an average O3'(Ade_{14})–P(Gua_{15})–OW(water) angle of ~176º suggested an ideal position for in-line attack on the O3'–P bond. Furthermore, the distance between the same OW atom in this location and both Nδ(His80) and O=C(Trp78) was also ≤ 4.0 Å. Therefore this water molecule got unbiasedly trapped for the rest of the simulation, with one of its oxygen's lone pairs properly oriented for attack, even though its protons (HW) swapped their interactions with the two hydrogen bond acceptor atoms of these two amino acids (Figure 2). When we compared this water location with that found for the hydrolytic water in the I-*Ppo*I–DNA complexes, they were found to be perfectly superimposable.

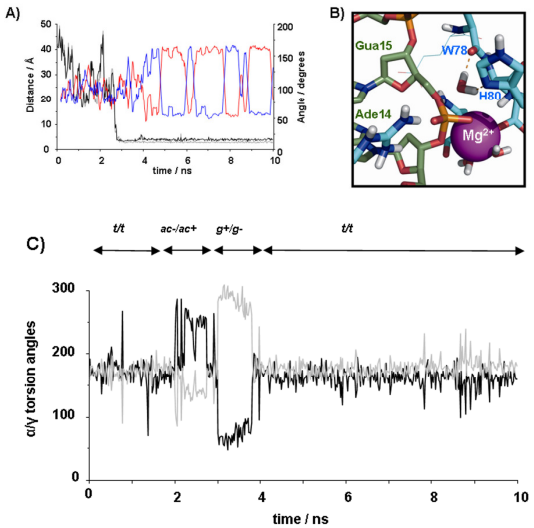


Figure 2. A) Time evolution of the distances between the oxygen (OW) of the nucleophilic water molecule and both the phosphorus atom (black) and Nδ(His80) (grey). The angle formed by each water hydrogen (HW1 and HW2, red and blue), the water oxygen (OW), and Nδ(His80) is also represented in the secondary vertical axis. B) Representative snapshot from our simulation of the active site of Vvn in its complex with the DNA octamer showing this water molecule correctly hydrogen-bonded to O=C(Trp78) and Nδ(His80) and suitably positioned for the in-line attack. C) Time evolution of the α and γ backbone torsion angles of Gua_{15} (*t*, *ac*, and *g* stand for *trans*, *anticlinal*, and *gauche*, respectively). Note that the stability achieved after ~4 ns is concomitant with the monotonic OW–P distance displayed in Figure 2A.

The α/γ torsion angles of the cleavable phosphate group linking Ade_{14} and Gua_{15} remained mostly constant in a *trans/trans* (*t/t*) conformation during the first 2 ns of simulation but then changed to negative and positive anticlinal (*ac–/ac+*) and later to positive and negative *gauche* (*g+/g–*) (Figure 2B). After these fast crankshaft motions of the DNA sugar-phosphate backbone, the α/γ torsionals returned to their original *t/t* configuration for the rest of the simulation. During all this time, Nδ(Asn127) remained hydrogen-bonded to O5'(Gua_{15}) except for the period when the α/γ torsion angles were in the ac–/ac+ DNA conformation. Interestingly, the return to the catalytically relevant *t/t* backbone configuration after the initial fluctuations was concomitant with the approach and proper positioning of the catalytic water molecule and fixation via hydrogen bonding to both Nδ(His80) and O=C(Trp78).

Moreover, when we docked deoxythymidine 3',5'-bis-(*p*-nitrophenylphosphate) into the active site of Vvn using an automated force-field-based method that allows for ligand flexibility,[26] the best-scoring pose adopted the same *t/t* conformation for the α/γ torsion angles of the cleavable phosphate (Supporting Figure S2). This artificial minimal substrate is cleaved by Smn,[11,13] NucA[14] and I-*Ppo*I[22] at its 5'-end, albeit at a much lower rate than DNA or RNA, to produce nitrophenol and deoxythymidine 3'-(*p*-nitrophenylphosphate) 5-phosphate. The scissile bond for these metallonucleases is therefore different from that attacked by staphylococcal nuclease (O5'–P) or pancreatic DNase I (O–P bond linking the other *p*-nitrophenol moiety to the 3'-phosphate on thymidine),[1] in agreement with the fact that Smn, NucA, I-*Ppo*I and, by extension, Vvn need at least a second phosphate group to be present 3' to the bond to be cleaved. This has also been shown for Smn using oligodeoxynucleotides substituted by a single phosphorothioate group.[11] In the case of Vvn, this second phosphate hydrogen bonds to the backbone NH of Trp85, a residue whose side-chain indole nitrogen, together with that of Trp94, is engaged in yet another hydrogen bond with the following phosphate in the oligonucleotide strand. On the other side of the cleavable bond, it is the side chains of Arg75 and Lys28 that establish hydrogen bonds with the O5' and OP2 of another nucleotide in the same strand, as well as Lys54 with the next phosphate.[6,15] Thus, in common with Smn,[11] a minimum of five phosphate groups would appear to be necessary for high-affinity DNA binding and moderately good cleavage activity by Vvn.

The simulated trajectory of the Vvn–DNA complex showed that the short hydrogen bond between Nε(His80) and O=C(Glu113), which is present in the crystal structure[6] and thought to make the imidazole a stronger base to facilitate water deprotonation, was also maintained. To proceed with the cleavage reaction, therefore, we subjected the active site region to QM treatment during the course of the MD simulation and defined two double reaction coordinates for simultaneously shortening the HW---Nε and OW---P distances and stretching the HW–OW and O3'–P bonds, as explained in Methods.

Nucleophilic attack and intermediate product generation

As a result of the QM reaction coordinate, (i) a hydroxide was generated *in situ* that attacked the phosphodiester bond, (ii) an imidazolium was formed on the side chain of His80, and (iii) the negative charge on the pentacoordinate trigonal bipyramidal TS was stabilized by Mg^{2+} (coordinated to OP1 and O3') and the guanidinium side chain of Arg99 (forming an ion pair with OP2) (Supporting Figure S3). This latter interaction has been previously observed only in the E–P complexes solved by X-ray crystallography.[6,15] Subsequent collapse of the TS yielded the intermediate products of this first part of the reaction, namely a 5'-phosphorylated dinucleotide, d(GA), and a d(TCTTAA) hexanucleotide with a terminal 3'-hydroxylate, both of them still hydrogen bonded to the intact complementary strand, d(TCTTAAGA) (Supporting movie M1).

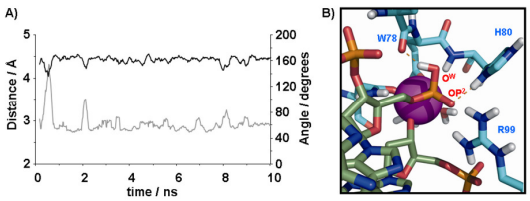


Figure 3. A) Time evolution of (i) the distance between the phosphate non-bridging oxygen OP2 (grey) and Nδ(His80), and (ii) the angle between OP2, HNδ(His80) and Nδ(His80) (black). B) The HNδ(His80) proton (originally HW) is proposed to be transferred to the OP2 phosphate on its way to the O3' hydroxylate.

An unrestrained MD simulation of this updated system configuration showed the immediate loss of the original hydrogen bond between the protonated Nδ(His80$^+$) and OW (now belonging to the new terminal phosphate) and formation of a new hydrogen bond between the imidazolium Nδ and OP2 of the same phosphate group. This latter interaction was maintained for the remaining part of the 10-ns simulation except for brief exchanges with OW (Figure 3). At this point, the active site architecture reproduced very accurately the overall geometry found in the crystallographic E–P complex containing the His80Ala mutant Vvn (PDB id. 1OUP) but showed an improved hydrogen-bonding interaction between the terminal phosphate O5' and Nδ(Asn127). On the other hand, the hydrogen bond between OP2 and NH2(Arg99) may ensure that this product does not evolve backwards to favour the reverse reaction[6] (see below). The equivalent residue in Smn is Arg57, whose mutation to Ala has been shown to decrease by >99% the activity of the wild-type enzyme.[11]

Deprotonation of His80

In view of the stability of the hydrogen bond established between OP2 and Nδ(His80$^+$), we reasoned that it was perfectly feasible for the proton on Nδ to be transferred to this negatively charged phosphate oxygen. A QM reaction coordinate was employed for this purpose (Supporting Figure S4) and the new connectivity was used to simulate the resulting system configuration for a further 10 ns. Interestingly, in this new simulation, the two hydroxyls in the terminal phosphate rotated about their O–P bonds after ~3 ns and exchanged hydrogen bond acceptors through water-mediated interactions (Figure 4). Thus, the original hydrogen bond between Nδ(His80$^+$) and OP2 was lost and replaced with a new one between OP2 and the negatively charged, Mg^{2+}-bound O3'. Reciprocally, Nδ(His80) became engaged in a hydrogen bond with the phosphate OW after losing its close interaction with O=C(Trp78), which in turn became solvated by water.

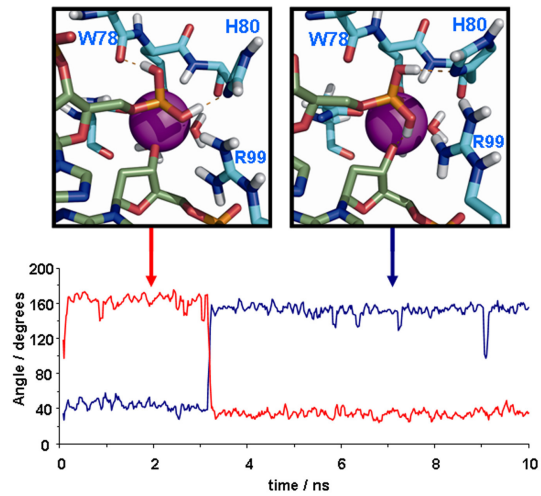


Figure 4. Swapping of hydrogen-bonding partners for the newly created terminal phosphate bearing two hydroxyl groups on the phosphorus atom after transfer of the proton from Nδ(His80) to OP2: the OP2–HW:Nδ(His80) hydrogen bond (red) is broken whereas a new OP2–HW:O3' hydrogen bond (blue) is formed.

Protonation of the leaving group

In this new setting it became apparent that OP2, still hydrogen-bonded to NH1(Arg99), was the most likely proton donor to O3'$^{(-)}$ (Figure 4). This shuttle mechanism would imply that the leaving group can be protonated by a hydrogen originating from the same water molecule that initiated the nucleophilic attack. The QM reaction coordinate showed the feasibility of this proposal (Supporting Figure S5) and rendered (i) a neutral hydroxyl whose oxygen was still coordinating the Mg^{2+}, and (ii) a terminal 5'-phosphate bearing a charge of –1 (Figure 5). Consequently, there appears to be neither structural support nor any need to invoke the unlikely deprotonation[9,27] by the metal cation of a second water molecule, as postulated in several unproven mechanistic proposals. On the one hand, O3'$^{(-)}$ itself is coordinating the metal, on the same plane as the two liganded water molecules; on the other hand, the only other water close to O3'$^{(-)}$ belongs to the second solvation shell of Mg^{2+} and is hydrogen bonded to OP2(Gua$_{15}$) and NH1(Arg99).

Exit of the products from the active site

During the MD simulation of the system containing the cleaved DNA strand, (i) both OP1(Gua$_{15}$) and O3'(Ade$_{14}$) retained their coordination to the Mg^{2+} cation, and (ii) the backbone O=C(Asn127) and carboxamide Nδ(Asn127) remained hydrogen bonded to O3' (acting as a donor) and O5' (acting as an acceptor),

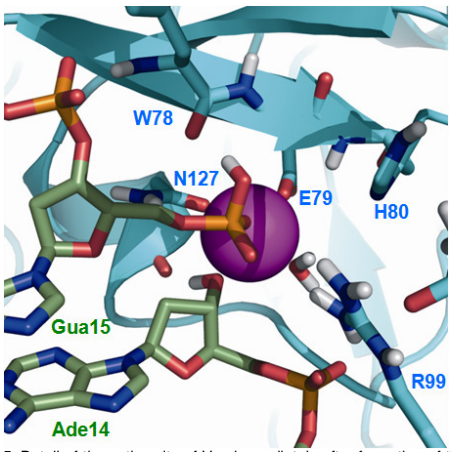
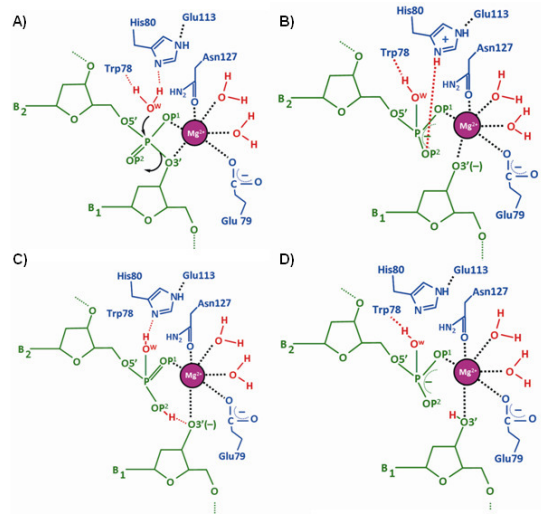


Figure 5. Detail of the active site of Vvn immediately after formation of the final products of the cleavage reaction. Note the singly protonated terminal 5'-phosphate, the free 3'-hydroxyl and the neutral imidazole in His80.

respectively. On the contrary, the guanidinium of Arg99 started to weaken its interaction with the just created terminal 5'-phosphate on Gua_{15}, which began to get fully solvated, but maintained its hydrogen bond with the phosphate group linking Ade_{14} to Ade_{13}. Although much longer simulation times could be expected to reproduce the separation of the DNA containing the cleaved strand from the Mg^{2+}–bound protein as a result of continued solvation of the binding partners, we did not pursue this issue any further after 25 ns. Instead, we made use of a normal mode analysis (NMA) to study the large-amplitude protein motions that most likely contribute to vacation of the active site and product release. Our rationale was that whereas a 10-ns MD simulation has been shown to be long enough for an enzymatic protein to visit several crystallographic conformations,[28] other larger-scale motions possibly related to catalysis occur on the microsecond-to-millisecond timescale. Fortunately, experimentally observed conformational changes connecting open and closed states, which usually limit the overall catalytic rate of many enzymes, are often correlated with low-frequency normal modes[23,25] that can be inexpensively calculated.[29] For Vvn, we found that the first nontrivial lowest-frequency mode (upon removal of the six rigid motions of rotation and translation) represented a clasping motion involving two domains on both sides of the DNA-binding cleft (Supporting movie M2). Vvn can then be likened to a hand that is able to grasp an incoming DNA when closed and to release a cleaved oligonucleotide when open. For the latter to occur, Arg72 (supported by Glu77) and Arg75 in one domain would be in charge of pulling the nucleotide containing the 5'-phosphate out of the active site whereas Arg99 (supported by Glu113) in the other domain would evacuate the strand containing the free 3'-hydroxyl by virtue of a strong electrostatic interaction with the preceding phosphate. This view is consistent with the finding that at the end of our 25-ns fully unrestrained MD simulation Arg99 started to lose its interaction with O5'(Gua_{15}) (>3.7 Å) whereas it maintained a short hydrogen bond with O3'(Ade_{14}).

Conclusions



Scheme 1. Full catalytic mechanism of Vvn in the light of our computer simulation results. Note that the hydrogen bond between O5'(Gua_{15}) and Nδ(Asn127) has been omitted for clarity.

Although the active site architecture of several non-specific ββα-metallonucleases and related enzymes is known in atomic detail, and despite the fact that site-directed mutagenesis experiments have shed light into the critical residues involved in DNA cleavage, their full catalytic mechanism has remained elusive so far. The main uncertainties concern the origin of the hydrolytic water and the source of the hydrogen that protonates the leaving group. We thought that theoretical methods that combine QM and MM in a dynamic context could be of value in this respect and chose the well-characterized Vvn as a representative example of this interesting class of nucleases.

Starting with a Vvn–DNA complex that was built using crystallographic information from the structures of the complexes of wild-type and mutant Vvn and I-*Ppo*I nuclease with different oligonucleotides, we carried out a series of simulations to dissect the catalytic mechanism of Vvn in a stepwise fashion. The stability of the DNA and protein conformations in the complex throughout the whole procedure was assessed by monitoring the root-mean-square deviation from the initial model (Supporting Figure S6). The results obtained have allowed us to characterize the putative roles that individual active site residues play in the reaction path and have provided plausible answers to some long-standing issues. The principal findings (Scheme 1) can be summarized as follows: (1) the imidazole of His80, aided by O=C(Glu113), functions as the general base that abstracts a proton from a water molecule and generates the hydroxide anion that then serves as the attacking nucleophile on the phosphorus atom; (2) this water molecule originates from the bulk solvent, and not from the hydration sphere of the metal cation, and is held in place by Nδ(His80) and O=C(Trp78); (3) proper positioning of this water is coupled to a sugar-phosphate rearrangement that is made possible by (i) coordination of Mg^{2+} with the non-bridging OP1 and the bridging O3' oxygens in the phosphate group to be cleaved, and (ii) hydrogen-bonding of O5' to Nδ(Asn127), (4) attack of the *in situ* generated hydroxide is in-line with the O3'–P phosphodiester bond and the associative (S_N2) reaction proceeds without a covalent intermediate; (5) the pentacoordinate trigonal

bipyramidal phosphorane-like TS is stabilized by Mg^{2+}, which is coordinated by OP1 and O3', and by the guanidinium side chain of Arg99, which forms an ion pair with OP2; (6) an OP2-mediated proton shuttle from Nδ(His80) to the O3' leaving group appears more probable than the abstraction of a proton from a second water molecule in the coordination sphere of Mg^{2+}, as previously proposed.

The catalytic mechanism that we have dissected for Vvn is likely to be shared by other members of the family and related ββα-metal nucleases that rely on the activation of a water molecule by the imidazole ring of a suitably positioned His to cleave both DNA and RNA in double- and single-stranded forms. The role of the only metal ion would be, on the one hand, to attract and stabilize the DNA substrate by forming a network of interactions with protein residues, water molecules and nucleotides that thereafter facilitates the positioning and activation of a water molecule for the nucleophilic attack and, on the other hand, to stabilize the TS of the reaction, which also interacts strongly with Arg99.

We are aware of the limitations of the semiempirical method that we have used, which was the only one currently implemented in the AMBER distribution that allowed us to study a molecular system containing both P atoms and a Mg^{2+} ion. For this reason we have not attempted to obtain the free energy profile of the reaction, which still needs to be done using a better Hamiltonian for the QM calculations. Nonetheless the structural and geometrical characterization provides a plausible mechanism and will hopefully open new avenues for further theoretical and experimental research.

Computational Methods

Construction of the Vvn-DNA complex

Since no structural data of the active form of Vvn in complex with DNA were available, a model was built by superimposing the ββα motif of wild-type, Mg^{2+}-bound Vvn (PDB id. 1OUO) in the absence of DNA onto the ββα of the homing endonuclease I-*Ppo*I in complex with a 21-mer (PDB id. 1CYQ),[17] of which we kept only the central self-complementary $d(TCTTAAGA)_2$ octanucleotide. As a result, $OP1(Gua_{15})$ and $O3'(Ade_{14})$ from the DNA octamer occupied the same positions as two crystallographic water molecules in the unbound Vvn structure. The Mg^{2+} was then coordinated by these two phosphate oxygens, Oδ1(Asn127), Oε1(Glu79), and two other water molecules. By superimposing the crystal structure of inactive, Ca^{2+}-bound Vvn complexed with DNA (PDB id. 1OUP) over our construction, the most noticeable difference affected the conformation of the O3'–P–O5'–C5'(α)/O5'–C5'–C4'–C3'(γ) torsion angles of the cleavable phosphate.

Molecular dynamics simulations

The Vvn-DNA complex was immersed in a truncated octahedron of TIP3P water molecules[30] that extended 15 Å away from any solute atom. Incorporation of the appropriate number of sodium ions[31] at random locations ensured electrical neutrality. The cutoff distance for the non-bonded interactions was 9 Å and periodic boundary conditions were used. Electrostatic interactions were treated using the smooth particle mesh Ewald (PME) method[32] with a grid spacing of 1 Å. The SHAKE algorithm[33] was applied to all bonds involving hydrogens and an integration step of 2.0 fs was used throughout. The molecular dynamics (MD) simulation protocol made use of *sander* (for the QM/MM part, see below) and *pmemd* modules in the AMBER 11 suite of programs.[34] Firstly, solvent molecules and counterions were relaxed by energy minimization and allowed to redistribute around the positionally restrained solute (25 $kcal \cdot mol^{-1} \cdot Å^{-2}$) during 50 ps of MD at constant temperature (300 K) and pressure (1 atm), essentially as described previously.[35] These initial harmonic restraints were gradually reduced in a series of progressive energy minimizations until they were completely removed. The resulting system, as well as those resulting from the hybrid QM/MM calculations described below, were heated from 100 to 300 K during 20 ps, equilibrated at 300 K for 1 ns and further simulated under the same conditions up to a total time of 10 ns during which system coordinates were collected every 20 ps for further analysis. NMR-type restraints were used during the initial classical MD for the Mg^{2+} coordinating sphere, i.e. two water oxygens, Oε(Glu79), $OP1(Gua_{15})$, and Oδ(Asn127) to avoid known problems of ligand exchange.[10] An additional restraint between one coordinating water oxygen and Oδ(Asn127) was also included but no restraints were applied between Mg^{2+} and O3'. The restraining values used were r_1= 1.30, r_2= 1.80, r_3= 2.19, and r_4= 2.69 Å with the force constants $k_2 = k_3$ = 20 kcal/mol·Å^2. The final E–P simulation, which lasted 25 ns, was fully unrestrained. Each ns of MD took ~17 hours of CPU time running in parallel on 32 IBM Power PC 970MP processors at the Barcelona Supercomputing Center (MareNostrum).

Hybrid QM/MM calculations.

Classical molecular mechanics (MM) methods lack the ability to treat fundamentally quantum processes, such as bond breaking/forming and charge fluctuations as a function of geometry,[36] but it is possible to treat a subsection of the system by quantum mechanical (QM) methods and use a coupling potential to connect the MM and QM regions. This hybrid QM/MM approach is seamlessly integrated in the sander module of AMBER 11[34] and also includes a complete treatment of long range electrostatics using a QM/MM modified PME method.[37,38] The QM region, defined by the *iqmatoms* keyword, encompassed the active site region where bonds are broken and formed, and the MM region included all the remaining protein, DNA and solvent atoms as well as the counterions. Care was taken not to cut any polar bonds when defining the QM/MM boundary. Thus, the QM region contained the whole of Ade_{14} (up to C4'), the phosphate and C5'of Gua_{15}, Mg^{2+} and its two coordinating water molecules, the catalytic water, and the side chains (starting from Cβ) of Glu79, His80, Trp78, Asn127 and Arg99. We selected the PM3 Hamiltonian[39] and full electrostatic interactions between the QM charge density (expanded in a STO-6G minimal basis set) and the point charges of the MM atoms. Three different reaction coordinates at constant velocities were implemented employing a steering MD protocol[40] and a linear combination of distances (Å) between pairs of atoms: (i) for activation of the water molecule by His80, attack of the resulting hydroxide onto the phosphorus atom and subsequent O3–P cleavage: $\xi = d_{OW\text{-}H} - d_{N\text{-}H}$ and $\xi = d_{PO3'} - d_{OWP}$; (ii) for transfer of the proton from His80 to OP1: $\xi = d_{N\text{-}H} - d_{OP1\text{-}H}$, and (iii) for proton transfer from OP2 to the O3' leaving group: $\xi = d_{OP2\text{-}H} - d_{O3'\text{-}H}$. The time frames were 20 ps for (i) and 5 ps for (ii) and (iii) using in all cases the recommended force constant of 1000 kcal/(mol·Å). During the QM/MM part of the MD simulation SHAKE (keyword *qmshake*) was turned off and the integration step was reduced to 0.5 fs.

Analysis of the molecular dynamics trajectories

Three-dimensional structures and trajectories were visually inspected using the computer graphics program PyMOL.[41] Interatomic distances and angles were monitored using the *ptraj* module in

AMBER and program Curves+ .[42] 360º degrees were added to all negative torsion values for plotting purposes.

Normal mode analysis

The X-ray crystal structure of Vvn (PDB entry 1OUO)[6] at 2.30 Å resolution was retrieved from the Protein Data Bank.[2] An elastic network model connected all Cα atoms by springs representing the interatomic force fields and non-hydrogen protein atoms (within a cutoff of 10 Å) were considered as point masses. Vvn was then analyzed as a large set of coupled harmonic oscillators using the NOMAD–Ref server[29] and default parameters. The 10 lowest-frequency normal modes were calculated and each mode was explored in its two opposite directions. For the first non-trivial mode corresponding to the largest amplitude, the 30 resulting conformations were refined by geometry optimization of the side chains using the sander module of AMBER 10.

Ligand docking

The CGRID and CDOCK programs,[43] as implemented in the VSDMIP platform,[44] were used to dock the minimal substrate deoxythymidine 3',5'-bis-(*p*-nitrophenylphosphate) into the active site of Vvn, as found in PDB entry 1OUO.[6]

Acknowledgements

We are grateful to Dr. Alberto Gómez and Mr. Alvaro Cortés for assistance with VSDMIP and to Prof. María José Camarasa and Dr. Sonsoles Velázquez (IQM, CSIC) for many fruitful discussions. We thankfully acknowledge the generous allowance of computer resources, technical expertise and assistance provided by the Red Española de Supercomputación at the Barcelona Supercomputing Center (MareNostrum). This work was partially supported by Comisión Interministerial de Ciencia y Tecnología (SAF2009-13914-C02-02) and Comunidad Autónoma de Madrid (S-BIO/0214/2006). J.A.B.-C. enjoys a research fellowship from PharmaMar S.A.U. (Colmenar Viejo, Madrid, Spain).

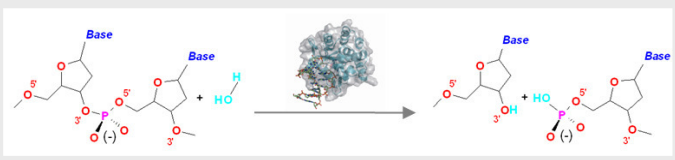
Keywords: enzyme catalysis · DNA cleavage · water splitting · molecular dynamics · quantum chemistry

Received: ((will be filled in by the editorial staff))
Published online: ((will be filled in by the editorial staff))

[1] E. S. Rangarajan, V. Shankar, *FEMS Microbiol. Rev.* **2001**, *25*, 583-613.
[2] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* **1977**, *112*, 535-542.
[3] S. V. Shlyapnikov, V. V. Lunin, M. Perbandt, K. M. Polyakov, V. Y. Lunin, V. M. Levdikov, C. Betzel, A. M. Mikhailov, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2000**, *56*, 567-572.
[4] V. Y. Lunin, V. M. Levdikov, S. V. Shlyapnikov, E. V. Blagova, V. V. Lunin, K. S. Wilson, A. M. Mikhailov, *FEBS Lett.* **1997**, *412*, 217-222.
[5] M. Ghosh, G. Meiss, A. Pingoud, R. E. London, L. C. Pedersen, *J. Biol. Chem.* **2005**, *280*, 27990-27997.
[6] C. L. Li, L. I. Hor, Z. F. Chang, L. C. Tsai, W. Z. Yang, H. S. Yuan, *EMBO J.* **2003**, *22*, 4014-4025.
[7] Y. T. Wang, W. J. Yang, C. L. Li, L. G. Doudeva, H. S. Yuan, *Nucleic Acids Res.* **2007**, *35*, 584-594.
[8] C. M. Dupureur, *Metallomics* **2010**, *2*, 609-620.
[9] J. A. Cowan, *BioMetals* **2002**, *15*, 225-235.
[10] C. Chen, B. W. Beck, K. Krause, B. M. Pettitt, *Proteins* **2006**, *62*, 982-995.
[11] P. Friedhoff, G. Meiss, B. Kolmes, U. Pieper, O. Gimadutdinow, C. Urbanke, A. Pingoud, *Eur. J. Biochem.* **1996**, *241*, 572-580.
[12] M. D. Miller, J. Cai, K. L. Krause, *J. Mol. Biol.* **1999**, *288*, 975-987.
[13] B. Kolmes, I. Franke, P. Friedhoff, A. Pingoud, *FEBS Lett.* **1996**, *397*, 343-346.
[14] G. Meiss, O. Gimadutdinow, B. Haberland, A. Pingoud, *J. Mol. Biol.* **2000**, *297*, 521-534.
[15] K. C. Hsia, C. L. Li, H. S. Yuan, *Curr. Opin. Struct. Biol.* **2005**, *15*, 126-134.
[16] B. Loll, M. Gebhardt, E. Wahle, A. Meinhart, *Nucleic Acids Res.* **2009**, *37*, 7312-7320.
[17] E. A. Galburt, B. Chevalier, W. Tang, M. S. Jurica, K. E. Flick, R. J. Monnat, Jr., B. L. Stoddard, *Nat. Struct. Biol.* **1999**, *6*, 1096-1099.
[18] E. A. Galburt, B. L. Stoddard, *Biochemistry* **2002**, *41*, 13851-13860.
[19] K. E. Flick, M. S. Jurica, R. J. Monnat, Jr., B. L. Stoddard, *Nature* **1998**, *394*, 96-101.
[20] M. Midon, P. Schafer, A. Pingoud, M. Ghosh, A. F. Moon, M. J. Cuneo, R. E. London, G. Meiss, *Nucleic Acids Res.* **2011**, *39*, 623-634.
[21] E. Rico, J. F. Alzate, A. A. Arias, D. Moreno, J. Clos, F. Gago, I. Moreno, M. Dominguez, A. Jimenez-Ruiz, *Mol. Biochem. Parasitol.* **2009**, *163*, 28-38.
[22] P. Friedhoff, I. Franke, K. L. Krause, A. Pingoud, *FEBS Lett.* **1999**, *443*, 209-214.
[23] E. Marco, F. Gago, *ChemMedChem* **2007**, *2*, 1388-1401.
[24] A. Negri, E. Marco, J. Damborsky, F. Gago, *J. Mol. Graphics Modell.* **2007**, *26*, 643-651.
[25] A. Negri, D. Rodríguez-Larrea, E. Marco, A. Jiménez-Ruiz, J. M. Sánchez-Ruiz, F. Gago, *Proteins* **2010**, *78*, 36-51.
[26] K. Wang, M. Murcia, P. Constans, C. Perez, A. R. Ortiz, *J. Comput. Aided Mol. Des.* **2004**, *18*, 101-118.
[27] J. A. Cowan, *Chem. Rev.* **1998**, *98*, 1067-1088.
[28] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hubner, D. Kern, *Nature* **2007**, *450*, 838-844.
[29] E. Lindahl, C. Azuara, P. Koehl, M. Delarue, *Nucleic Acids Res.* **2006**, *34*, W52-56.
[30] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[31] J. Aqvist, *J. Phys. Chem.* **1990**, *94*, 8021-8024.
[32] T. A. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
[33] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
[34] D. A. Case, T. A. Darden, I. Cheatham, T.E., C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, in *University of California, San Francisco*, **2010**.
[35] E. Marco, A. Negri, F. J. Luque, F. Gago, *Nucleic Acids Res.* **2005**, *33*, 6214-6224.
[36] G. M. de Seabra, R. C. Walker, M. Elstner, D. A. Case, A. E. Roitberg, *J. Phys. Chem. A* **2007**, *111*, 5655-5664.
[37] R. C. Walker, M. F. Crowley, D. A. Case, *J. Comput. Chem.* **2008**, *29*, 1019-1031.
[38] K. Nam, J. Gao, D. M. York, *J. Chem. Theory Comput.* **2004**, *1*, 2-13.
[39] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209-220.
[40] A. Crespo, M. A. Marti, D. A. Estrin, A. E. Roitberg, *J. Am. Chem. Soc.* **2005**, *127*, 6940-6941.
[41] W. D. Lano, in *DeLano Scientific LLC*, **2006**, p. URL: http://www.pymol.org/.
[42] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, *Nucleic Acids Res.* **2009**, *37*, 5917-5929.
[43] C. Perez, A. R. Ortiz, *J. Med. Chem.* **2001**, *44*, 3768-3785.
[44] R. Gil-Redondo, J. Estrada, A. Morreale, F. Herranz, J. Sancho, A. R. Ortiz, *J. Comput. Aided Mol. Des.* **2009**, *23*, 171-184.

Entry for the Table of Contents

FULL PAPERS



Parsimony in water splitting at the active site of ββα-metallonucleases. Although two water molecules are usually depicted in the catalytic mechanism of these enzymes, one to cleave the phosphodiester bond and another to protonate the leaving hydroxylate, our computational results strongly suggest that it is the same water molecule from bulk solvent that provides both the nucleophilic hydroxide and the proton that are necessary to accomplish the complete reaction.

*Juan A. Bueren-Calabuig, Claire Coderch, Eva Rico, Antonio Jiménez-Ruiz, and Federico Gago**

Page No. – Page No.

Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: *Vibrio vulnificus* periplasmic nuclease as a test case

References (EndNote)
[1] E. S. Rangarajan, V. Shankar, *FEMS Microbiol. Rev.* **2001**, *25*, 583-613.
[2] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* **1977**, *112*, 535-542.
[3] S. V. Shlyapnikov, V. V. Lunin, M. Perbandt, K. M. Polyakov, V. Y. Lunin, V. M. Levdikov, C. Betzel, A. M. Mikhailov, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2000**, *56*, 567-572.
[4] V. Y. Lunin, V. M. Levdikov, S. V. Shlyapnikov, E. V. Blagova, V. V. Lunin, K. S. Wilson, A. M. Mikhailov, *FEBS Lett.* **1997**, *412*, 217-222.
[5] M. Ghosh, G. Meiss, A. Pingoud, R. E. London, L. C. Pedersen, *J. Biol. Chem.* **2005**, *280*, 27990-27997.
[6] C. L. Li, L. I. Hor, Z. F. Chang, L. C. Tsai, W. Z. Yang, H. S. Yuan, *EMBO J.* **2003**, *22*, 4014-4025.
[7] Y. T. Wang, W. J. Yang, C. L. Li, L. G. Doudeva, H. S. Yuan, *Nucleic Acids Res.* **2007**, *35*, 584-594.
[8] C. M. Dupureur, *Metallomics* **2010**, *2*, 609-620.
[9] J. A. Cowan, *BioMetals* **2002**, *15*, 225-235.
[10] C. Chen, B. W. Beck, K. Krause, B. M. Pettitt, *Proteins* **2006**, *62*, 982-995.
[11] P. Friedhoff, G. Meiss, B. Kolmes, U. Pieper, O. Gimadutdinow, C. Urbanke, A. Pingoud, *Eur. J. Biochem.* **1996**, *241*, 572-580.
[12] M. D. Miller, J. Cai, K. L. Krause, *J. Mol. Biol.* **1999**, *288*, 975-987.
[13] B. Kolmes, I. Franke, P. Friedhoff, A. Pingoud, *FEBS Lett.* **1996**, *397*, 343-346.
[14] G. Meiss, O. Gimadutdinow, B. Haberland, A. Pingoud, *J. Mol. Biol.* **2000**, *297*, 521-534.
[15] K. C. Hsia, C. L. Li, H. S. Yuan, *Curr. Opin. Struct. Biol.* **2005**, *15*, 126-134.
[16] B. Loll, M. Gebhardt, E. Wahle, A. Meinhart, *Nucleic Acids Res.* **2009**, *37*, 7312-7320.
[17] E. A. Galburt, B. Chevalier, W. Tang, M. S. Jurica, K. E. Flick, R. J. Monnat, Jr., B. L. Stoddard, *Nat. Struct. Biol.* **1999**, *6*, 1096-1099.
[18] E. A. Galburt, B. L. Stoddard, *Biochemistry* **2002**, *41*, 13851-13860.
[19] K. E. Flick, M. S. Jurica, R. J. Monnat, Jr., B. L. Stoddard, *Nature* **1998**, *394*, 96-101.
[20] M. Midon, P. Schafer, A. Pingoud, M. Ghosh, A. F. Moon, M. J. Cuneo, R. E. London, G. Meiss, *Nucleic Acids Res.* **2011**, *39*, 623-634.
[21] E. Rico, J. F. Alzate, A. A. Arias, D. Moreno, J. Clos, F. Gago, I. Moreno, M. Dominguez, A. Jimenez-Ruiz, *Mol. Biochem. Parasitol.* **2009**, *163*, 28-38.
[22] P. Friedhoff, I. Franke, K. L. Krause, A. Pingoud, *FEBS Lett.* **1999**, *443*, 209-214.
[23] E. Marco, F. Gago, *ChemMedChem* **2007**, *2*, 1388-1401.
[24] A. Negri, E. Marco, J. Damborsky, F. Gago, *J. Mol. Graphics Modell.* **2007**, *26*, 643-651.
[25] A. Negri, D. Rodríguez-Larrea, E. Marco, A. Jiménez-Ruiz, J. M. Sánchez-Ruiz, F. Gago, *Proteins* **2010**, *78*, 36-51.
[26] K. Wang, M. Murcia, P. Constans, C. Perez, A. R. Ortiz, *J. Comput. Aided Mol. Des.* **2004**, *18*, 101-118.
[27] J. A. Cowan, *Chem. Rev.* **1998**, *98*, 1067-1088.
[28] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hubner, D. Kern, *Nature* **2007**, *450*, 838-844.
[29] E. Lindahl, C. Azuara, P. Koehl, M. Delarue, *Nucleic Acids Res.* **2006**, *34*, W52-56.
[30] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[31] J. Aqvist, *J. Phys. Chem.* **1990**, *94*, 8021-8024.
[32] T. A. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
[33] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
[34] D. A. Case, T. A. Darden, I. Cheatham, T.E., C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C.

Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, in *University of California, San Francisco*, **2010**.
[35] E. Marco, A. Negri, F. J. Luque, F. Gago, *Nucleic Acids Res.* **2005**, *33*, 6214-6224.
[36] G. M. de Seabra, R. C. Walker, M. Elstner, D. A. Case, A. E. Roitberg, *J. Phys. Chem. A* **2007**, *111*, 5655-5664.
[37] R. C. Walker, M. F. Crowley, D. A. Case, *J. Comput. Chem.* **2008**, *29*, 1019-1031.
[38] K. Nam, J. Gao, D. M. York, *J. Chem. Theory Comput.* **2004**, *1*, 2-13.
[39] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209-220.
[40] A. Crespo, M. A. Marti, D. A. Estrin, A. E. Roitberg, *J. Am. Chem. Soc.* **2005**, *127*, 6940-6941.
[41] W. D. Lano, in *DeLano Scientific LLC*, **2006**, p. URL: http://www.pymol.org/.
[42] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, *Nucleic Acids Res.* **2009**, *37*, 5917-5929.
[43] C. Perez, A. R. Ortiz, *J. Med. Chem.* **2001**, *44*, 3768-3785.
[44] R. Gil-Redondo, J. Estrada, A. Morreale, F. Herranz, J. Sancho, A. R. Ortiz, *J. Comput. Aided Mol. Des.* **2009**, *23*, 171-184.

Supporting Information for cbic.201100485

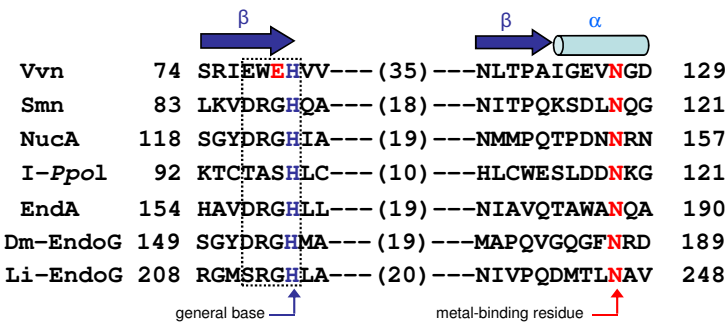


Figure S1. Multiple sequence alignment of the ββα motifs from *Vibrio vulnificus* periplasmic nuclease (Vvn), *Serratia marcescens* nuclease (Smn), nuclease A from *Anabaena sp.* (NucA), *Physarum polycephalum* His-Cys box endonuclease I-*Ppo*l, *Streptococcus pneumoniae* DNA-entry nuclease EndA, and apoptotic endonucleases G from *Drosophila melanogaster* (Dm-endoG) and *Leishmania infantum* (Li-endoG).

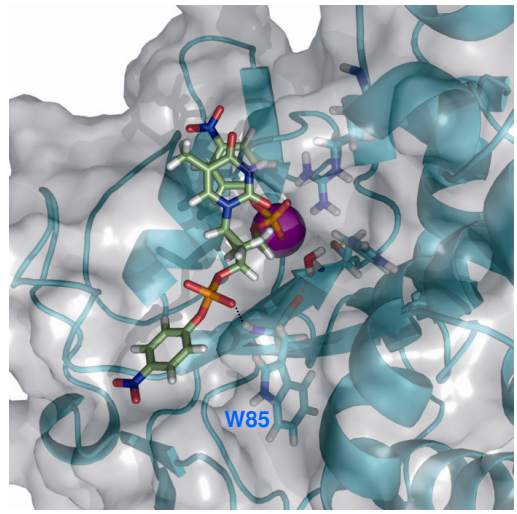
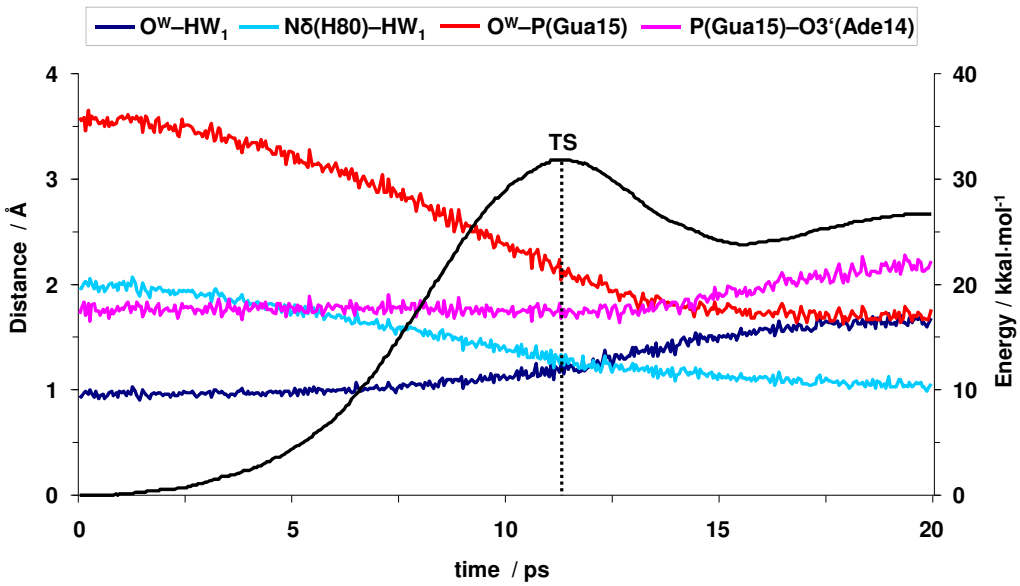


Figure S2. Result of the automated docking of the minimal substrate 3',5'-bis-(*p*-nitrophenylphosphate) in the active site of Vvn. Note that the conformation of the phosphate close to the Mg^{2+} is the same as that in the DNA strand that was cleaved in our simulations.



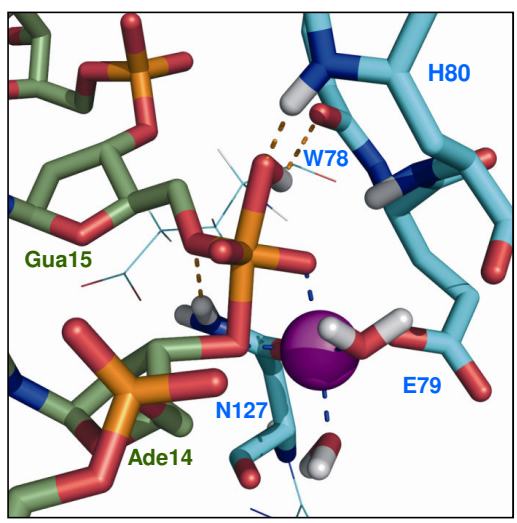


Figure S3: Energy profile (top) of the activation of the water molecule by His80 and subsequent nucleophilic attack of the resulting hydroxide on the phosphodiester bond through the transition state (TS) shown at the bottom.

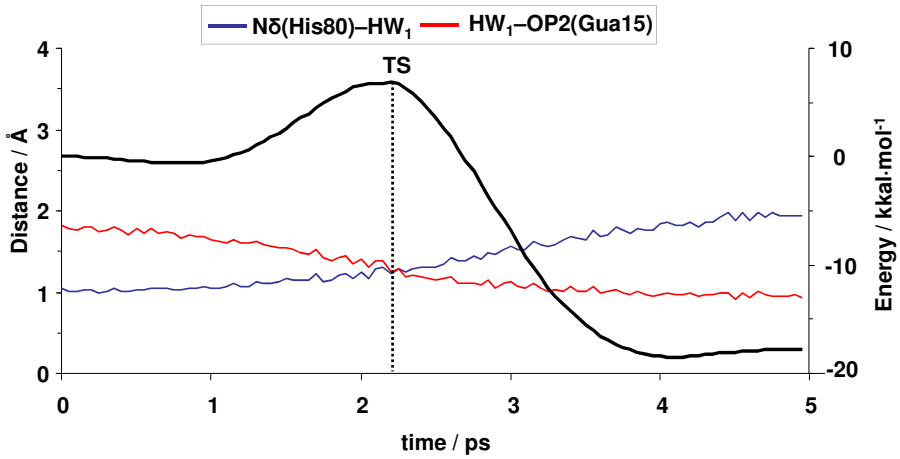


Figure S4. Energy profile of the proposed proton transfer from Nδ(His80^{+}) to OP2.

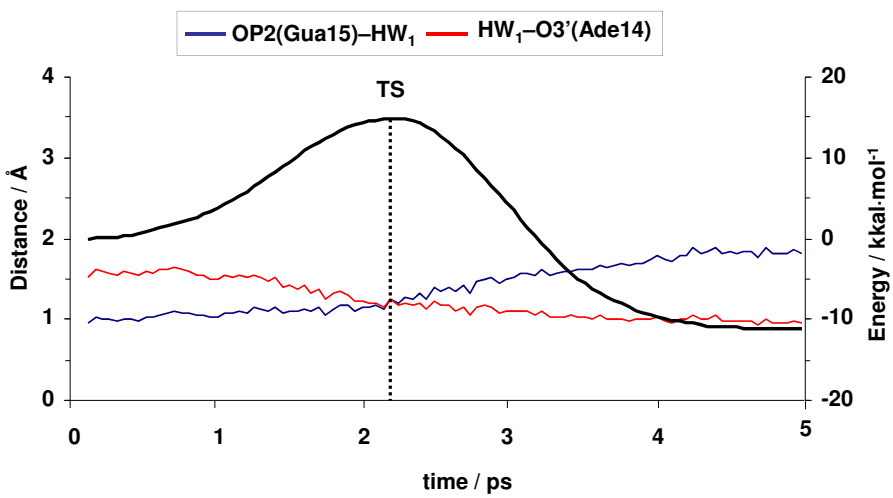


Figure S5. Energy profile of the proposed proton transfer from OP2 to O3'.

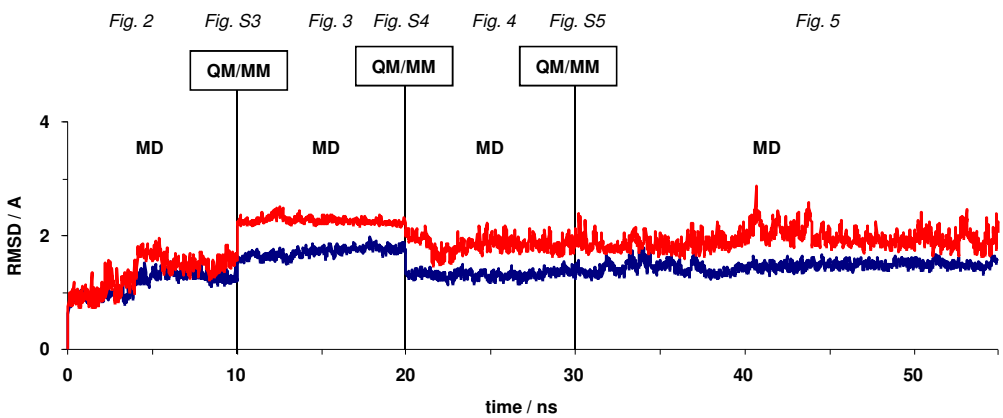


Figure S6. Time evolution of the root-mean-square deviation (RMSD) of protein Cα (blue) and DNA atoms (red, terminal basepairs excluded) from the initial modeled structure of the protein-DNA complex along the whole set of simulations. The QM/MM boxes indicate the times at which the hybrid method was used, and representative snapshots are shown in the figures indicated at the top.

Additional AMBER PREP and parameter files

AMBER PREP file for Deoxyadenine with O3' (-)

```
0    0    2

DAC
DAC.data
DAC INT 1
CORRECT OMIT DU BEG
0.00000
   1   DUMM  DU    M    0  -1  -2    0.0000   0.0000   0.0000     0.000000
   2   DUMM  DU    M    1   0  -1    1.7811   0.0000   0.0000     0.000000
   3   DUMM  DU    M    2   1   0    1.7798  59.9875 -90.0000     0.000000
   4   P     P     M    3   2   1    2.8766  73.9119 107.9319                1.04498
   5   O1P   O2    E    4   3   2    1.4980 132.8966 219.0422               -0.80519
   6   O2P   O2    E    4   3   2    1.5039  59.5185 114.4633               -0.80519
   7   O5'   OS    M    4   3   2    1.5956 119.7242  24.2876               -0.22399
   8   C5'   CT    M    7   4   3    1.3975 127.7034 250.6975                0.04717
   9   H5'1  H1    E    8   7   4    1.0896 109.8094  49.5871                0.02085
  10   H5'2  H1    E    8   7   4    1.0897 110.4215 285.3800                0.02085
  11   C4'   CT    M    8   7   4    1.5208 112.4541 167.0964                0.13722
  12   H4'   H1    E   11   8   7    1.0902 110.6990 160.0623                0.02819
  13   O4'   OS    S   11   8   7    1.3688 111.5316 283.9090               -0.40381
  14   C1'   CT    B   13  11   8    1.4239 103.4265 146.2173                0.10927
  15   H1'   H2    E   14  13  11    1.0905 116.8979 132.7737                0.08124
  16   N9    N*    S   14  13  11    1.5175 111.7769 251.8873               -0.17997
  17   C8    CK    B   16  14  13    1.3619 130.5709  32.0968                0.37879
  18   H8    H5    E   17  16  14    1.0799 119.7250 353.4804                0.08590
  19   N7    NB    S   17  16  14    1.2754 116.8419 177.7201               -0.69159
  20   C5    CB    S   19  17  16    1.3894 103.4200 356.2863                0.03800
  21   C6    CA    B   20  19  17    1.4556 134.5525 180.6452                0.66927
  22   N6    N2    B   21  20  19    1.2685 129.0068 358.5350               -0.87878
  23   H61   H     E   22  21  20    1.0098 123.1359 186.9327                0.39724
  24   H62   H     E   22  21  20    1.0099 115.3163   3.3941                0.39724
  25   N1    NC    S   21  20  19    1.3283 116.8111 180.4991               -0.78010
  26   C2    CQ    B   25  21  20    1.2957 119.2749 357.5679                0.55189
  27   H2    H5    E   26  25  21    1.0800 112.8722 185.3878                0.01943
  28   N3    NC    S   26  25  21    1.3460 129.8083   0.4220               -0.76494
  29   C4    CB    E   28  26  25    1.3143 110.9432   4.3047                0.45661
  30   C3'   CT    M   11   8   7    1.5474 112.9654  50.9812                0.38876
  31   H3'   H1    E   30  11   8    1.0905 113.1364 304.9388               -0.03141
  32   C2'   CT    B   30  11   8    1.5199  92.5517 190.4345               -0.16328
  33   H2'1  HC    E   32  30  11    1.0896 103.6693 170.5033                0.04962
  34   H2'2  HC    E   32  30  11    1.0904 112.4342 285.1287                0.04962
  35   O3'   OS    M   30  11   8    1.3615 110.2145  71.0783               -0.95681

IMPROPER
 C8   C4   N9   C1'
 C6   H61  N6   H62
 N7   N9   C8   H8
 N1   N3   C2   H2
 C5   N1   C6   N6

LOOP CLOSING EXPLICIT
 C1'  C2'
 C4   C5
 C4   N9

DONE
STOP
```

AMBER PREP file for Deoxyguanosine with $PO_3H^{(-)}$

```
0    0    2

DGC
DGC.data
DGC INT 1
CORRECT OMIT DU BEG
0.00000
   1   DUMM DU     M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2   DUMM DU     M   1   0  -1   1.7147   0.0000   0.0000   0.00000
   3   DUMM DU     M   2   1   0   1.7146  60.0192 -90.0000   0.00000
   4   P    P      M   1   2   3   8.6876 140.2860 289.5900   1.207059
   5   O1P  O2     E   4   1   2   1.5213 107.5112 163.6074  -0.839828
   6   O2P  O2     E   4   1   2   1.5348  94.7367  45.9562  -0.839828
   7   O3P  OH     S   4   1   2   1.5266 113.8217 289.2082  -0.643021
   8   H3O  HO     E   7   4   1   0.9599 109.5095 257.0438   0.418718
   9   O5'  OS     M   4   1   2   1.6304  11.2913 254.2238  -0.551044
  10   C5'  CT     M   9   4   1   1.4851 124.3732  38.0536   0.059414
  11   H5'1 H1     E  10   9   4   1.0901  97.2346 123.9555   0.080739
  12   H5'2 H1     E  10   9   4   1.0897 104.0440 352.7751   0.080739
  13   C4'  CT     M  10   9   4   1.4994 130.5334 229.9825   0.285823
  14   H4'  H1     E  13  10   9   1.0904 111.0754 353.3489   0.056689
  15   O4'  OS     S  13  10   9   1.4237 102.5262  99.9776  -0.310025
  16   C1'  CT     B  15  13  10   1.4006 106.4869 146.7101  -0.073420
  17   H1'  H2     E  16  15  13   1.0897 109.4277 120.2111   0.133929
  18   N9   N*     S  16  15  13   1.5243 113.9649 241.8752   0.068001
  19   C8   CK     B  18  16  15   1.3846 132.6795   3.6537   0.105503
  20   H8   H5     E  19  18  16   1.0807 120.7781   9.4751   0.211218
  21   N7   NB     S  19  18  16   1.3250 114.7104 179.3168  -0.575350
  22   C5   CB     S  21  19  18   1.4175 102.6578   0.9649   0.108815
  23   C6   C      B  22  21  19   1.4201 131.4558 172.9435   0.532472
  24   O6   O      E  23  22  21   1.2314 125.5322   9.3698  -0.609208
  25   N1   NA     B  23  22  21   1.3667 114.6581 187.0651  -0.470229
  26   H1   H      E  25  23  22   1.0095 117.2186 182.0397   0.311318
  27   C2   CA     B  25  23  22   1.3118 123.8561 356.3380   0.767491
  28   N2   N2     B  27  25  23   1.3065 117.2897 196.7158  -0.984023
  29   H21  H      E  28  27  25   1.0095 118.0074 354.4198   0.427853
  30   H22  H      E  28  27  25   1.0097 123.1304 159.6867   0.427853
  31   N3   NC     S  27  25  23   1.3358 125.3174  13.2727  -0.634522
  32   C4   CB     E  22  21  19   1.3613 110.6402   4.8004   0.177756
  33   C3'  CT     M  13  10   9   1.5247 117.7052 221.3521   0.055001
  34   H3'  H1     E  33  13  10   1.0900 107.6580 327.2738   0.092067
  35   C2'  CT     B  33  13  10   1.4871 108.2856   4.9862  -0.133683
  36   H2'1 HC     E  35  33  13   1.0905 112.0917 216.5148   0.097925
  37   H2'2 HC     E  35  33  13   1.0905 114.0808  92.3737   0.097925
  38   O3'  OS     M  33  13  10   1.3866 113.2052  89.7513  -0.480375


IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 C2   H21  N2   H22
 N7   N9   C8   H8
 N1   N3   C2   N2

LOOP CLOSING EXPLICIT
 C1'  C2'
 C4   C5
 C4   N9

DONE
STOP
```

AMBER PREP file for Deoxyguanosine with PO_3H_2

```
0    0    2

DGP
DGP.data
DGP INT 1
CORRECT OMIT DU BEG
0.00000
   1   DU   DU     M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2   DU   DU     M   1   0  -1   1.7148   0.0000   0.0000   0.00000
   3   DU   DU     M   2   1   0   1.7142  60.0263 -90.0000   0.00000
   4   O1P  O2     M   1   2   3   4.0813 122.0506  64.9488  -0.881468
   5   P    P      M   4   1   2   1.5100  44.3971  67.0882   1.668473
   6   O3P  OH     S   5   4   1   1.5759 101.1264 163.8170  -0.699679
   7   H3O  HO     E   6   5   4   0.9566 108.7221  55.2564   0.468245
   8   O2P  OH     S   5   4   1   1.4876 118.7039  40.1964  -0.699679
   9   H4O  HO     E   8   5   4   0.9601 109.5135  98.5128   0.468245
  10   O5'  OS     M   5   4   1   1.6538 108.9483 268.6385  -0.533943
  11   C5'  CT     M  10   5   4   1.4329 119.5530 206.9304   0.145860
  12   H5'1 H1     E  11  10   5   1.0907 107.7971  73.8953   0.066273
  13   H5'2 H1     E  11  10   5   1.0901 106.1532 302.1521   0.066273
  14   C4'  CT     M  11  10   5   1.5796 109.3042 186.0621   0.511890
  15   H4'  H1     E  14  11  10   1.0904 106.3072 320.3178   0.035717
  16   O4'  OS     S  14  11  10   1.4932 110.2038  77.7843  -0.430655
  17   C1'  CT     B  16  14  11   1.4008 104.7104 156.8073  -0.006115
  18   H1'  H2     E  17  16  14   1.0897 108.8112  88.2330   0.125871
  19   N9   N*     S  17  16  14   1.5275 108.8223 214.6232   0.055031
  20   C8   CK     B  19  17  16   1.3779 129.4000   4.3191   0.068033
  21   H8   H5     E  20  19  17   1.0799 120.8271 352.7295   0.192901
  22   N7   NB     S  20  19  17   1.3406 115.4840 171.4810  -0.571344
  23   C5   CB     S  22  20  19   1.4038 101.1943   1.3561   0.186617
  24   C6   C      B  23  22  20   1.4423 129.1174 178.6452   0.499104
  25   O6   O      E  24  23  22   1.2690 126.3075   4.2315  -0.581842
  26   N1   NA     B  24  23  22   1.4320 115.5668 180.6107  -0.555091
  27   H1   H      E  26  24  23   1.0099 117.9577 175.7939   0.361123
  28   C2   CA     B  26  24  23   1.4370 119.3491   9.8112   0.712628
  29   N2   N2     B  28  26  24   1.2891 122.6782 177.0022  -0.830521
  30   H21  H      E  29  28  26   1.0108 118.0591 346.4717   0.429368
  31   H22  H      E  29  28  26   1.0095 114.6866 196.9871   0.429368
  32   N3   NC     S  28  26  24   1.3421 120.8992 358.1603  -0.644677
  33   C4   CB     E  32  28  26   1.3347 118.4994 354.0624   0.192310
  34   C3'  CT     M  14  11  10   1.5775 110.1973 196.3531   0.053605
  35   H3'  H1     E  34  14  11   1.0899 106.8135 326.8879   0.122825
  36   C2'  CT     B  34  14  11   1.5030  97.8170 208.2813  -0.015635
  37   H2'1 HC     E  36  34  14   1.0904 108.1341 136.1652   0.045979
  38   H2'2 HC     E  36  34  14   1.0901 113.7496 252.5842   0.045979
  39   O3'  OS     M  34  14  11   1.4294 112.8419  90.0268  -0.492321

IMPROPER
 C8   C4   N9   C1'
 C5   N1   C6   O6
 C6   C2   N1   H1
 C2   H21  N2   H22
 N7   N9   C8   H8
 N1   N3   C2   N2

LOOP CLOSING EXPLICIT
 C1'  C2'
 C4   C5
 C4   N9

DONE
STOP
```

Modifications to the parmbsc0 force field for DGP

```
ANGLE
OH-P -OH   200.0      108.23    PO3H2
```

AMBER PREP file for Mg^{2+}

```
0   0   2

COUNTERION (+2)
MG.data
MG  INT     1
CORRECT OMIT DU  BEG
   0.00000
   1   DU   DU    M   0  -1  -2   0.0000   0.0000   0.0000   0.00000
   2   DU   DU    M   1   0  -1   3.7992   0.0000   0.0000   0.00000
   3   DU   DU    M   2   1   0   2.5176 144.8384 -90.0000   0.00000
   4   MG   MG    M   3   2   1   2.7389 144.7882 179.8838   2.00000

DONE
STOP
```

MovieM1. Water splitting, hydroxide attack on the phosphodiester bond, and intermediate product formation (Bueren-et-al_movie1.wmv, 20 MB).

MovieM2. First non-trivial normal mode of Vvn showing the clasping motion that involves two domains on both sides of the DNA-binding cleft. The DNA cartoon is displayed to facilitate understanding but was not included in the NMA calculation. The side-chains of some representative residues are displayed as sticks (Bueren-et-al_movie2.wmv, 4.5 MB)

4. CONCLUSIONES

"*Now this is not the end. It is not even the beginning of the end. But it is perhaps, the end of the beginning.*" — Winston Churchill

1. ***Molecular pharmacology and antitumor activity of Zalypsis in several human cancer cell lines.***

a) Zalypsis pertenece a una familia de compuestos que se unen al grupo amino exocíclico de guaninas en la doble hélice del ADN a través del iminio generado in situ por deshidratación de la carbinolamina presente en el anillo A.

b) El aducto resultante se estabiliza adicionalmente a través de interacciones de van der Waals y el establecimiento de uno o más enlaces de hidrógeno con los nucleótidos adyacentes.

c) Estos enlaces de hidrógeno parecen determinar la especificidad de secuencia de la unión de esta familia de fármacos.

d) Los modelos moleculares equilibrados de los complejos precovalentes de Zalypsis-ADN en solución acuosa aportan una explicación racional para las diferencias en la unión del fármaco a los tres tripletes centrales principales que representan sitios de unión malos (AGA), intermedios (AGG) y buenos (AGC).

2. ***PM01183, a new DNA minor groove covalent binder with potent in vitro and in vivo anti-tumour activity***

a) Los tripletes de ADN preferidos para la unión de triptamicidina (PM01183), son AGC, CGG, AGG y TGG.

b) Los detalles moleculares de dichas preferencias de unión pueden ser racionalizadas mediante modelado molecular demostrando que un patrón específico de enlaces de hidrógeno es necesario tanto para el reconocimiento de la secuencia como para la activación del fármaco.

c) El complejo covalente formado por un oligonucleótido representativo con un triplete central CGG apenas mostró tener distorsiones estructurales, siendo la más significativa siendo un ensanchamiento del surco menor en la región donde el fármaco está unido, así como una curvatura de la molécula de ADN hacia el surco mayor.

d) La guanina unida al fármaco establece interacciones no enlazantes (van der Waals y electrostáticas, incluyendo enlaces de hidrógeno) con los nucleótidos adyacentes de ambas hebras.

e) Una probable consecuencia, apoyada por los experimentos de fusión (*melting*), es que uno o más aductos de triptamicidina-guanina pueden dificultar o prevenir la separación de las hebras y, por lo tanto, bloquear las horquillas de transcripción y replicación.

3. *XPF-dependent DNA breaks and RNA polymerase II arrest induced by antitumor DNA interstrand crosslinking-mimetic tetrahydroisoquinoline alkaloids*

a) Zalypsis y Yondelis favorecen el corte de ADN al que están unidos covalentemente por la endonucleasa por XPF e inhiben la transcripción por la RNA polimerasa II. Los resultados de modelado molecular son capaces de proponer una explicación a estos fenómenos en términos atomísticos.

b) Los resultados de dinámica molecular mostraron que la fusión del oligonucleótido diana en el triplete central CGG y en las bases vecinas está impedida por la presencia de los fármacos unidos a la guanina.

c) Estas observaciones implican que Yondelis y Zalypsis contribuyen a extender la longitud de la región central de doble cadena del ADN y, por lo tanto, favorecen el reconocimiento por XPF y su unión.

d) Desde un punto de vista estructural, hemos aportado una explicación con un nivel de detalle atomístico de las diferencias en el bloqueo de la síntesis por la RNA polimerasa II basándonos en los cristales disponibles de los complejos de elongación de la RNA polimerasa II.

4. *Temperature-induced melting of double-stranded DNA in the absence and presence of covalently bonded antitumour drugs: insight from molecular dynamics simulations*

a) La fusión del ADN en ausencia y presencia de fármacos unidos puede aportar información acerca de la estabilización debida a la unión de estos fármacos.

b) Empleando simulaciones a una temperatura elevada junto con altas concentraciones salinas, fuimos capaces de detectar una desnaturalización completa de un oligodesoxinucleótido en ausencia de fármacos en menos de 200 ns.

c) Por el contrario, después de 200 ns de simulación por dinámica molecular, una región de doble cadena formada por pares de bases de Watson-Crick, incluyendo el triplete central CGG, se mantuvo en todos los complejos fármaco-ADN.

d) Nuestros resultados muestran un punto de vista atomístico del proceso de fusión de ADN en ausencia y presencia de diversos fármacos antitumorales y confirman la gran estabilización que los monodaductos de Yondelis®, Zalypsis® y PM01183 han demostrado tener sobre la doble hélice de ADN.

e) Por el contrario, mitomicina C, el principal representantede los agentes capaces de formar entrecruzamientos intercatenarios en el ADN, evita la separación de las hebras debido a que forma un enlace covalente con cada cadena del ADN, pero su efecto en la estabilización en el dúplex no parece extenderse más allá del paso CpG al cual está covalentemente unida.

5. ***Rationale for the opposite stereochemistry of the major monoadducts and interstrand crosslinks formed by mitomycin C and its decarbamoylated analogue at CpG steps in DNA and the effect of cytosine modification on reactivity***

a) Al simular cada paso de la reacción usando una combinación de MD y QM/MM hemos sido capaces de obtener información con un nivel de detalle atomístico sobre los fenómenos de reconocimiento previos a la formación del monoaducto así como sobre los requerimientos estructurales y electrónicos que posiblemente determinan la mayor reactividad de MMC en los pasos CpG cuando las citosinas están metiladas.

b) Aportamos además una posible explicación al origen de la formación de monoaductos C1α or C1β dependiendo de los puentes de hidrógeno establecidos en los complejos precovalentes ADN-MMC^+ y ADN-DMC^+.

c) La estructura helicoidal del ADN en el paso central de unión al fármaco parece estar preorganizada para la unión de MMC^+ y además presenta un MEP más negativo cunado las citosinas están metiladas.

d) Nuestras simulaciones de MD sin restricciones reproducen fielmente las interacciones fármaco-ADN previamente descritas en las soluciones de RMN del monoaducto y aportan más detalles sobre i) el complejo precovalente, ii) la reacción del monoaducto activado con una segunda guanina de la hebra complementaria, y iii) el ICL final.

e) Igualmente aportamos información estructural sobre el α-ICL y el β-ICL. Éste último complejo no había sido descrito hasta ahora: mientras que el primero apenas modifica la doble hélice, el segundo provoca una gran distorsión en el paso central que puede tener consecuencias en cuanto a la reparación mediada por NER y HR.

6. ***Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: Vibrio vulnificus periplasmic nuclease as a test case***

a) A partir de un complejo *Vvn*-ADN que se construyó usando información cristalográfica de las estructuras de los complejos de *Vvn* de tipo silvestre y mutante y de la nucleasa I-PpoI con diferentes oligonucleótidos, se llevó a cabo una serie de simulaciones para analizar el mecanismo catalítico de *Vvn* de forma escalonada.

b) Los resultados obtenidos nos han permitido caracterizar los papeles que desempeña cada residuo del sitio activo en el transcurso de la reacción y han respondido a algunas dudas que quedaban sin resolver. En particular, se demuestra que el mecanism4o más probable es aquel en el que una sola molécula de agua es la responsable de hidrolizar el enlace fosofodiéster y protonar el grupo hidroxilo saliente.

c) El mecanismo catalítico que hemos descrito para *Vvn* es probablemente el mismo en otros miembros de la familia de nucleasas ββα-metal que se basan en la activación de una molécula de agua por el anillo de imidazol de una histidina correctamente posicionada para hidrolizar el ADN y el ARN.

5. BIBLIOGRAFÍA

1. Hurley, L.H. and Boyd, F.L. (1988) DNA as a target for drug action. *Trends in pharmacological sciences*, **9**, 402-407.
2. Hurley, L.H. (1989) DNA and associated targets for drug design. *Journal of medicinal chemistry*, **32**, 2027-2033.
3. Hurley, L.H. (2002) DNA and its associated processes as targets for cancer therapy. *Nature reviews*, **2**, 188-200.
4. Karplus, M. and McCammon, J.A. (2002) Molecular dynamics simulations of biomolecules. *Nature structural biology*, **9**, 646-652.
5. Waring, M. (2006) Sequence-specific DNA Binding agents. *RSC Publishing*.
6. Pérez, A. (2008) Estudio de mecansimos de interacción macromolecular. *Universidad de Barcelona*.
7. Leach, A.R. (2001) Molecular Modelling. *Person Prentice Hall*.
8. Singh, U.C. and Kollman, P.A. (1984) An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, **5**, 129-145.
9. Bayly, C.I., Cieplak, P., Cornell, W. and Kollman, P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, **97**, 10269-10280.
10. Cornell, W.D., Cieplak, P., Bayly, C.I. and Kollmann, P.A. (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society*, **115**, 9620-9631.

11. 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. and Kollman,P.A. . (1995) A second-generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
12. Verlet, L. (1967) Ground State of Liquid Helium-4 and Helium-3. *Phys Rev*, **160**, 208-218.
13. Ryckaert, J.P., Ciccoti,G. and Berendsen,H.J.C. . (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys*, **23**, 327-341.
14. Woodcock, L.V. (1971) Isothermal Molecular Dynamics Calculations for Liquid Salts. *Chemical Physics Letters*, **10**, 257-261.
15. H J C Berendsen, J.P.M.P., W F Van Gunsteren, A Dinola, J R Haak. (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690.
16. Perez, A. and Orozco, M. (2010) Real-time atomistic description of DNA unfolding. *Angewandte Chemie (International ed*, **49**, 4805-4808.
17. Onufriev A., B.D., Case D. A.,. (2000) Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem.*, **104**, 3712.
18. Sitkoff, D., Sharp, K.A. and Honig, B. (1994) Correlating solvation free energies and surface tensions of hydrocarbon solutes. *Biophysical chemistry*, **51**, 397-403; discussion 404-399.
19. Seabra, M., Walker, R.C., Elstner, M., Case, D.A. and Roitberg, A.E. (2007) Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the amber molecular dynamics package. *J Phys Chem A*, **111**, 5655-5664.

20. Spiegel, K. and Magistrato, A. (2006) Modeling anticancer drug-DNA interactions via mixed QM/MM molecular dynamics simulations. *Organic & biomolecular chemistry*, **4**, 2507-2517.

21. Mulholland, A.J. (2007) Chemical accuracy in QM/MM calculations on enzyme-catalysed reactions. *Chemistry Central journal*, **1**, 19.

22. Levitt, A.W.a.M. (1976) Theoretical Studies of Enzymic reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *Journal of Molecular Biology*, **103**, 227-249.

23. Hershey, A.D. and Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*, **36**, 39-56.

24. Hershey, A.D., Dixon, J. and Chase, M. (1953) Nucleic acid economy in bacteria infected with bacteriophage T2. I. Purine and pyrimidine composition. *The Journal of general physiology*, **36**, 777-789.

25. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.

26. Calladine C.R & Drew, H.R. (1992) Understanding DNA. *Academic Press Limited, San Diego*.

27. Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, **6**, 201-209.

28. Wong, K.Y. and Pettitt, B.M. (2008) The pathway of oligomeric DNA melting investigated by molecular dynamics simulations. *Biophysical journal*, **95**, 5618-5626.

29. Lescrinier, E., Froeyen, M. and Herdewijn, P. (2003) Difference in conformational diversity between nucleic acids with a six-membered 'sugar' unit and natural 'furanose' nucleic acids. *Nucleic acids research*, **31**, 2975-2989.

30. Negri, A., Rodriguez-Larrea, D., Marco, E., Jimenez-Ruiz, A., Sanchez-Ruiz, J.M. and Gago, F. (2008) Protein-protein interactions at an enzyme-substrate interface: characterization of transient reaction intermediates throughout a full catalytic cycle of Escherichia coli thioredoxin reductase. *Proteins*, **78**, 36-51.

31. Marco, E. and Gago, F. (2007) Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. *ChemMedChem*, **2**, 1388-1401.

32. Chothia, C. and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *The EMBO journal*, **5**, 823-826.

33. Chothia, C., Lesk, A.M., Levitt, M., Amit, A.G., Mariuzza, R.A., Phillips, S.E. and Poljak, R.J. (1986) The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science (New York, N.Y*, **233**, 755-758.

34. Lesk, A.M., Levitt, M. and Chothia, C. (1986) Alignment of the amino acid sequences of distantly related proteins using variable gap penalties. *Protein engineering*, **1**, 77-78.

35. Sander, C. and Schneider, R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, **9**, 56-68.

36. Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J Mol Biol*, **147**, 195-197.

37. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, **22**, 4673-4680.

38. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, **302**, 205-217.

39. Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics (Oxford, England)*, **14**, 892-893.

40. Fiser, A. and Sali, A. (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics (Oxford, England)*, **19**, 2500-2501.

41. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *Journal of molecular graphics*, **8**, 52-56, 29.

42. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check stereochemical quality of protein structures *J. Appl. Cryst.*, **26**, 283-291.

43. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)*, **22**, 195-201.

44. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, **234**, 779-815.

45. Fiser, A., Do, R.K. and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci*, **9**, 1753-1773.

46. Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, **4**, 363-371.

47. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, **9**, 40.

48. Gooford, P. (1995) “GRID Molecular Discovery Ltd. *Oxford, England,*.

49. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic acids research*, **34**, W116-118.

50. Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*, **47**, 1739-1749.

51. Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T. and Banks, J.L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry*, **47**, 1750-1759.

52. Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., Sanschagrin, P.C. and Mainz, D.T. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of medicinal chemistry*, **49**, 6177-6196.

53. http://www.pharmamar.com/inicio.aspx.

54. Casado, J.A., Rio, P., Marco, E., Garcia-Hernandez, V., Domingo, A., Perez, L., Tercero, J.C., Vaquero, J.J., Albella, B., Gago, F. *et al.* (2008) Relevance of the Fanconi anemia pathway in the response of human cells to trabectedin. *Molecular cancer therapeutics*, **7**, 1309-1318.

55. Tomasz, M., Chowdary, D., Lipman, R., Shimotakahara, S., Veiro, D., Walker, V. and Verdine, G.L. (1986) Reaction of DNA with chemically or enzymatically activated mitomycin C: isolation and structure of the major covalent adduct. *Proc. Natl. Acad. Sci. U. S. A.*, **83**, 6702-6706.

56. Tomasz, M., Lipman, R., Chowdary, D., Pawlak, J., Verdine, G.L. and Nakanishi, K. (1987) Isolation and structure of a covalent cross-link adduct between mitomycin C and DNA. *Science*, **235**, 1204-1208.

57. Tomasz, M. and Palom, Y. (1997) The mitomycin bioreductive antitumor agents: cross-linking and alkylation of DNA as the molecular basis of their activity. *Pharmacol. Ther.*, **76**, 73-87.

58. Leal, J.F., Garcia-Hernandez, V., Moneo, V., Domingo, A., Bueren-Calabuig, J.A., Negri, A., Gago, F., Guillen-Navarro, M.J., Aviles, P., Cuevas, C. *et al.* (2009) Molecular pharmacology and antitumor activity of Zalypsis in several human cancer cell lines. *Biochemical pharmacology*, **78**, 162-170.

59. Leal, J.F., Martinez-Diez, M., Garcia-Hernandez, V., Moneo, V., Domingo, A., Bueren-Calabuig, J.A., Negri, A., Gago, F., Guillen-Navarro, M.J., Aviles, P. *et al.* (2010) PM01183, a new DNA minor groove covalent binder with potent in vitro and in vivo anti-tumour activity. *British journal of pharmacology*, **161**, 1099-1110.

60. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. and Sharma, R.A. (2008) DNA repair pathways as targets for cancer therapy. *Nature reviews*, **8**, 193-204.

61. Andreassen, P.R. and Ren, K. (2009) Fanconi anemia proteins, DNA interstrand crosslink repair pathways, and cancer therapy. *Current cancer drug targets*, **9**, 101-117.

62. Guirouilh-Barbat, J., Antony, S. and Pommier, Y. (2009) Zalypsis (PM00104) is a potent inducer of gamma-H2AX foci and reveals the importance of the C ring of trabectedin for transcription-coupled repair inhibition. *Molecular cancer therapeutics*, **8**, 2007-2014.

63. Cervenka, J., Arthur, D. and Yasis, C. (1981) Mitomycin C test for diagnostic differentiation of idiopathic aplastic anemia and Fanconi anemia. *Pediatrics*, **67**, 119-127.

64. Damia, G., Silvestri, S., Carrassa, L., Filiberti, L., Faircloth, G.T., Liberi, G., Foiani, M. and D'Incalci, M. (2001) Unique pattern of ET-743 activity in different cellular systems with defined deficiencies in DNA-repair pathways. *International journal of cancer*, **92**, 583-588.

65. Tsodikov, O.V., Enzlin, J.H., Scharer, O.D. and Ellenberger, T. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 11236-11241.

66. Tripsianes, K., Folkers, G., Ab, E., Das, D., Odijk, H., Jaspers, N.G., Hoeijmakers, J.H., Kaptein, R. and Boelens, R. (2005) The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure*, **13**, 1849-1858.

67. Feuerhahn, S., Giraudon, C., Martinez-Diez, M., Bueren-Calabuig, J.A., Galmarini, C.M., Gago, F. and Egly, J.M. XPF-Dependent DNA Breaks and RNA Polymerase II Arrest Induced by Antitumor DNA Interstrand Crosslinking-Mimetic Alkaloids. *Chemistry & biology*, **18**, 988-999.

68. Mulholland, A.J. (2005) Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug discovery today*, **10**, 1393-1402.