



Facultad de Química

Departamento de Química Orgánica

**Centro Singular de Investigación en Química Biológica y
Materiales Moleculares (CIQUS)**

**Modulación de la interacción de
péptidos derivados de factores de
transcripción con secuencias
específicas de ADN**

Jesús Mosquera Mosquera

Diciembre 2014



D. JOSÉ LUIS MASCAREÑAS CID, CATEDRÁTICO DEL DEPARTAMENTO DE QUÍMICA ORGÁNICA DE LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA Y D. MARCO EUGENIO VÁZQUEZ SENTÍS, PROFESOR TITULAR DEL DEPARTAMENTO DE QUÍMICA ORGÁNICA DE LA UNIVERSIDAD DE SANTIAGO,

CERTIFICAN: Que la memoria adjunta titulada, "*Modulación de la interacción de péptidos derivados de factores de transcripción con secuencias específicas de ADN*", que para optar al grado de Doctor en Química presenta Don Jesús Mosquera Mosquera, ha sido realizada bajo nuestra dirección, en los laboratorios del departamento de Química Orgánica y del Centro Singular de Investigación en Química Biológica y Materiales Moleculares (CIQUS) de la USC.

Considerando que constituye trabajo de tesis, autorizamos su presentación en la Universidad de Santiago de Compostela.

Y para que así conste, expiden el presente certificado en Santiago de Compostela, a 30 de Septiembre de 2014.

Fdo.: José Luis Mascareñas Cid

Fdo.: M. Eugenio Vázquez Sentís



A mi esposa, mis padres, mi hermana y mi madrina





ÍNDICE

| | |
|---|-----------|
| PRESENTACIÓN..... | 1 |
| ABREVIATURAS Y ACRÓNIMOS | 3 |
| LA QUÍMICA BIOLÓGICA..... | 11 |
| INTRODUCCIÓN..... | 15 |
| 1. CONCEPTOS BÁSICOS: ADN Y PROTEÍNAS | 17 |
| 1.1. Función biológica del ADN..... | 19 |
| 1.2. Las proteínas..... | 20 |
| 1.2.1. La hélice α | 22 |
| 1.2.2. La lámina β | 23 |
| 2. LA EXPRESIÓN GÉNICA | 25 |
| 2.1. Etapas generales del proceso de expresión génica | 25 |
| 2.2. La transcripción..... | 25 |
| 2.3. La traducción | 27 |
| 3. LOS FACTORES DE TRANSCRIPCIÓN | 29 |
| 3.1. Bases generales del reconocimiento molecular entre FT y proteínas..... | 29 |
| 3.2. Familias de factores de transcripción | 31 |
| 3.2.1. Hélice-Giro-Hélice (HTH: Helix-turn-Helix) y Homeodominio..... | 31 |
| 3.2.2. Proteínas de Cremallera de Leucinas (bZIP) y Hélice-Lazo-Hélice (HLH: Helix-Loop-Helix)..... | 32 |
| 3.2.3. Dedos de Zinc (Zinc Finger Protein)..... | 34 |
| 4. REGULACIÓN DE LA EXPRESIÓN GÉNICA..... | 35 |
| 4.1. Mecanismos naturales de regulación génica..... | 35 |
| 4.2. Regulación de la actividad de los factores de transcripción | 36 |
| 5. INTERACCION DE MOLÉCULAS PEQUEÑAS CON EL ADN | 39 |
| 5.1. Agentes que se unen mediante contactos electrostáticos..... | 39 |
| 5.2. Agentes intercalantes..... | 40 |
| 5.3. Agentes que se unen de forma covalente al ADN..... | 41 |
| 5.4. Moléculas que se insertan en el surco menor..... | 42 |
| 6. RECONOCIMIENTO DEL ADN MEDIANTE PÉPTIDOS SINTÉTICOS..... | 47 |
| 6.1. Dimerización artificial de proteínas bZIP..... | 47 |
| 6.2. Motivos helicoidales estabilizados..... | 50 |
| 6.2.1. Uso de estructuras plantilla preorganizadas (grafting de residuos)..... | 50 |
| 6.2.2. Restricción conformacional (péptidos grapados) | 51 |
| 6.2.3. Estabilización mediante aminoácidos inductores de helicidad..... | 51 |

| | |
|---|------------|
| 6.3. Conjugados entre fragmentos de FTs y moléculas que se unen al surco menor del ADN..... | 52 |
| 6.4. Conjugados entre intercalantes y fragmentos de factores de transcripción..... | 54 |
| 6.5. Estudio de la interacción entre péptidos y el ADN | 55 |
| 7. PÉPTIDOS CUYA INTERACCIÓN CON EL ADN PUEDE MODULARSE..... | 59 |
| 7.1. Derivados conmutables mediante el uso de luz..... | 59 |
| 7.1.1. Grupos fotoisomerizables..... | 59 |
| 7.1.2. Grupos protectores fotolábiles..... | 60 |
| 7.2. Derivados conmutables por la acción de metales..... | 63 |
| OBJETIVOS | 65 |
| CAPÍTULO 1..... | 71 |
| CAPÍTULO 2..... | 77 |
| CAPÍTULO 3..... | 83 |
| CAPÍTULO 4..... | 91 |
| CONCLUSIONS | 101 |
| RESUMEN..... | 105 |
| SUMMARY..... | 113 |
| PARTE EXPERIMENTAL..... | 121 |
| <i>PARTE EXPERIMENTAL CAPÍTULO 1.....</i> | 123 |
| <i>PARTE EXPERIMENTAL CAPÍTULO 2.....</i> | 133 |
| <i>PARTE EXPERIMENTAL CAPÍTULO 3.....</i> | 151 |
| <i>PARTE EXPERIMENTAL CAPÍTULO 4.....</i> | 165 |

PRESENTACIÓN

Esta tesis se presenta por compendio de artículos de investigación aceptados en revistas de ámbito y prestigio internacional con revisión anónima por pares, de acuerdo con el artículo 41 del *Reglamento de los estudios de doctorado de la Universidad de Santiago de Compostela*.

De acuerdo con este reglamento sólo se incluyen aquellos artículos publicados, llevados a cabo durante los estudios de doctorado y que no se incluyeron como parte de otras tesis.





ABREVIATURAS Y ACRÓNIMOS





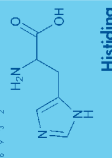





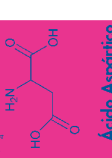
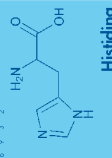





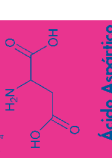
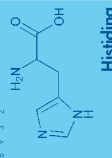





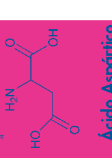
| | |
|-----------------------|--|
| δ | Desplazamiento químico (ppm) |
| ϵ | Coefficiente de extinción |
| μ | Momento dipolar |
| μL | Microlitro |
| μM | Micromolar |
| θ_{MRE} | Elipticidad molar por residuo |
| A | Adenina |
| Aba | Ácido p-acetamidobenzoico |
| Ac | Acetilo |
| Ac_2O | Anhídrido acético |
| AcOH | Ácido acético |
| AcOEt | Acetato de etilo |
| Aib | 2-aminoisobutírico |
| Alloc | Aliloxicarbonil |
| aPP | Péptido pancreático de ave |
| ADN | ácido desoxirribonucleico |
| APN | ácido peptidonucleico |
| ARN | ácido ribonucleico |
| ARNm | Ácido ribonucleico mensajero |
| ARNpol | ARN polimerasa |
| ARNr | Ácido ribonucleico ribosomal |
| ARNt | Ácido ribonucleico de transferencia |
| ATP | Adenosíntrifosfato |
| a. u. | unidades arbitrarias |
| BAPPA | N^1, N^3 -bis(4-amidinofenil)propano-1,3-diamina |
| Boc | tert-Butiloxicarbonilo |
| BSA | Seroalbúmina bovina |
| tBu | tert-Butilo |

| | |
|-------------------|---|
| C | Citosina |
| calcd. | Calculada |
| CDCl ₃ | Cloroformo deuterado |
| d | Doblete |
| DC | Dicroísmo circular |
| DEDTC | Dietilditiocarbamato |
| DIEA | N,N-Diisopropiletilamina |
| DMAP | N,N-Dimetilaminopiridina |
| DMF | N,N-Dimetilformamida |
| DMSO | Dimetilsulfóxido |
| DTNB | Ácido 5, 5'-ditiobis-2-nitrobenzónico |
| DTT | 1, 4-Ditiotreitol |
| EDT | 1, 2-Etanoditiol |
| AEDT | Ácido etilendiaminotetraacético |
| EMSA | Ensayos de desplazamiento en gel (Electrophoretic mobility shift assay) |
| equiv | Equivalente |
| ESI-MS | Espectrometría de masas mediante ionización por electroespray |
| Et | Etilo |
| Et ₃ N | Trietilamina |
| Et ₂ O | Éter dietílico |
| EtOH | Etanol |
| Fmoc | 9-Fluorenilmetoxicarbonilo |
| FR | Fase reversa |
| FT | Factor de transcripción |
| FTG | Factor de transcripción general o basal |
| G | Guanina |
| GPF | Grupos protectores fotolábiles |

| | |
|----------|--|
| GR | Receptor glucocorticoides |
| HATU | 2-(1H-7-Azabenzotriazol-1-il)-1,1,3,3-tetrametiluronio hexafluorofosfato |
| HBTU | 2-(1H-Benzotriazol-1-il)-1,1,3,3-tetrametiluronio hexafluorofosfato |
| HD | homeodominio |
| HOBt | 1-Hidroxibenzotriazol |
| HPLC-MS | Espectrometría de masas y cromatografía líquida de alta resolución |
| HLH | Familia de factores de transcripción Hélice-lazo-hélice |
| HTH | Familia de factores de transcripción Hélice-giro-hélice |
| IE | Impacto electrónico |
| Im | Imidazol |
| J | Constante de acoplamiento |
| Kd | Constante de disociación |
| LC/MS | Cromatografía líquida / Espectrometría de masas |
| LED | Diodo emisor de luz |
| M | Molar |
| m | Multiplete |
| MALDI-MS | Ionización mediante desorción por láser asistida mediante matriz |
| Me | Metilo |
| MeOH | Metanol |
| mg | Miligramo |
| MHz | Megahertzios |
| mL | Mililitro |
| mM | Milimolar |
| mmol | Milimol |
| m/z | Masa/carga |
| NAP | nitroacetofenona |

| | |
|-------|--|
| nm | Nanómetro |
| nM | Nanomolar |
| NP-40 | Octiletoxipolietoxietanol |
| PAGE | Electroforesis en gel de poliacrilamida (PoliAcrilamide gel electrophoresis) |
| PDB | protein data bank |
| Pbf | 2,2,4,6,7-Pentametildihidrobenzofuran-5-sulfonilo |
| Ph | Fenilo |
| Phi | 9,10-fenantrenoquinona diamina |
| Phen | 1,10-fenantrolina |
| ppm | Partes por millón |
| Py | Pirrol |
| Rho | rodamina |
| RMN | Resonancia magnética nuclear |
| s | Singlete |
| SPPS | Síntesis de péptidos sobre soporte sólido |
| t | Triplete |
| T | Timina |
| TAMRA | Tetrametilrodamina |
| TBE | Disolución de TRIS-borato-EDTA |
| TCEP | Triscarboxietilfosfina |
| TFA | Ácido trifluoroacético |
| TIS | Triisopropilsilano |
| TO | Thiazole orange |
| Tpy | Terpiridina |
| TRIS | Tris(hidroximetil)aminometano |
| Trt | Trifenilmetilo (tritilo) |
| UV | Ultravioleta |

Tabla periódica de los aminoácidos

| | | | | | | |
|--|--|---|--|--|--|---|
| H 155.16 137.14 $C_9H_9N_3O_2$  His Histidina | F 165.19 147.18 $C_9H_9NO_2$  Phe Fenilalanina | A 89.09 71.08 $C_6H_9NO_2$  Ala Alanina | C 121.16 103.14 $C_3H_7NO_2S$  Cys Cisteína | G 75.07 57.05 $C_2H_5NO_2$  Gly Glicina | G 146.15 128.13 $C_5H_9NO_3$  Gln Glutamina | D 133.10 115.09 $C_4H_7NO_4$  Asp Ácido Aspártico |
| R 174.20 156.19 $C_6H_{12}N_4O_2$  Arg Arginina | L 131.18 113.16 $C_6H_{11}NO_2$  Leu Leucina | M 149.21 131.20 $C_5H_9NO_2S$  Met Metionina | N 132.12 114.10 $C_4H_7NO_3$  Asn Asparagina | S 105.09 87.08 $C_3H_7NO_3$  Ser Serina | T 119.12 101.10 $C_4H_9NO_3$  Thr Treonina | E 147.13 129.11 $C_5H_9NO_4$  Glu Ácido Glutámico |
| K 146.19 128.17 $C_6H_{12}NO_2$  Lys Lisina | I 131.18 113.16 $C_6H_{11}NO_2$  Ile Isoleucina | P 115.13 97.12 $C_5H_9NO_2$  Pro Prolina | V 117.15 99.13 $C_6H_9NO_2$  Val Valina | Y 181.19 163.17 $C_9H_9NO_3$  Tyr Trosina | S 105.09 87.08 $C_3H_7NO_3$  Ser Serina | W 204.23 186.21 $C_{11}H_{12}N_2O_2$  Trp Triptófano |

Grupos protectores comunes en síntesis Fmoc de péptidos

| | | | |
|--|--|---|--|
| Fmoc 9-Fluorenilmetoxi-carbonil $M_r = 223.25$  | Mt 4-Metilnitil $M_r = 257.36$  | Pnc 2,2,5,7,8-Pentametil-croman-6-sulfonil $M_r = 267.37$  | t-Bu t-Butil $M_r = 57.12$  |
|--|--|---|--|



LA QUÍMICA BIOLÓGICA





La química, en sus principios, era considerada como una rama de la ciencia encargada del estudio de la materia inerte, por lo tanto muy alejada de la biología y la medicina. Sin embargo, a lo largo de la historia esta separación ha ido disminuyendo y actualmente la frontera entre ambas ciencias está muy difuminada. Uno de los momentos claves en la química, fue la síntesis de la urea llevada a cabo por Friedrich Wöhler en 1828,^[1] que demostró que moléculas biológicas podían ser sintetizadas a partir de material considerado no biológico.

Con el paso de los años la química, como ciencia responsable de analizar y sintetizar moléculas, fue demostrando su utilidad para entender los procesos biológicos. Esto provocó el nacimiento en el siglo XIX de una nueva disciplina encargada de aunar los conocimientos de la química y la biología, que fue denominada *bioquímica*. Sin embargo, la química y la biología continuaron evolucionando de forma bastante separada hasta la mitad del siglo XX. Fue en este momento cuando relevantes hitos, como el descubrimiento del código genético, la posibilidad de sintetizar biomoléculas complejas y el mayor entendimiento del reconocimiento molecular, permitieron la integración entre la química y la biología.

La comunidad científica comprendió que la química proporciona herramientas excelentes para entender las interacciones entre biomoléculas y para manipularlas de forma programada y selectiva. En este ámbito nace la *química biológica*,^[2] que sirve como puente entre la síntesis, la química molecular, la biología celular y las ciencias biomédicas. Esta nueva disciplina se define como la rama de la ciencia enfocada en el desarrollo y uso de técnicas y estrategias químicas para estudiar y manipular sistemas biológicos. La química biológica abarca campos tan diversos como el reconocimiento biomolecular, la catálisis enzimática, la señalización celular, los procesos de internalización y transporte, la síntesis de biomoléculas o la regulación de la expresión génica.

Esta nueva rama ha tenido un gran impacto en la ciencia actual, tal como se pone de manifiesto con la aparición, a partir de la década de los noventa, de varias revistas especializadas en esta temática tales como: *Chemistry & Biology*, *ChemBioChem*, *Current Opinion in Chemical Biology* y otras aun más recientes como *ACS Chemical Biology* y *Nature Chemical Biology*. En este contexto también merece la pena destacar la reciente creación del Grupo Especializado en Química Biológica de la Real Sociedad Española de Química (<http://quimicabiologica.es/>).

[1] Wöhler, F. *Annalen der Physik und Chemie* **1828**, 88, 253.

[2] Famulok, M; Waldmann, H. *ChemBioChem* **2001**, 2, 3.



INTRODUCCIÓN





1. CONCEPTOS BÁSICOS: ADN Y PROTEÍNAS

El ADN es una molécula polimérica formada por condensación de desoxirribonucleótidos, los cuales a su vez están compuestos por un núcleo de β -D-2'-desoxirribosa unido a una base nitrogenada mediante un enlace β -glicosídico en el carbono 1' y a un grupo fosfato a través del carbono 5' (Figura 1). Por convenio, las cadenas de ADN se nombran según su secuencia de bases leídas en la dirección del extremo 5' al 3'.

La presencia del grupo fosfato, que se encuentra disociado a pH fisiológico, convierte al ADN en un polímero polianiónico.^[3] La variabilidad química del ADN se concentra en las bases, que son las responsables directas de la codificación genética, mientras que los grupos de azúcar y fosfato tienen un papel fundamentalmente estructural. Las bases nitrogenadas son compuestos heterocíclicos que pueden ser de dos tipos: *púrico* que contienen un par de anillos fusionados (imidazol y pirimidina) y *pirimidínico* que contiene un único anillo de pirimidina. Las bases púricas presentes en el ADN son la adenina (A) y guanina (G); las bases pirimidínicas son la timina (T) y citosina (C).

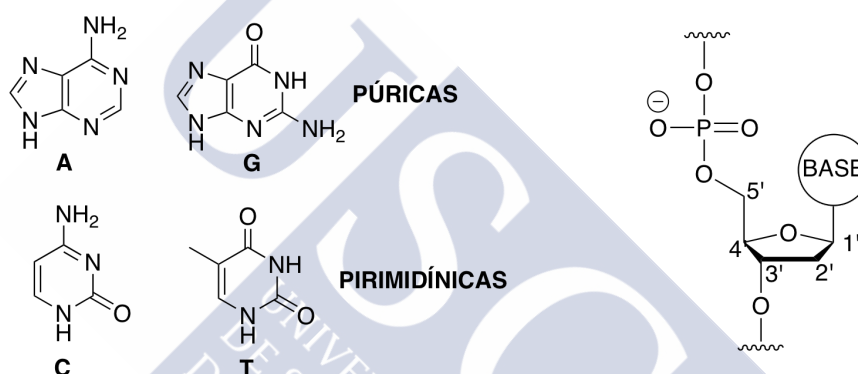


Figura 1. Izquierda: Bases nitrogenadas presentes en el ADN. Derecha: Estructura general de los nucleótidos que forman parte del ADN, con la numeración estándar de los átomos de carbono del azúcar.

El ADN se encuentra normalmente en forma de dímeros antiparalelos de oligonucleótidos; estos dímeros están estabilizados por puentes de hidrógeno entre las bases nitrogenadas y el efecto hidrofóbico resultante del apilamiento de las bases y su exclusión del medio acuoso, que son capaces de compensar las repulsiones electrostáticas entre los fosfatos cargados negativamente—normalmente los fosfatos en el ADN se encuentran asociados a cationes metálicos que neutraliza parcialmente su carga. Las restricciones geométricas y la disposición de los grupos dadores y aceptores de enlace de hidrógeno, provocan que la adenina se empareje con la timina (A-T, -13 Kcal/mol en fase gas) y la guanina con la citosina (G-C, -21 Kcal/mol en fase gas). Debido a esto se dice que la adenina-timina y citosina-guanina son pares de bases complementarias.^[4]

[3] a) Lavery, R.; Zadecki, C.; Westbrook, J. *Oxford Handbook of Nucleic Acid Structure* **1999**, Oxford. b) Neidle, S. *Principles of Nucleic Acid Structure* **2008**, Elsevier Inc., UK.

[4] a) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, P.; Darnell, J. E. *Molecular Cell Biology* **2000**, W. H. Freeman & Co., New York. b) Calladine, C. R.; Drew, H. R.; Luisi, B. F.; Travers, A.A. *Understanding DNA: The molecule & How it works* **2004**, Elsevier Ltd., UK. c) Yanson, I. K.; Teplitskii A. B.; Sukhodub, L. F. *Biopolymers* **1970**, *18*, 1149.

La conformación más común del ADN es la denominada *forma B*, que presenta una estructura helicoidal dextrógira en la cual la hélice da un giro completo cada 3.4 nm. Cada vuelta de la hélice se compone de 10 bases, que se encuentran prácticamente perpendiculares al eje de la doble hélice. Esta conformación origina dos surcos de diferente tamaño, uno estrecho y profundo denominado surco menor (5 Å de ancho y 8.5 Å de profundidad) y otro más ancho y superficial denominado surco mayor (12 Å de ancho y 7.5 Å de profundidad).^[5]

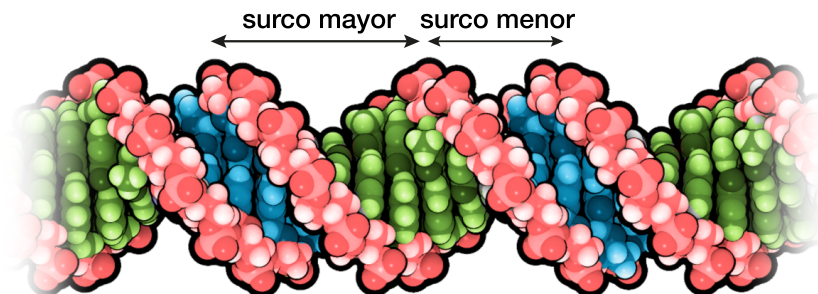


Figura 2. Estructura del ADN en *forma B*. El esqueleto de ribosa-fosfato se muestra en rojo y las bases nitrogenadas en el interior de cada surco se muestran en verde (surco mayor) y azul (surco menor).

Debido a la disposición de las bases en el ADN, el surco mayor presenta un patrón dador-aceptor de enlaces de hidrógeno más variable que el menor y por lo tanto contiene mayor información (Figura 3).^[6] Un claro ejemplo es que los pares de bases A/T y T/A en el surco menor son degenerados en relación a su habilidad para formar enlaces de hidrógeno. Debido a este motivo y a su mayor accesibilidad, las proteínas reconocen el ADN preferentemente a través del surco mayor.

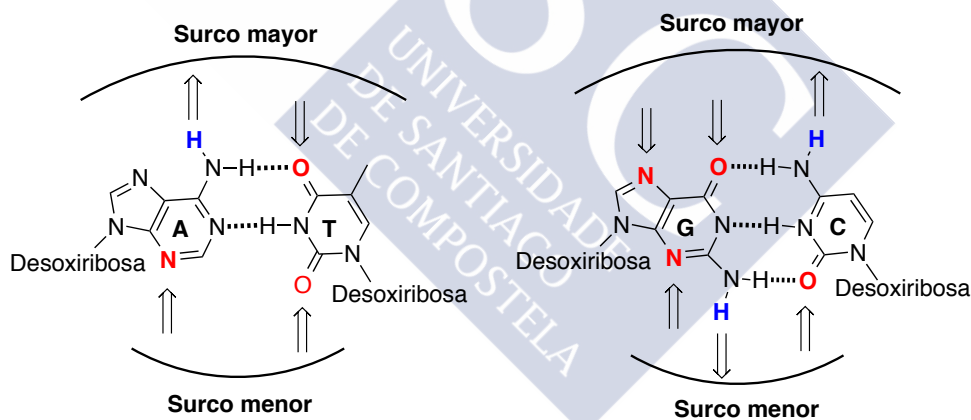


Figura 3. Apareamiento de bases A/T y G/C en el ADN. Las flechas indican los átomos que pueden formar enlaces de hidrógeno con otros componentes con los dadores en azul y los aceptores en rojo.

[5] a) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737. b) Blackburn, G. M.; Gait, M. J.; Loakes, D.; Williams, D. M. *Nucleic Acids in Chemistry and Biology* **2006**, RSC Publishing, Cambridge. c) Banaszak, L. J. *Foundations of Structural Biology* **2000**, Academic Press, San Diego.

[6] Luscombe, N. M.; Laskowski, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H. *Nucl. Acids Res.* **2000**, *28*, 235.

1.1. Función biológica del ADN

El ADN es el biopolímero que almacena la información genética. Esta información está codificada en la secuencia de bases de la cadena del ácido nucleico. Cada base puede considerarse como una letra de un alfabeto de cuatro letras que se utiliza para componer mensajes biológicos. La concordancia entre el alfabeto de nucleótidos del ADN (4 letras) y el de aminoácidos (20 letras) se manifiesta en el proceso de la expresión génica,^[7] a través del cual la célula transcribe la secuencia de nucleótidos del ADN a una molécula de ARN (transcripción), que luego traduce la información en la secuencia de aminoácidos de una proteína (traducción).^[8] No todos los fragmentos del ADN pueden transcribirse, solamente unos pequeños segmentos denominados *genes*, que pueden definirse como secuencias de ADN que codifican proteínas específicas. El conjunto de todos los genes en el organismo recibe el nombre de *genotipo* y el resultado de su expresión se conoce como *fenotipo*.^[9]

Las células eucariotas poseen varias cadenas muy largas de ADN (10^7 – 10^9 pb) almacenado en el núcleo celular, que se compactan en estructuras denominadas *cromosomas*. El ADN en estas estructuras se encuentra formando parte de un complejo proteína-ADN denominado *cromatina* (Figura 4). El *nucleosoma* es la unidad estructural mínima de la cromatina y está constituido por una secuencia de ADN (146 pb) enrollado a un octámero de proteínas denominadas *histonas*.^[10] Además de un papel estructural, la cromatina también juega un papel fundamental en la regulación de la expresión facilitando o restringiendo el acceso de la maquinaria transcripcional a las hebras de ADN y, consecuentemente, desencadenando o inhibiendo el proceso de la transcripción.

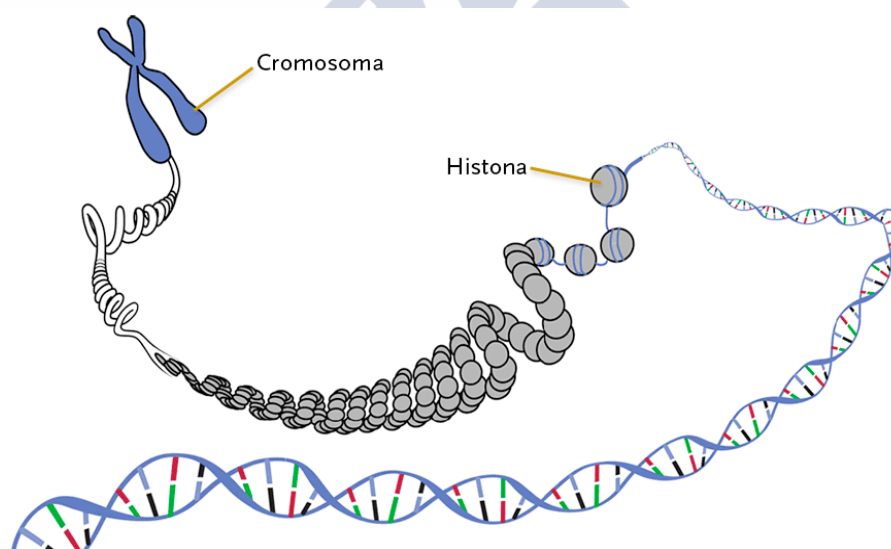


Figura 4. Empaquetamiento del ADN desde su forma condensada (cromosoma) hasta una cadena.

[7] Mathews, C. K.; Van Holde, K. E.; Ahern, K. G. *Bioquímica* **2003**, Addison Wesley, Madrid, cap. 4.

[8] Crick, F.H.C. *Nature* **1970**, 227, 561.

[9] a) Pearson, H. *Nature* **2006**, 441, 398. b) Pennisi, E. *Science* **2007**, 316, 1556.

[10] a) Li, Y. J.; Fu, X. H.; Liu, D. P.; Liang, C. C. *Int. J. Biochem. Cell. Biol.* **2004**, 36, 1411. b) Khorasanizadeh, S. *Cell*, **2004**, 116, 259.

1.2. Las proteínas

Al igual que el ADN es un biopolímero formado por nucleótidos, las proteínas también son biopolímeros, aunque en este caso formados por α -aminoácidos. Un α -aminoácido es una molécula que contiene un grupo carboxilo y un grupo amino unidos al mismo átomo de carbono. La mayor parte de las proteínas están formadas por la combinación de 20 de estos monómeros, conocidos como *aminoácidos estándar* que difieren entre sí por la naturaleza de los grupos funcionales presentes en sus cadenas laterales (Figura 5). Estos grupos varían en tamaño, forma, carga, hidrofobicidad y reactividad, agrupándose en función de la naturaleza de la misma en *hidrofóbicos* (G, A, V, L, I, M, F, P, W), *polares* (S, T, N, Q, Y, C), *ácidos* (D, E) y *básicos* (K, R, H). Los aminoácidos estándar se nombran habitualmente por abreviaturas de tres letras o símbolos de una sola letra, que se utilizan para indicar de manera abreviada la secuencia de una proteína.^[11]

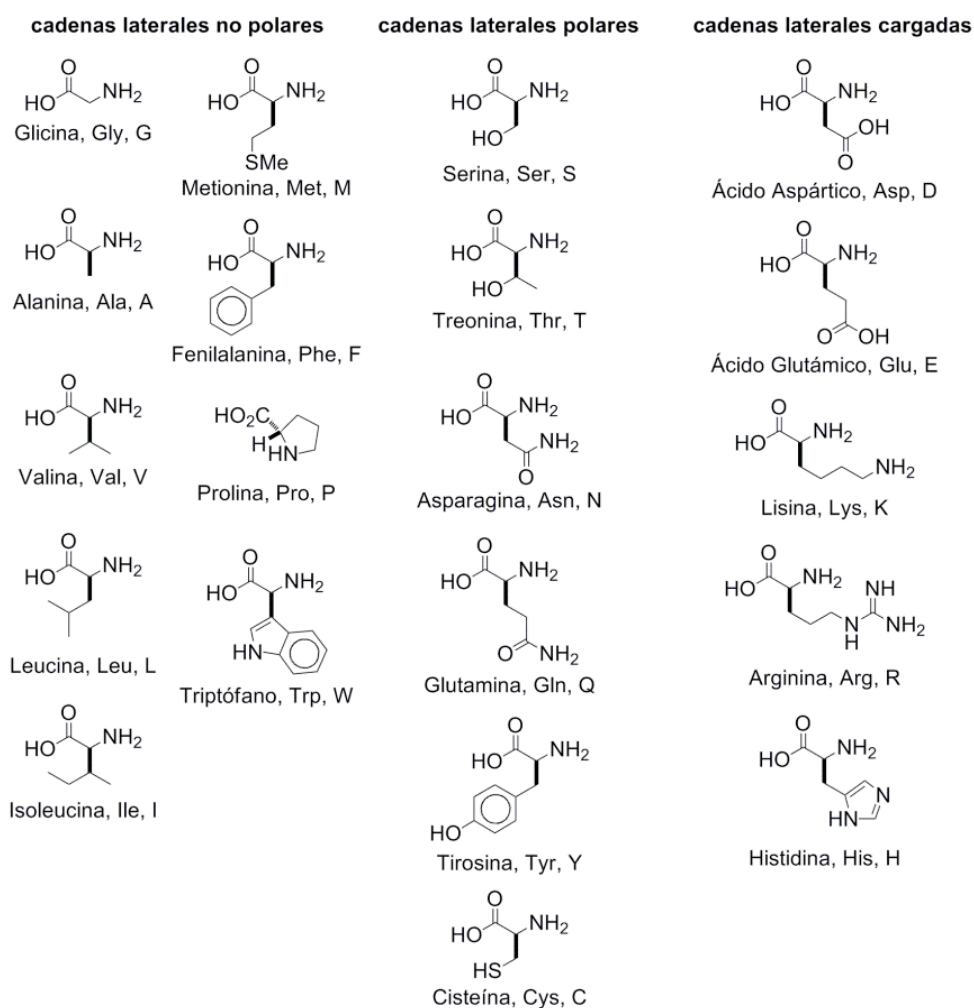


Figura 5. Estructura y códigos de una y tres letras para los aminoácidos estándar, agrupados según la naturaleza de su cadena lateral.

[11] a) Voet, D.; Voet, J. G. *Biochemistry* **1995**, John & Sons, New York, cap. 7. b) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell* **2008**, Garland Publishing, New York; cap. 3. c) Mathews, C. K.; Van Holde, K. E. *Bioquímica* **2002**, McGraw-Hill/Interamericana, Madrid, cap. 5.

Las proteínas son polímeros formados por la condensación del grupo carboxilo de un aminoácido con el grupo amino de otro aminoácido en forma de amidas, este tipo de unión se conoce comúnmente como *enlace peptídico* (Figura 6). El equilibrio de esta reacción está desplazado hacia la hidrólisis, por lo tanto la síntesis de proteínas requiere un aporte de energía. Sin embargo, el enlace amida es cinéticamente estable una vez formado.^[12]

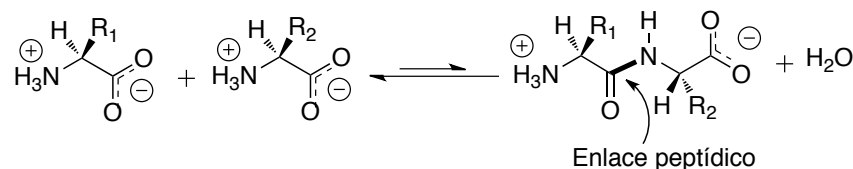


Figura 6. Formación del enlace peptídico o amida.

Estudios de difracción de rayos X llevados a cabo por Linus Pauling y Robert Corey demostraron que el enlace peptídico C–N es ligeramente más corto que el enlace C–N de una amina simple y que los átomos asociados con el enlace son coplanares. A partir de estos descubrimientos, Pauling y Corey dedujeron que el enlace peptídico C–N no puede girar libremente a causa de su carácter parcial de doble enlace, por lo que el esqueleto polipeptídico de una proteína puede visualizarse como una serie de planos rígidos consecutivos que comparten puntos comunes de giro alrededor de los C α . Esta restricción en el número de enlaces que pueden girar libremente, reduce drásticamente el número de conformaciones posibles en una cadena polipeptídica y permite definir su conformación únicamente mediante dos ángulos diedros denominados ϕ (N–C α) y ψ (C α –C) (Figura 7).^[13]

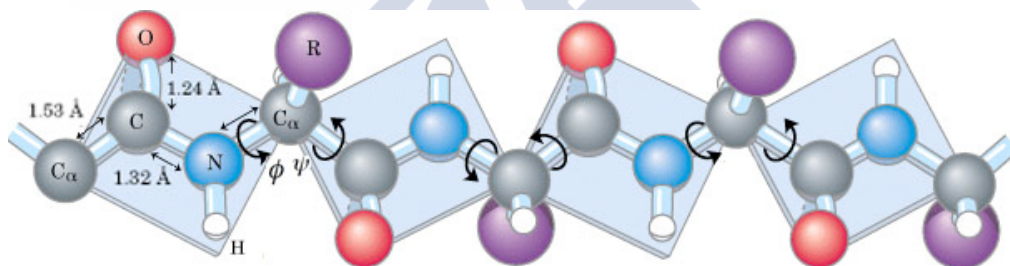


Figura 7. Rotación alrededor de los enlaces de un polipéptido y ángulos diedros que definen la conformación de la cadena polipeptídica

La estructura de las proteínas se describe tradicionalmente en términos de **cuatro niveles de organización**:

- *Estructura primaria*: secuencia de aminoácidos del polipéptido.
- *Estructura secundaria*: distribución espacial local de los átomos de la cadena principal sin tener en cuenta las cadenas laterales. Estos plegamientos locales pueden seguir unos patrones conocidos: hélices α , giros, láminas β , etc.

[12] Martin, R. B. *Biopolymers*, **1998**, *45*, 351.

[13] a) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry* **2008**, W. H. Freeman and Company, NY; cap.4. b) Stryer, L.; Berg, J. M.; Tymoczko, J. L. *Bioquímica con aplicaciones clínicas* **2012**, Editorial Reverté, cap. 2.

- *Estructura terciaria*: disposición tridimensional de todos los átomos de una proteína resultante de la organización espacial de los motivos de estructura secundaria. Suele estar estabilizada por interacciones hidrofóbicas entre las cadenas laterales no polares, o en algunos casos por enlaces disulfuro.
- *Estructura cuaternaria*: disposición de subunidades proteicas en complejos tridimensionales. Sólo las proteínas constituidas por más de una cadena polipeptídica poseen este tipo de estructura.

1.2.1. La hélice α

La hélice α es una estructura cilíndrica en la cual el esqueleto polipeptídico se encuentra estrechamente enrollado dextrógicamente alrededor de un eje imaginario dibujado longitudinalmente por el centro de la hélice y los grupos de las cadenas laterales se orientan hacia fuera de la hélice. Esta estructura se estabiliza por el efecto hidrofóbico resultante de la compactación del esqueleto peptídico y por el establecimiento de puentes de hidrógeno entre los grupos N-H y C=O de la cadena principal, concretamente el grupo C=O forma un enlace de hidrógeno con el grupo N-H de un aminoácido situado cuatro residuos más adelante. Esta conformación corresponde a unos ángulos $\phi = -57^\circ$ y $\psi = -47^\circ$ y cada giro de hélice incluye 3.6 aminoácidos. ^[13, 14]

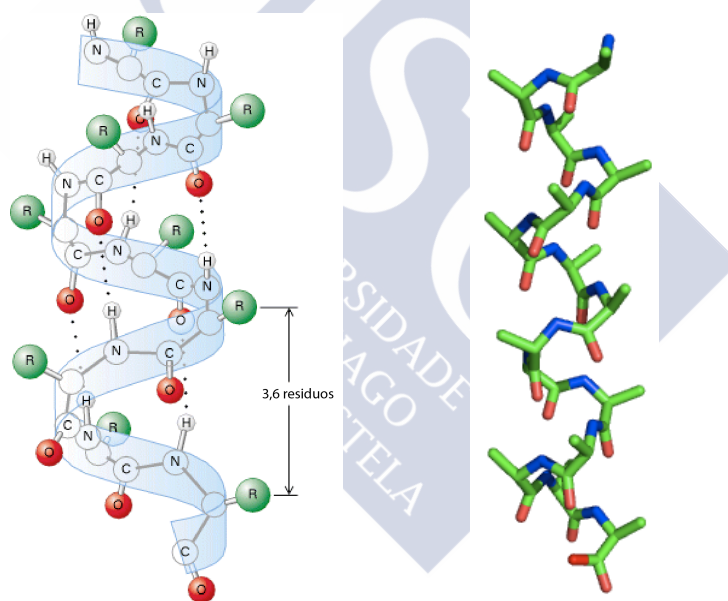


Figura 8. Izquierda: Modelo de bolas de una hélice α en el que se observan las interacciones claves entre N-H y C=O. Derecha: Estructura cristalográfica de una hélice α .

[14] a) Branden, C.; Tooze, J. *Introduction to Protein Structure* **1999**, Garland Publishing, Inc., New York.
b) Sewald, N.; Jakubke, H-D. *Peptides: Chemistry and Biology* **2002**, Wiley-VCH GmbH & Co, Weinheim.

- Factores de estabilización de las hélices α :

Cada aminoácido de un polipéptido tiene una tendencia intrínseca a formar hélice α , según el tipo de cadena lateral. Los aminoácidos con menos tendencia a encontrarse en estructuras secundarias de este tipo son la prolina y la glicina. En el caso de la prolina, se debe a la falta del grupo N-H y restricción conformacional impuesta por su estructura cíclica, y en el caso de la glicina es debido a su alta flexibilidad conformacional que penaliza entrópicamente el plegamiento en forma de hélice α . Por otra parte, los aminoácidos con ramificaciones en el átomo de carbono β , como es el caso de valina, treonina y isoleucina, tienden a desestabilizar las hélices debido a impedimentos estéricos; también son desestabilizantes los aminoácidos con cadenas que contienen dadores o aceptores de puentes de hidrógeno muy próximos a la cadena principal, pues compiten por los grupos C=O y N-H de la cadena principal, es el caso de serina, aspartato y asparagina. Así pues, el aminoácido con mayor tendencia a formar hélices α es la alanina, seguida de cerca por la arginina.^[15]

Otro factor importante en la estabilización de la hélice es la posición de un residuo con respecto a sus vecinos, puesto que las interacciones entre cadenas laterales pueden estabilizar o desestabilizar esta estructura. Estas interacciones son especialmente importantes entre las cadenas laterales que se encuentran separadas por tres aminoácidos, debido a su proximidad en el espacio en la hélice plegada. Los casos más importantes son las interacciones electrostáticas, que dan lugar a pares iónicos entre aminoácidos con diferente carga y la presencia de residuos aromáticos que pueden dar lugar a interacciones hidrofóbicas.

Finalmente, la estabilidad de una hélice también depende de los residuos localizados cerca de los extremos. Cada hélice presenta un pequeño dipolo eléctrico neto debido a la alineación de dipolos de sus residuos, por lo tanto las cargas parciales positivas y negativas del dipolo de la hélice residen en los grupos amino y carbonilo peptídicos situados cerca de los extremos. Por este motivo normalmente se encuentran residuos cargados positivamente cerca del extremo C-terminal y negativos en el N-terminal, los cuales ejercen una interacción estabilizadora con la carga del dipolo. Este tipo de interacción estabilizadora se conoce como “N-capping” o “C-capping” según en el extremo en el que se produce.^[13]

1.2.2. La lámina β

Esta estructura secundaria esta formada por dos o más cadenas polipeptídicas cortas (5-8 residuos) denominadas hebras β . Estas hebras se encuentran completamente extendidas en forma de zig-zag, lo que genera una distancia entre aminoácidos de 3.5 Å. Las cadenas polipeptídicas se encuentran unidas a través de enlaces de hidrógeno entre átomos del esqueleto de cadenas adyacentes. Estas hebras pueden estar orientadas una respecto a otra de forma paralela o antiparalela, siendo la antiparalela la forma más estable debido a la menor distorsión de los enlaces de hidrógeno. La cadenas laterales de aminoácidos se disponen perpendicularmente al plano de la lámina y en direcciones opuestas para residuos continuos, dando lugar a un patrón alterno en las cadenas.^[13,14]

[15] Pace, C. N.; Scholtz, J. M. *Biophysical Journal* **1998**, 75, 422.

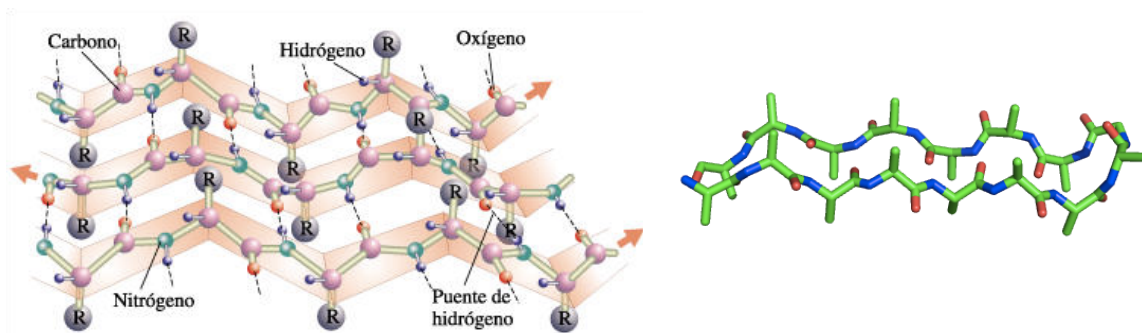


Figura 9. Izquierda: Modelo de una lámina β en el que se observan las interacciones claves entre N-H y C=O. Derecha: Estructura cristalográfica de una lámina β .



2. LA EXPRESIÓN GÉNICA

2.1. Etapas generales del proceso de expresión génica

El término expresión génica se refiere al proceso completo mediante el cual la información contenida en un gen es descodificada, dando lugar a una determinada proteína. De forma simplificada el proceso consta de dos etapas básicas: *transcripción* y *traducción*. La primera etapa ocurre en el núcleo celular y en ella hay una transferencia de información desde un de ácido nucleico (ADN) a otro (ARN), de ahí el nombre de transcripción. La segunda etapa tiene lugar en el citoplasma, en ella un hay cambio de “lenguaje” y la información contenida en el ARN se transforma en una secuencia de aminoácidos, de ahí el nombre de *traducción*. Este flujo direccional de información se conoce como el *dogma central de la biología molecular*.^[8]

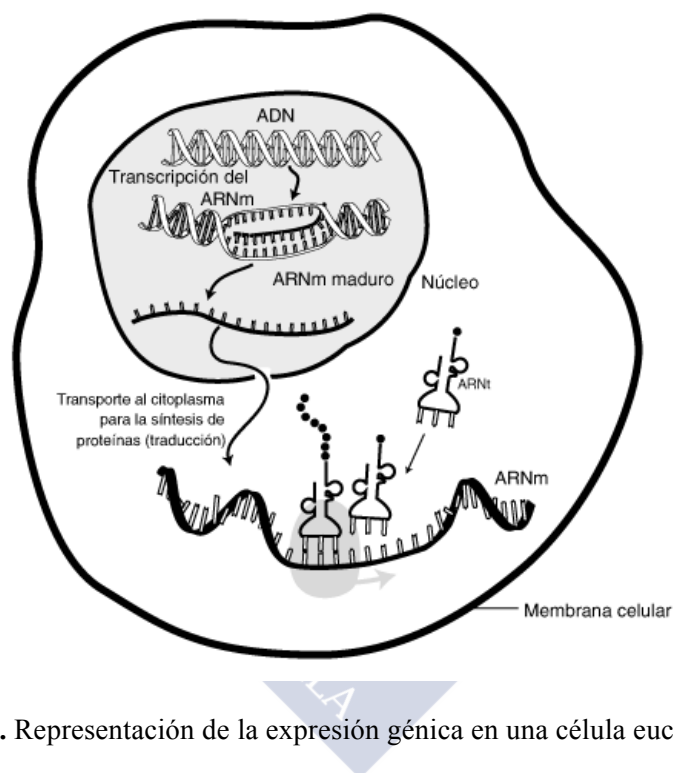


Figura 10. Representación de la expresión génica en una célula eucariota.

2.2. La transcripción

En esta etapa se genera el ARN, el cual es químicamente similar al ADN excepto en que como azúcar contiene ribosa en lugar de desoxirribosa, y tiene la base uracilo en lugar de la timina. La síntesis del ARN es el proceso más regulado de la expresión génica. Un ejemplo de el alto grado de regulación es el hecho de que todas la células de un organismo pluricelular tienen el mismo material genético, sin embargo la diferente regulación en la etapa de transcripción produce la enorme variabilidad morfológica y funcional que da lugar a los diferentes tejidos.

La generación del ARN es catalizada por la *ARN polimerasa*, que sintetiza tres tipos de ARN: ARNm (ARN mensajero), ARNt (ARN de transferencia) y ARNr (ARN ribosómico).

El ARNm es el encargado de llevar la información genética del ADN a los ribosomas, donde tiene lugar la síntesis de las proteínas. El ARNt lleva el aminoácido adecuado al ribosoma según la secuencia del ARNm y el ARNr forma parte del ribosoma.^[16] En células eucariotas existen tres tipos distintos de enzimas ARN polimerasa: la *ARN polimerasa I* que transcribe los ARNr,^[17] la *ARN polimerasa III* que codifica un número pequeño de ARNs entre los que se encuentran los ARNt,^[18] y la *ARN polimerasa II* (ARNpol II),^[19] responsable de la transcripción del ARNm. La polimerización catalizada por ARNpol II consta de tres etapas:^[16]

- *Iniciación*: en la que la ARNpol II se une al promotor y desencadena el desenrollamiento local de la doble hélice de ADN.
- *Elongación*: usando una de las dos cadenas de ADN como molde, la ARNpol II forma la cadena de ARNm en dirección 5' a 3', al mismo tiempo que genera el ARN va desenrollando la doble hélice de ADN.
- *Terminación*: se libera el ARNm tras incorporar en el extremo 3' una cola de poliA.

La ARNpol II se asocia a un gran número de proteínas denominadas *factores de transcripción generales (FTGs)*^[20] para formar el complejo de transcripción activo. Estos complejos buscan en el ADN centros promotores, que son los lugares donde va a comenzar la transcripción de un gen y que se encuentran generalmente situados del lado 5' del punto de inicio de la transcripción. Cuando este conjunto FTGs/ARNpol II se une al núcleo promotor se produce la fosforilación de este dominio, lo que provoca un cambio substancial en las propiedades iónicas de la ARNpolII que permite el comienzo de la síntesis del ARNm. A pesar de su gran complejidad, esta maquinaria molecular por si sola no es capaz de producir la transcripción de manera eficiente y normalmente necesita la presencia adicional de cofactores y sobre todo de otras proteínas denominadas *factores de transcripción específicos (FT)*,^[21] que cuando se unen a secuencias reguladoras del gen promueven el inicio de su expresión.^[22]

[16] a) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry* **2008**, W. H. Freeman and Company, NY; cap. 26. b) Stryer, L.; Berg, J. M.; Tymoczko, J. L. *Bioquímica con aplicaciones clínicas* **2012**, Editorial Reverté, cap. 29.

[17] Russell, J.; Zomerdiijk, J. C. *Biochem. Soc. Symp.* **2006**, 73: 203.

[18] Ishiguro, G. A.; Kassavetis, G. A.; Peter E. *Mol. Cell. Biol.* **2002**, 22, 3264.

[19] Svejstrup, J. Q.; Conaway, R. C.; Conaway, J. W. *Molecular Cell* **2006**, 24, 637.

[20] a) Buratowski S. *Cell* **1994**, 77, 1. b) Juven-Gershon, T.; Hsu, J-Y.; Theisen, J. W. M.; Kadonaga, J. T. *Curr. Opin. Cell Biol.* **2008**, 20, 253.

[21] a) Lemon, B.; Tjian, R. *Genes Dev.* **2000**, 14, 2551. b) Flanagan, P. *Nature* **1991**, 350, 436.

[22] Ptashne, M.; Gann A. *Nature* **1997**, 386, 569.

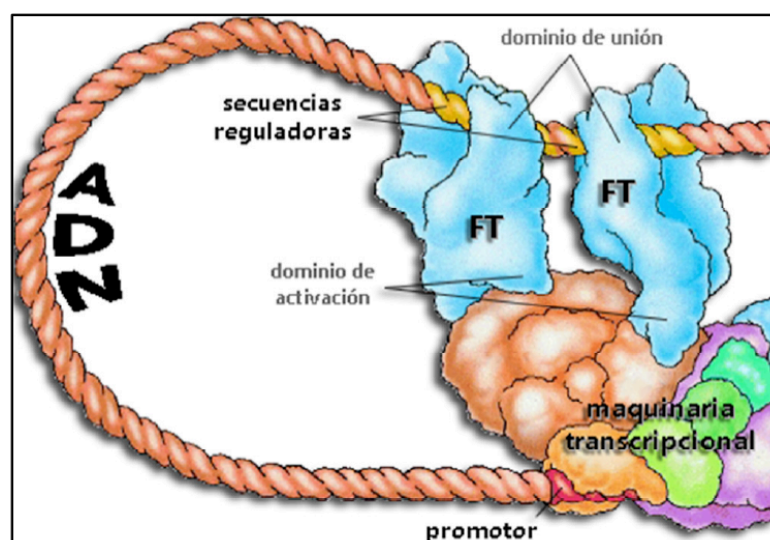


Figura 11. Representación de la interacción de los FT con la maquinaria transcripcional.

2.3. La traducción

Una vez que la molécula de ARNm fue sintetizada por la ARN polimerasa, esta sufre un etapa de maduración en el núcleo. En este proceso el ARN experimenta varios procesos de corte y empalme, en los que se eliminan unas regiones no codificantes denominadas *intrones*.^[16a] A continuación el ARNm, ya maduro, es exportado al citoplasma a través del poro nuclear.^[23] El ARNm una vez en el citoplasma se asocia con los ribosomas, los cuales son partículas ribonucleoproteicas compuestas por dos subunidades formadas por varias proteínas y moléculas de ARNr.^[24] Estas dos subunidades se encuentran separadas y se asocian cuando la subunidad más pequeña se une al ARNm y a un ARNt. La molécula de ARNt tiene como función transportar el aminoácido que va a ser acoplado a la cadena polipeptídica creciente. El ARNt posee en un extremo el aminoácido y en el otro extremo tiene un triplete de bases llamado *anticodón*, que se aparea por complementariedad de bases con el codón presente en el ARNm. Por lo tanto cada triplete de bases del ARNm corresponde a un determinado aminoácido, esta equivalencia entre los tripletes de bases y los aminoácidos se denomina el *código genético*.^[25]

[23] Köhler, A.; Hurt, E. *Nature Reviews Molecular Cell Biology* **2007**, *8*, 761.

[24] a) Thomas A. S. *Nature Reviews Molecular Cell Biology* **2008**, *9*, 242. b) Ramakrishnan, V. *Cell* **2002**, *108*, 557. c) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679.

[25] a) Rodnina, M.V.; Beringer, M.; Wintermeyer, W. *Trends Biochem. Sci.* **2007**, *32*, 20. b) Lodish, H.; Berk, A.; Zipursky, L. S.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology* **2000**, W. H. Freeman, New York, capítulo 10.

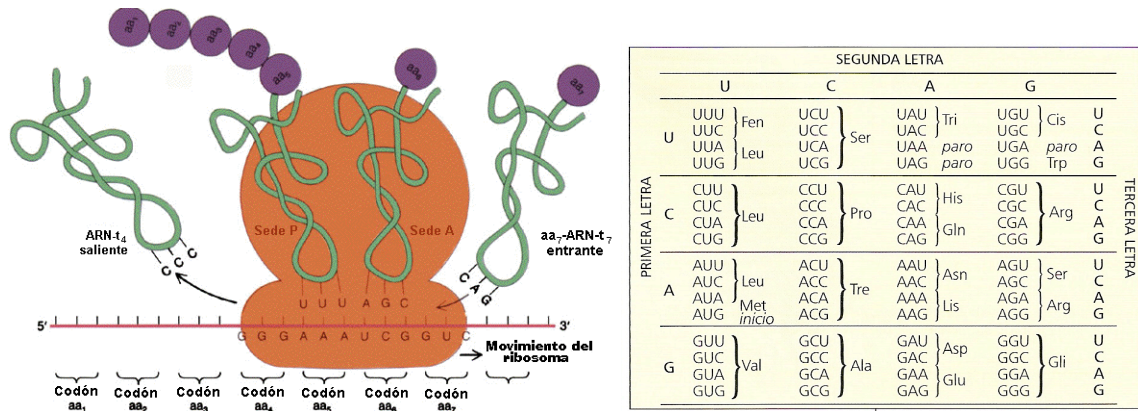


Figura 12. Izquierda: Mecanismo de síntesis de las proteínas. En naranja se muestra el ribosoma. Derecha: El código genético.



3. LOS FACTORES DE TRANSCRIPCIÓN

3.1. Bases generales del reconocimiento molecular entre FT y proteínas

Como se ha comentado antes, los factores de transcripción específicos (FT) son proteínas especializadas que interactúan de forma selectiva con secuencias genéticas específicas (zonas reguladoras proximales o distales) y de esta forma inducen la formación de la maquinaria transcripcional.^[26] Los FTs logran su especificidad de secuencia debido a que poseen una superficie que es complementaria químicamente a las zonas expuestas de los pares de bases.^[27] Además de estas interacciones específicas, la mayoría de proteínas que unen al ADN también hacen un gran número de contactos no específicos con los fosfatos del esqueleto del ADN. Las interacciones claves para el reconocimiento específico del ADN por parte de los FT suelen tener lugar en el surco mayor y en la mayoría de los casos consisten en contactos entre las cadenas laterales de los aminoácidos de una hélice con los grupos funcionales expuestos en las bases del ADN.^[28]

Los contactos directos entre proteínas y el ADN se pueden agrupar en cuatro tipos:^[29]

- *Enlaces de hidrógeno e interacciones electroestáticas con la cadena de fosfatos:* estos contactos ocurren con las cadenas laterales de aminoácidos básicos (Lys, Arg y His). Normalmente este tipo de interacciones no confieren especificidad, pero aumentan la estabilidad termodinámica del complejo y ayudan a situar la unidad de reconocimiento específico en la orientación adecuada.

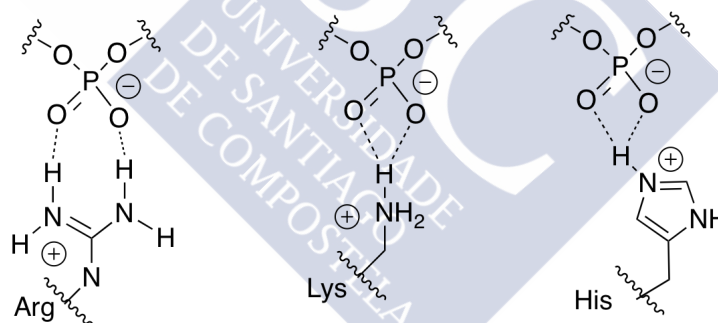


Figura 13. Ejemplo de las interacciones que pueden ocurrir entre los fosfatos del ADN y las cadenas laterales de aminoácidos básicos.

- *Enlaces de hidrógeno con las bases o los azúcares:* estas interacciones se establecen con las cadenas laterales polares de los aminoácidos y, debido a la direccionalidad del enlace de hidrógeno, son claves para la especificidad. Las estructuras de rayos X muestran que los residuos que más participan en este tipo de interacción son la Arg, Lys, Ser y Thr. Los enlaces de hidrógeno bidentados, que implican la formación de

[26] a) Latchman, D. S. *Eukaryotic Transcription Factors* **2003**, Elsevier, London. b) Luscombe, N. M.; Austin, S. E.; Berman, H. M.; Thornton, J. M. *Genome Biol.* **2000**, *1*, 1.

[27] Pazos, E.; Mosquera, J.; Vázquez, M. E.; Mascareñas J. L. *ChemBioChem* **2011**, *12*, 1958.

[28] Suzuki, M.; Gerstein, M. *Proteins* **1995**, *23*, 525.

[29] Sidney, M. H. *Bioorganic Chemistry: Nucleic Acids* **1996**, Oxford University Press, New York.

dos enlaces de hidrógeno entre una cadena lateral de un aminoácido y una o dos bases de ADN, son especialmente importantes en este tipo de reconocimiento (Figura 14).

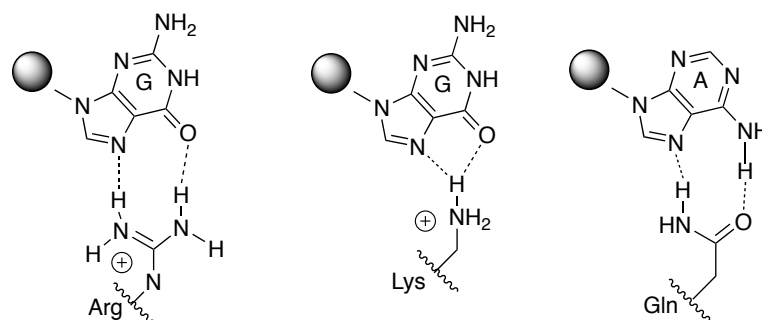


Figura 14. Enlaces de hidrógeno bidentados más comunes en el reconocimiento del ADN.^[30]

- *Interacciones hidrofóbicas con las bases:* cuando una molécula se inserta en un surco del ADN se produce un desplazamiento de las moléculas de agua; este proceso puede ir también acompañado de la pérdida de contraiones que rodean al ADN. Estos efectos contribuyen a compensar la pérdida de entropía asociada a la formación de un complejo intermolecular y son especialmente importantes en la unión a los surcos menores ricos en regiones de A y T, que son profundos y estrechos y están muy solvatados.^[31]
- *Enlaces de hidrógeno mediados por moléculas de agua:* son muy comunes en las proteínas, representando alrededor de un 15% de las interacciones totales entre el ADN y las proteínas.^[30] Este tipo de interacción se da cuando una molécula de agua forma un enlace de hidrógeno a la vez con el ADN y la proteína. Debido a la direccionalidad tienen un importante papel en el reconocimiento específico.

Además de interactuar con el ADN mediante contactos directos con los pares de bases, los FTs también hacen un reconocimiento denominado *indirecto*, relacionado con pequeñas variaciones en la estructura y flexibilidad del ADN dependientes de la secuencia.^[32]

Es importante indicar que la mayor parte de los FTs aislados presentan una baja afinidad por sus secuencias diana de ADN y es necesario la **homo o hetero-oligomerización** para lograr altas especificidades o reconocer secuencias largas en el genoma. Esta interacción cooperativa entre FTs se logra a través de otros dominios diferentes a los de unión del ADN, los cuales participan en contactos proteína-proteína.^[33] Debido al gran número de variables que la naturaleza usa para reconocer el ADN, no existen unas reglas generales que relacionen los aminoácidos y estructura secundaria de una proteína con una secuencia específica de reconocimiento en el ADN.^[34]

[30] Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. *Nucl. Acids Res.* **2001**, *29*, 2860.

[31] Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. *Structure* **2000**, *8*, 1015.

[32] Vázquez, M. E.; Caamaño, A. M.; Mascareñas, J. L. *Chem. Soc. Rev.* **2003**, *32*, 338.

[33] a) Nelson, H. C. *Curr. Opin. Genet. Dev.* **1995**, *5*, 180. b) Moretti, R.; Ansari, A. Z. *Biochimie* **2008**, *90*, 1015. c) Segal, D. J.; Barbas, C. F. *Curr. Opin. Chem. Biol.* **2000**, *4*, 34.

[34] Pabo, C. O.; Neklodova, L. *J. Mol. Biol.* **2000**, *301*, 597.

3.2. Familias de factores de transcripción

A pesar de que los factores de transcripción reconocen el ADN usando una gran variedad de motivos estructurales, muchos de ellos comparten características comunes, lo que ha permitido clasificarlos en familias según la estructura secundaria del dominio de reconocimiento del ADN.^[35]

3.2.1. Hélice-Giro-Hélice (HTH: Helix-turn-Helix) y Homeodominio

El dominio HTH contiene tres hélices que se encuentran próximas en el espacio y forman una estructura globular compacta. Las hélices I y II son paralelas y casi perpendiculares a la hélice III, que reconoce el ADN a través del surco mayor y por eso se denomina *hélice de reconocimiento*. Este motivo suele presentar además una cola en el extremo *N*-terminal que interacciona con el surco menor adyacente.^[36] El motivo HTH es muy común en los factores de transcripción procariotas en los que suele operar en forma de oligómeros con otros FT. Cuando este motivo se encuentra en FT eucariotas se denomina *homeodominio* y típicamente reconoce secuencias implicadas en la regulación del desarrollo.^[37]

Uno de los factores de transcripción más representativos de esta familia es el *Engrailed*. Homólogos de este FT se encuentran presentes en varias especies, diferenciándose sólo en pequeñas variaciones en la secuencia y suelen estar implicados en el desarrollo del cerebro.^[38] Como modelo para esta proteína se usa el *Engrailed* presente en el organismo *Drosophila*, de cuyo complejo con el ADN se obtuvo la estructura de rayos X.^[39] El motivo de unión al ADN de esta proteína consta de 60 aminoácidos y reconoce de forma específica la secuencia *ERE* (TAATTA). Sin embargo, se ha encontrado que la mutación de uno de los aminoácidos de la hélice de reconocimiento, concretamente la Gln→Lys, da lugar a un mutante llamado *Engrailed Q50K* que reconoce la secuencia *QRE* (TAATCC) con una afinidad y especificidad mayores que las que muestra la proteína natural por su secuencia consenso (Figura 15).^[40]

-
- [35] a) Latchman, D. S. *Eukaryotic Transcription Factors* **2004**, Elsevier Ltd, Italy, cap. 4. b) Luscombe, N. M.; Austin, S. E.; Berman, H. M.; Thornton, J. M. *Genome Biology* **2000**, *1*, 1. c) Pabo, C. O. *Annu. Rev. Biochem.* **1992**, *61*, 1053.
- [36] a) Banerjee-Basu, S.; Baxevanis, A. D. *Nucleic acids reserach* **2001**, *29*, 3258. b) Gehring, W. J.; Qian, Y. Q.; Billeter, M.; Furukubo-Tokunaga, K.; Schier, A. F.; Resendez-Perez, D.; Affolter, M. M.; Orring, G.; Wüthrich, K. *Cell* **1994**, *78*, 221. c) Otting, G.; Qian, Y. Q.; Billeter, M.; Muller, M.; Affolter, M.; Gehring, W. J.; Wuthrich, K. *EMBO J.* **1990**, *9*, 3085.
- [37] Holland, P. W. *WIREs Dev. Biol.* **2013**, *2*, 31.
- [38] a) Jacobs, D. K.; Wray, C. G.; Wedeen, C. J.; Kostriken, R.; Desalle, R.; Staton, J. L.; Gates, R. D.; Lindberg, D. R. *Evolution & Development* **2000**, *2*, 340. b) Morgan, R. *FEBS Letters* **2006**, *580*, 2531.
- [39] a) Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579. b) Noyes, M. B.; Christensen, R. G.; Wakabayashi, A.; Stormo, G. D.; Brodsky, M. H.; Wolfe, S. A. *Cell* **2008**, *133*, 1277.
- [40] Tucker-Kellogg, L.; Rould, M. A.; Chambers, K. A.; Ades, S. E.; Sauer, R. T.; Pabo, C. O. *Structure* **1997**, *15*, 1047.

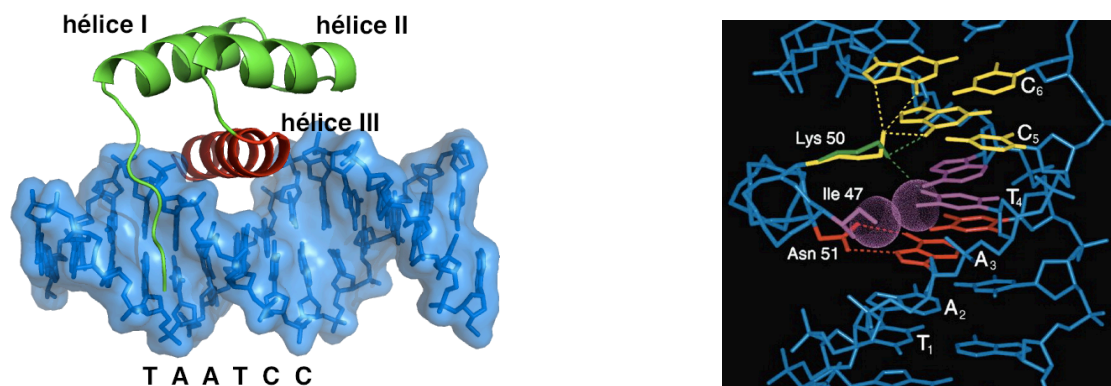


Figura 15. Izquierda: Estructura del homeodominio Engrailed Q50K interaccionando con el ADN. La hélice de reconocimiento se representada en rojo. (PDB: 2HDD). Derecha: Contactos más importantes de la hélice de reconocimiento del Engrailed Q50K con el ADN.^[37]

3.2.2. Proteínas de Cremallera de Leucinas (bZIP) y Hélice-Lazo-Hélice (HLH: Helix-Loop-Helix)

El nombre de este motivo estructural proviene de su unidad de dimerización, que es una cremallera de leucinas. El motivo bZIP es uno de los motivos estructurales más sencillos y consiste en homo- o heterodímeros helicoidales (coiled-coil) no covalentes, cuyas unidades monoméricas se pliegan en forma de hélices α ininterrumpidas en presencia del ADN consenso.^[41] Estas hélices poseen dos regiones muy diferenciadas:^[42]

- *Región básica:* constituida por unos 20-25 aminoácidos en el extremo *N*-terminal, contacta con el surco mayor del ADN y presenta múltiples residuos básicos (Arg y Lys). Una característica importante de la unión al ADN de las proteínas bZIP es que la región básica carece de una estructura secundaria definida en ausencia de su secuencia diana y sólo adopta la conformación de hélice α al complejarse con el ADN.^[43]
- *Cremallera de leucinas:* se encuentra en el extremo *C*-terminal y esta compuesta por aproximadamente por 30 aminoácidos. Se caracteriza por un patrón de repetición de residuos de Leu e Ile cada siete aminoácidos, lo que genera una superficie hidrofóbica a lo largo de la hélice que permite la dimerización.^[44] Aunque la dimerización de estas proteínas es un proceso que podría ocurrir en ausencia de ADN, ha sido demostrado que la unión al ADN se produce de forma secuencial, en el que primero se une uno de los monómeros y luego se produce la dimerización.^[45]

[41] Lee, K. A.; *J. Cell Sci.* **1992**, *103*, 9.

[42] a) Kerppola, T. K.; Curran, T. *Curr. Opin. Struct. Biol.* **1991**, *1*, 71. b) Hu, J. C.; Sauer, R. T. *Nucl. Acids Mol. Biol. Vol. VI* **1992**, 82. c) Vinson, C.; Acharya, A.; Taparowsky, E.J. *Biochim. Biophys. Acta* **2006**, *4*, 1759. d) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539. e) Glover, J. N.; Harrison, S.C. *Nature* **1995**, *373*, 257.

[43] a) Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J.P.; Harrison, S. C.; Struhl, K. *Nature* **1990**, *347*, 575. b) O'Neil, K. T.; Shuman, J.D.; Ampe, C.; DeGrado, W. F. *Biochemistry* **1991**, *30*, 9030.

[44] a) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538. b) O'Shea, E. K.; Rutkowski, R.; Stafford III, W. F.; Kim, P. S. *Science* **1989**, *245*, 646.

[45] Kohler, J. J.; Schepartz, A. *Bioorg. Med. Chem.* **2001**, *9*, 2435.

Una de las proteínas bZIP mejor caracterizadas es el factor de transcripción GCN4,^[46] un activador transcripcional de levaduras que interacciona preferentemente con las secuencias palindrómicas ATF/CREB y CRE (5'-ATGAcgTCAT-3') y la pseudopalindrómica AP1 (5'-ATGAcTCAT-3') en forma de homodímeros.^[47] Diversos estudios han permitido la resolución estructural del dominio de unión a ADN del GCN4, pudiendo observarse aquellos residuos que efectúan contactos específicos (no polares y enlaces de hidrógeno, Figura 16).^[48]

El motivo HLH es una variante del bZIP en el cual las dos regiones se encuentran separadas por el motivo de dimerización HLH. Este motivo consiste en dos hélices unidas por un lazo polipeptídico no helicoidal, en el que las hélices se encuentran unidas ininterrumpidamente a la región básica y la cremallera de leucinas.

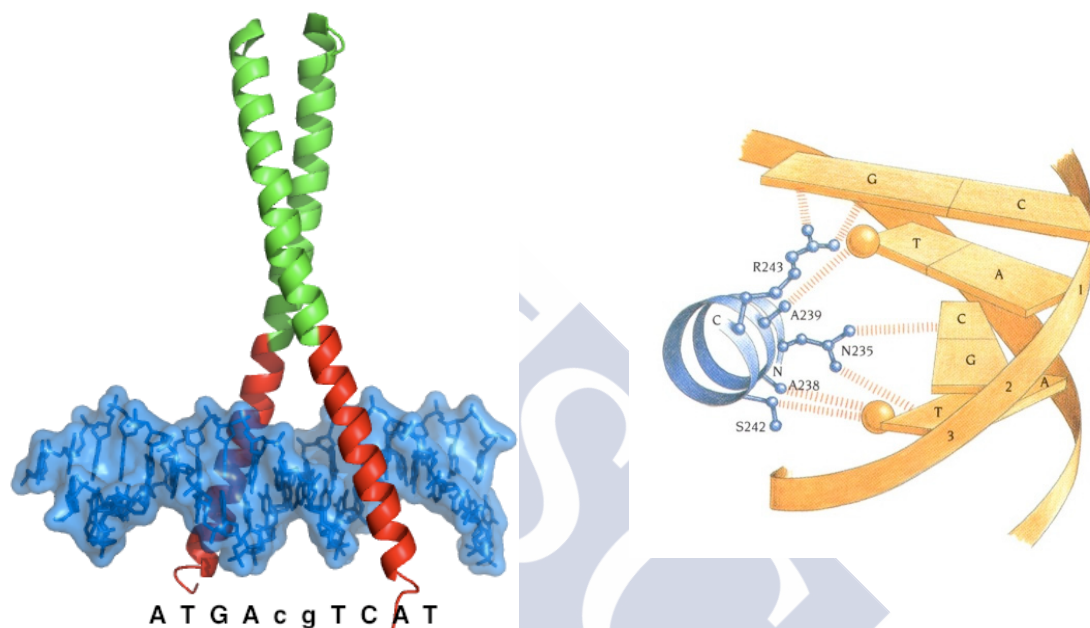


Figura 16. Izquierda: Estructura del dominio bZIP de la proteína GCN4 (PDB: 1YSA), la región básica se muestra en rojo.^[48a] Derecha: Contactos más reseñables de la región básica de GCN4 con su ADN consenso.^[48]

Ambas familias se encuentran exclusivamente en organismos eucariotas donde tienen papeles muy relevantes, especialmente relacionados con procesos de proliferación y diferenciación celular.^[49]

[46] Hope, I. A.; Struhl, K. *Cell* **1986**, *46*, 885.

[47] Keller, W.; König, P.; Richmond, T. J. *J. Mol. Biol.* **1995**, *254*, 657.

[48] a) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223. b) König, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139.

[49] a) Benz, C. C. *Endocrine-Related Cancer* **1998**, *5*, 271. b) Soucek, L.; Whitfield, J.; Martins, C. P.; Finch, A. J.; Murphy, D. J.; Sodir, N. M.; Karnezis, A. N.; Swigart, L. B.; Nasi, S.; Evan, G. I. *Nature* **2008**, *455*, 679.

3.2.3. Dedos de Zinc (Zinc Finger Protein)

Los *dedos de zinc* están formados por conjuntos de pequeños dominios dispuestos en tándem; cada uno de esos dominios independientes también se conoce por el mismo nombre de dedos de zinc. Su estructura esta estabilizada por la coordinación a un ión Zn^{2+} de forma tetraédrica a través de las cadenas laterales de aminoácidos His y Cys. Este tipo de FTs se suelen clasificar según el tipo y orden de los aminoácidos involucrados en la coordinación del zinc. El motivo más común es el Cys_2His_2 que consiste en dos láminas β antiparalelas, un giro formado por dos Cys y una hélice α que posee dos His. El zinc en estas proteínas tiene un papel estructural clave, ya que en su ausencia el motivo se encuentra desestructurado.^[50] La interacción con el ADN ocurre por la inserción de las hélices de reconocimiento en el surco mayor del ADN. La mayor parte de los dedos de cinc interactúan en forma de oligómeros—algunas proteínas contienen series de 10 o más dominios de este tipo—con el ADN, formado una hilera que se dispone a lo largo del surco mayor.^[51]

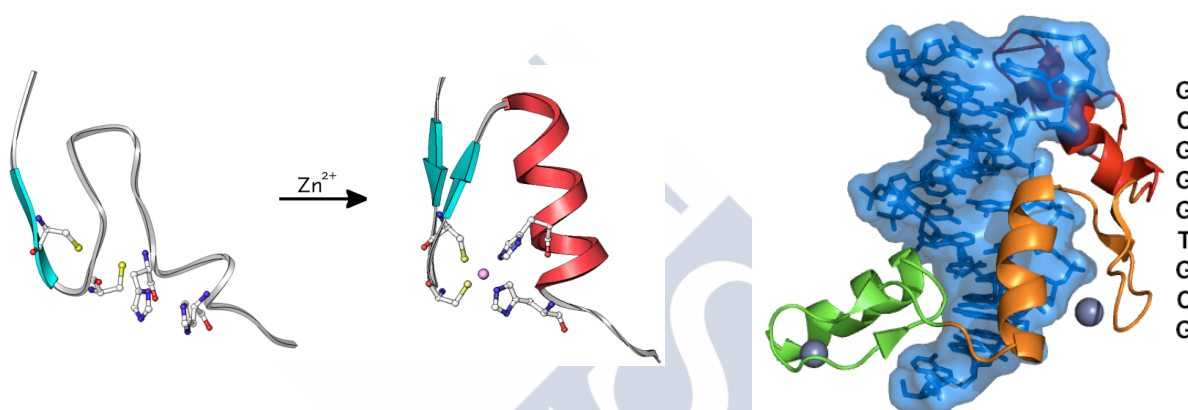


Figura 17. Izquierda: cambio conformacional inducido por la unión de Zn^{2+} en un dedo de zinc. Derecha: Estructura del dominio dedo de zinc de la proteína tipo Zif-268 (PDB: 1AAY).^[52]

[50] a) Dhanasekaran, D.; Negi, S.; Sugiura, Y. *Acc. Chem. Res.* **2006**, *39*, 45. b) Krishna, S.; Majumdar, I.; Grishin, N. *Nucleic Acids Res.* **2003**, *31*, 532.

[51] Iuchi, S.; Kuldell, N. *Zinc Finger Proteins: From atomic contact to cellular function* **2005**, Kluwer Academic/Plenum Publishers, New York, cap. 1.

[52] Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. *Structure* **1996**, *4*, 1171.

4. REGULACIÓN DE LA EXPRESIÓN GÉNICA

4.1. Mecanismos naturales de regulación génica

Los genes no pueden estar continuamente expresándose, sino que sólo generan las proteínas cuando y donde las necesitan. Esto se consigue gracias al control del proceso de la expresión génica. La regulación de este proceso se produce fundamentalmente durante la iniciación de la etapa transcripción a través de las siguientes estrategias:

- *Modificación química del ADN*: algunos genes pueden ser activados o desactivados por modificaciones químicas en las bases del ADN. La modificación más común es la adición de grupos metilo a las citosinas situadas en el extremo 5' de los genes, lo que provoca una menor transcripción de estos.^[53]
- *Modificaciones químicas en las histonas*: las cadenas laterales de los aminoácidos de las histonas están sujetas a diferentes modificaciones: acetilaciones, metilaciones o fosforilaciones. Estas alteraciones en las histonas modifican el grado de compactación de la cromatina, lo que repercute en una mayor o menor expresión de los genes situados en las regiones modificadas.^[54]
- *Regulación desde elementos proximales*: la mayoría de los genes que codifican proteínas están regulados por secuencias cortas que cambian la eficacia del promotor. Los elementos de este tipo se encuentran dentro de los 100-200 nucleótidos del punto de iniciación y se denominan elementos de control proximal. La regulación se realiza mediante la interacción de dichas secuencias con factores de transcripción, lo que produce un incremento o una disminución en la transcripción del gen.^[55]
- *Regulación desde elementos distales*: son secuencias de ADN que se encuentran en regiones lejanas al promotor y pueden activar o silenciar la expresión de un gen interaccionando con factores de transcripción específicos.^[55]
- *Acción de cofactores*: los cofactores no se unen selectivamente a secuencias específicas de ADN, pero son reclutados mediante interacciones con factores de transcripción, o con la maquinaria de transcripción general. Entre los distintos cofactores es de especial importancia la figura del *complejo mediador*, ya que este gran complejo proteico (1MDa, 22–28 subunidades) actúa como puente entre el complejo general de la transcripción, que incluye la ARNpol II y los activadores o represores que transmiten las señales desde los elementos distales y proximales del ADN (Figura 18). De este modo, la naturaleza modular del complejo mediador permite la integración de las señales reguladoras positivas y negativas.^[56]

[53] a) Slotkin, R. K.; Martienssen R. *Nature Reviews Genetics* **2007**, *8*, 272. b) Slotkin R. K.; Martienssen R. *Nature Genetics* **2003**, *33*, 245.

[54] Bannister, A. J.; Kouzarides, T. *Cell Research* **2011**, *21*, 381.

[55] a) Ptashne, M. *Nature* **1988**, *335*, 683. b) Ogbourne, S.; Antalis, T. M. *Biochem. J.* **1998**, *331*, 1.

[56] a) Kornberg, R. D. *Trends Biochem Sci.* **2005**, *30*, 235. b) Taatjes, D. J. *Trends Biochem Sci.* **2010**, *35*, 315. c) Thomas, M. C.; Chiang, C. M. *Critical Reviews in Biochemistry and Molecular Biology* **2006**, *41*, 105.

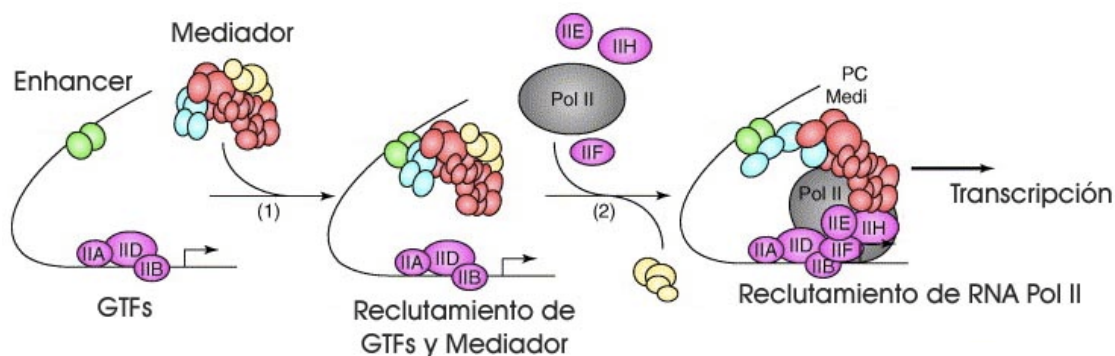


Figura 18. Representación esquemática de la función llevada a cabo por el complejo mediador en la activación transcripcional.^[57]

4.2. Regulación de la actividad de los factores de transcripción

Como se ha comentado anteriormente, en muchos casos la expresión de un determinado gen se inicia por la interacción de factores de transcripción con secuencias reguladoras específicas (proximales o distales). Por ello la presencia de factores de transcripción funcionales es determinante para que se produzca la expresión de distintos genes. **La naturaleza ejerce un control riguroso sobre la actividad de los distintos factores de transcripción, de tal forma que solo sean funcionales cuando y donde la célula los necesita.**^[58] Una de las estrategias para regular la actividad de los FT se basa en el control de su concentración, lo que se consigue regulando su síntesis. Esta táctica es económica desde un punto de vista metabólico pero tiene un tiempo de respuesta muy lento, por ello la célula suele utilizar con frecuencia estrategias ligadas a la activación controlada de factores de transcripción preexistentes, puesto que así se consigue una respuesta rápida a un determinado estímulo externo. Esta activación puede producirse de distintas formas:

- *Interacción con un ligando o proteína:* este mecanismo permite tanto la activación como la desactivación y en la mayor parte de los casos se basa en producir un cambio conformacional en el FT a través de la interacción de una molécula pequeña o una proteína. La nueva conformación puede ser más o menos activa que la inicial, e incluso a veces puede provocar que el FT reconozca una secuencia de ADN diferente. Un ejemplo de este tipo es el represor de triptófano.^[59]
- *Modificación covalente:* en esta estrategia el cambio conformacional se produce como consecuencia de la modificación covalente de alguna cadena lateral de los aminoácidos. Las modificaciones típicas son: fosforilación/desfosforilación, acetilación, metilación o ubiquitinación.^[60]

[57] Malik, S.; Roeder, R. *Trends in Biochem. Sci.* **2005**, *30*, 256.

[58] a) Lemon, B.; Tjian, R. *Genes Dev.* **2000**, *14*, 2551. b) Calkhoven, C. F.; Geert, A. B. *Biochem. J.* **1996**, *317*, 329. c) Latchman, D. S. *Eukaryotic Transcription Factors* **2004**, Elsevier Ltd, Italy; cap. 4.

[59] Sigler, P.B.; Haran, T.E.; Joachimiak, A. *EMBO J.* **1992**, *11*, 3021.

[60] Hill, C.S.; Treisman, R. *Cell* **1995**, *80*, 199.

- *Disociación de una proteína inhibidora*: táctica muy usada por la naturaleza para regular la actividad de FTs específicos, fundamentalmente su capacidad para unirse al ADN; como es el caso de NFκB, que está inhibido por IκB.^[61]
- *Degradación o fragmentación del factor de transcripción*: la proteólisis puede inducir la degradación del FT, pero mediante una proteólisis controlada también es posible la fragmentación de un precursor inactivo y consiguiente liberación de la forma activa del factor de transcripción.^[62]

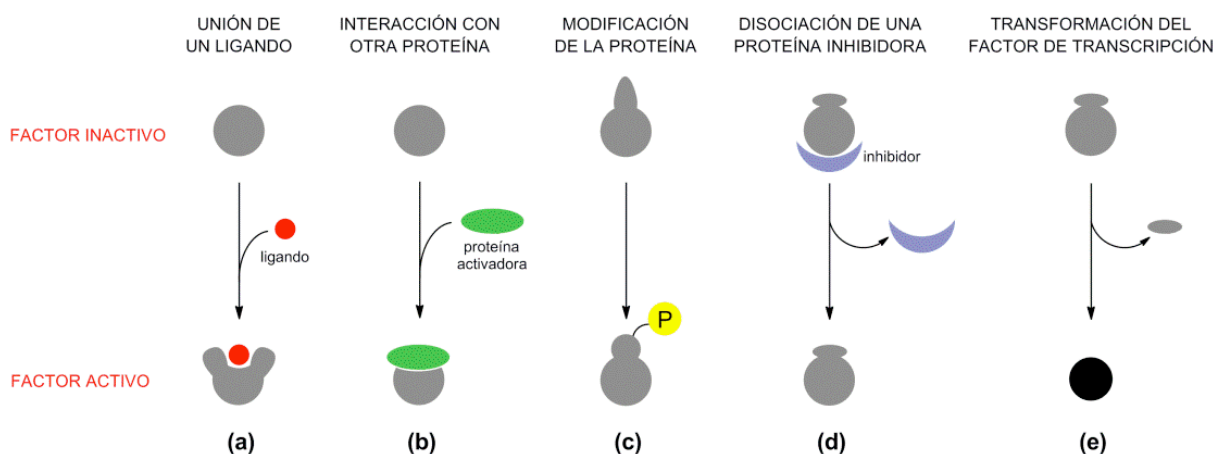


Figura 19. Métodos de activación de un factor de transcripción en respuesta a un estímulo externo concreto.

- *Localización*: los factores de transcripción al igual que el resto de proteínas se sintetizan en el citoplasma. Sin embargo para llevar a cabo su actividad, una vez sintetizados deben transportarse al núcleo a través de los poros nucleares. El control de este transporte puede ser aprovechado para regular su actividad.
- *Control combinatorio*: se entiende por control combinatorio de la transcripción como la combinación modular de un número limitado de FTs para controlar los procesos de transcripción celular de una gran cantidad de genes.^[63] Un ejemplo de este tipo de regulación lo encontramos en la familia bZIP. Por ejemplo, la proteína *Fos* homodimeriza muy débilmente, al heterodimerizar con *Jun* reconoce con elevada afinidad la secuencia AP1 (ATGAcTCAT) y si la heterodimerización la realiza con *CREB* va a unir preferentemente al sitio CRE (ATGAcgTCAT).^[64]

[61] Perkins, N.D. *Trends Biochem. Sci.* **2000**, 25, 434.

[62] a) Schmitz, M. L.; Bacher, S.; Kracht, M. *Trends Biochem. Sci.* **2001**, 26, 186. b) Pomerantz, J. L.; Baltimore, D. *Mol. Cell* **2002**, 10, 693. c) Brown, M.S.; Goldstein, J. L. *Cell* **1997**, 89, 331.

[63] Tjian, R.; Muniatis, T. *Cell* **1994**, 77, 5.

[64] Jones, N. *Cell* **1990**, 61, 9.

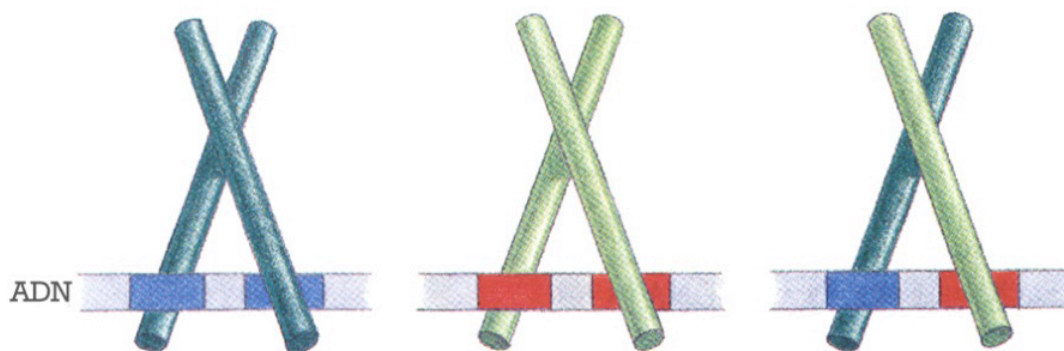


Figura 20. Representación esquemática de como el modo de dimerización influye en la selectividad de unión al ADN de FTs bZIP.^[11b] En el esquema se representa cómo mediante el control combinatorio pueden regularse tres secuencias de DNA a través de la acción de dos factores de transcripción que se asocian de diferente manera entre si.

En los últimos años se han intentado desarrollar estrategias moleculares para el control artificial de la expresión génica, pues ello podría permitir controlar determinadas actividades de la célula. Para lograr este objetivo lo ideal sería diseñar FT sintéticos que, al igual que los naturales, respondan a señales químicas o físicas externas. A pesar de la relevancia de este objetivo, el trabajo realizado hasta la fecha ha sido relativamente escaso. Sin embargo, este tipo de diseños se consideran de gran relevancia por su posible repercusión en el tratamiento de enfermedades como el cáncer o incluso el control de la diferenciación celular.

5. INTERACCIÓN DE MOLÉCULAS PEQUEÑAS CON EL ADN

Existen una serie de moléculas de pequeño tamaño que son capaces de interactuar con el ADN, pero la mayor parte lo hacen con una afinidad y especificidad relativamente bajas. Estas moléculas se pueden clasificar en cuatro familias según el tipo interacción con el ADN.^[65]

5.1. Agentes que se unen mediante contactos electrostáticos

La superficie polianiónica del ADN hace que sean necesarios cationes metálicos que neutralicen la carga para que la *forma B* sea estable (especialmente Na^+ , K^+ y Mg^{+2}). Sin embargo, también se da esta estabilización con moléculas policatiónicas como la *espermidina* y la *espermina*.^[66] Este tipo de biomoléculas poseen cadenas alifáticas con grupos amino, que a pH fisiológico se encuentran protonadas y por tanto pueden interactuar favorablemente de forma electrostática con la cadena de fosfatos del ADN. Esta interacción se ve incrementada por una contribución entrópica proveniente de la liberación de los contraiones metálicos asociados al ADN, por lo que su constante de afinidad por el ADN es típicamente muy alta ($2.3 \times 10^5 \text{ M}^{-1}$ para la espermina y $1.4 \times 10^5 \text{ M}^{-1}$ para la espermidina).^[67]

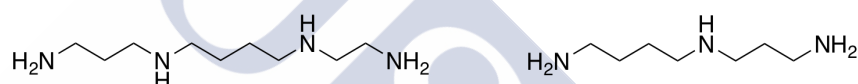


Figura 21. Izquierda: Estructura de la espermina. Derecha: Estructura de la espermidina.

Existen también un grupo de moléculas sintéticas basadas en complejos de Pt(II), conocidas como “*phosphate clamps*”, que interactúan con el ADN a través de los fosfatos de una forma análoga a las oligoaminas naturales. Este tipo de moléculas tienen una estructura similar a las poliaminas naturales en las que los grupos aminos son sustituidos por complejos de platino con carga +2.^[68]

[65] a) Wemmer, D.E. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439. b) Chaires, J.B. *Curr. Opin. Struct. Biol.* **1998**, *8*, 314.

[66] a) Deng, H.; Bloomfield, V. A.; Benevides, J. M.; Thomas, G. J. *Nucleic Acids Research* **2000**, *28*, 3379. b) Marx, K. A.; Reynolds, T. C. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6484.

[67] Ouameur, A. A.; Tajmir-Riahi, H. J. *Biol. Chem.* **2004**, *279*, 42041.

[68] a) Komeda, S.; Moulaei, T.; Woods, K. K.; Chikuma, M.; Farrell, N. P.; Williams, L. D. *J. Am. Chem. Soc.* **2006**, *128*, 16092. b) Rauter, H.; Domenico, R. D.; Menta, E.; Oliva, A.; Qu, Y.; Farrell, N. *Inorg. Chem.* **1997**, *36*, 3919. c) Komeda, S.; Moulaei, T.; Chikuma, M.; Odani, A.; Kipping, R.; Farrell, N. P.; Williams, L. D. *Nucleic Acids Res.* **2011**, *39*, 325.

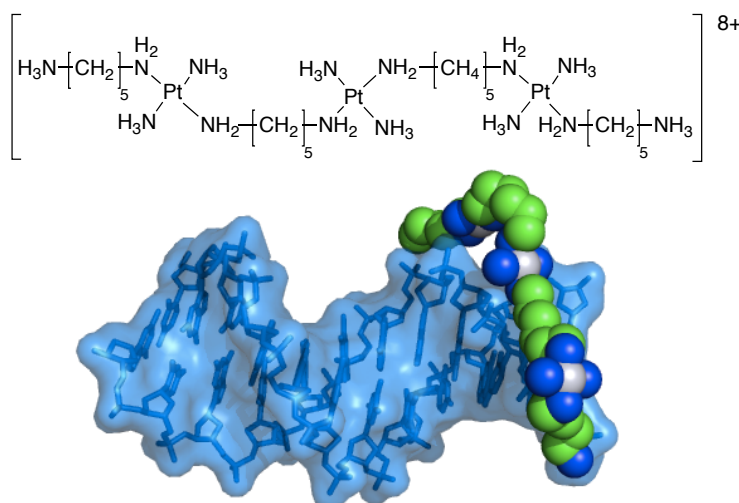


Figura 22. Arriba: Estructura de un complejo de platino de la familia de los “*phosphate clamps*”. Abajo: Estructura de rayos X del complejo de platino interactuando con el ADN (PDB: 2DYW).^[68a]

5.2. Agentes intercalantes

Estas moléculas se caracterizan por poseer sistemas aromáticos extendidos deficientes en carga. En su interacción con el ADN los agentes intercalantes se insertan entre pares de bases consecutivos de la doble cadena de ADN. Este tipo de interacción se estabiliza por una combinación de factores: apilamientos π , transferencia de carga, enlaces de hidrógeno e interacciones electrostáticas.^[69] La intercalación conlleva una deformación del ADN, que se alarga unos 3 Å por molécula insertada. Aunque éstas moléculas se unen con una cierta preferencia a pares G/C, no suelen tener una selectividad de secuencia clara.^[70]

Este tipo de moléculas se usan como agentes antitumorales debido a que atrapan a la topoisomerasa de tipo II en un complejo ternario con el ADN, el intercalante y la proteína.^[71] Además son ampliamente usados como marcadores de ácidos nucleicos debido a que algunas moléculas aromáticas se vuelven fluorescentes al intercalarse, un ejemplo de esto es el *bromuro de etidio*. En muchos casos se suelen dimerizar dos agente intercalantes para mejorar su afinidad y propiedades, como por ejemplo el *TOTO*, el cual es un *bis*-intercalante selectivo para ADN de doble cadena.^[72]

[69] a) Nakamoto, K.; Tsuboi, M.; Strahan, G. D. *Intercalating Drugs, in Drug-DNA Interactions: Structures and Spectra* **2008**, John Wiley & Sons, Inc., Hoboken, New York, USA. b) Rescifina, A.; Zagni, Z.; Varrica, M. G.; Pistrà, V.; Corsaro, A. *European Journal of Medicinal Chemistry* **2014**, *74*, 95.

[70] a) Neto, B. A. D.; Lapis, A. A. M. *Molecules* **2009**, *14*, 1725. b) Berman, H.M.; Young, P. R. *Annu. Rev. Biophys. Bioeng.* 1981, *10*, 87-114.

[71] Braña, M. F.; Cacho, M.; Gradillas, A.; de Pascual, T. B.; Ramos, A. *Curr Pharm Des.* **2001**, *7*, 1745.

[72] Bunkenborg, J.; Stidsen, M. M.; Jacobsen, J. P. *Bioconjug Chem.* **1999**, *10*, 824.

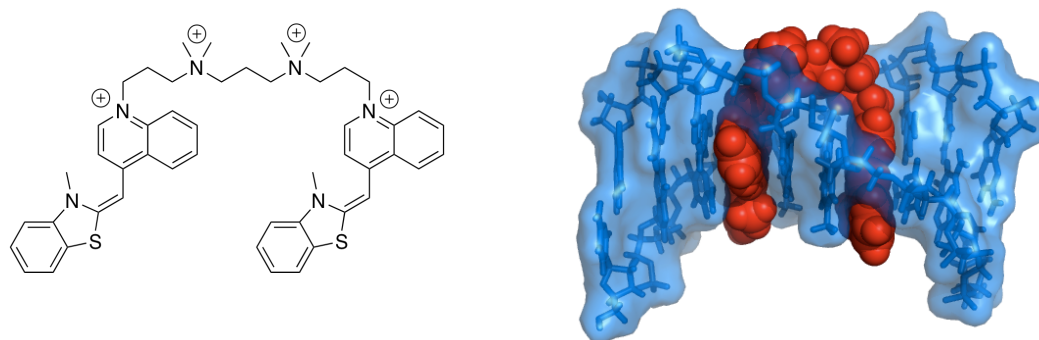


Figura 23. Izquierda: Estructura del bis-intercalante TOTO. Derecha: Estructura obtenida por RMN de TOTO interaccionando con ADN de doble cadena (PDB: 108D).^[73]

5.3. Agentes que se unen de forma covalente al ADN

Son moléculas que se unen al ADN mediante enlaces covalentes con las bases. La posición más susceptible de alquilación es el N⁷ de la guanina. Sin embargo también el C²-NH₂ de la guanina y los átomos de nitrógeno N³ y N⁷ de la adenina pueden participar en este proceso.^[74] Algunas de estas moléculas se unen a más de una base, lo que produce un entrecruzamiento entre hebras del ADN que impide su replicación. Este tipo de moléculas son muy usadas para tratar procesos tumorales y existe una gran variedad.^[75]

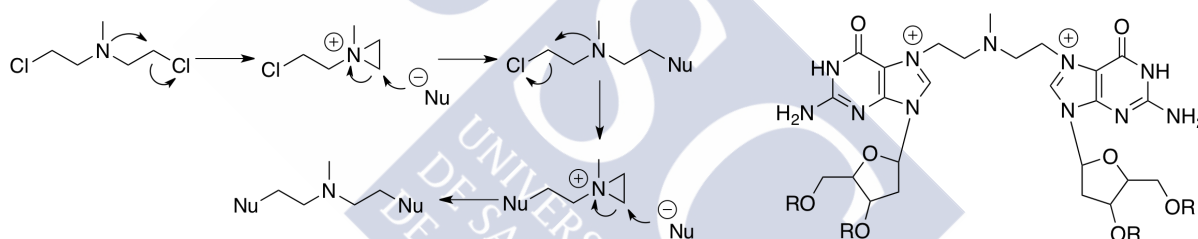


Figura 24. Izquierda: mecloretamina y mecanismo de alquilación. Derecha: estructura resultante de la alquilación de guaninas debido a la mecloretamina.^[76]

Además de moléculas orgánicas, existen un grupo de complejos metálicos que se unen a las bases del ADN de forma irreversible. La unión suele conllevar la coordinación entre el átomo N⁷ de la guanina y el metal. La mayor parte de este tipo de complejos están basados en platino, aunque también hay ejemplos con rutenio y paladio. Uno de los análogos más conocidos es el *cisplatino* y el *carboplatino*, que son ampliamente usados en el tratamiento del cáncer.^[77]

[73] Spielmann, H. P.; Wemmer, D. E.; Jacobsen, J. P. *Biochemistry* **1995**, *34*, 8542.

[74] Rajski, S. R.; Williams, R. M. *Chem. Rev.* **1998**, *98*, 2723.

[75] a) Fu, D.; Calvo, J. A.; Samson, L. D. *Nature Reviews Cancer* **2012**, *12*, 104. b) Kondo, N.; Takahashi, A.; Ono, K.; Ohnishi, T. *Journal of Nucleic Acids* **2010**, *1*, 543531.

[76] a) Brookes, P.; Lawley, P. D. *Biochem. J.* **1961**, *80*, 496. b) Balcome, S.; Park, S.; Dorr, D. R. Q.; Hafner, L.; Phillips, L.; Tretyakova, N. *Chem. Res. Toxicol.* **2004**, *17*, 950.

[77] Alderden, R. A.; Hall, M. D.; Hambley, T. W. *J. Chem. Educ.* **2006**, *83*, 728.

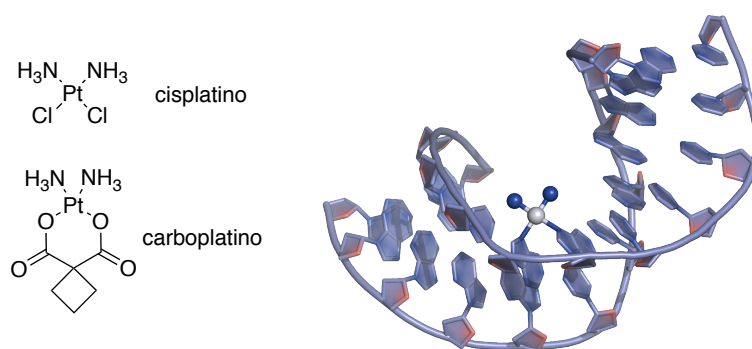


Figura 25. Izquierda: estructura del cisplatino y carboplatino. Derecha: estructura resultante de la alquilación de guaninas debido al cisplatino (PDB: 3LPV).

5.4. Moléculas que se insertan en el surco menor

Existen una gran variedad de moléculas pequeñas capaces de interactuar con el ADN insertándose en su surco menor.^[78] Dicha inserción está mediada por la acción de diferentes fuerzas intermoleculares, incluyendo enlaces de hidrógeno, contactos de van der Waals e hidrofóbicos—debido al desplazamiento de moléculas de agua en la inserción—e interacciones electrostáticas. Además la mayor parte de estas moléculas presentan una forma cóncava que favorece su acoplamiento con la geometría del surco menor.^[79]

Las moléculas que interactúan con el surco menor suelen tener preferencia por la inserción en regiones ricas en A/T, debido a que los surcos menores en esas regiones son más estrechos y profundos que en las regiones ricas en pares G/C y así se maximizan las interacciones de van der Waals.^[65] Además, en las regiones ricas en pares G/C el grupo amino de las guaninas sobresale en el fondo del surco menor, interfiriendo estéricamente con la inserción.^[80] Dos ejemplos clásicos de este tipo de moléculas son la distamicina A y la netropsina (Figura 26).^[81]

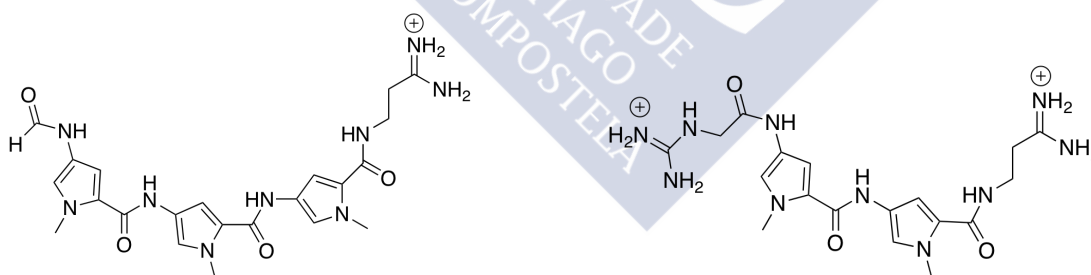


Figura 26. Izquierda: Estructura de la distamicina. Derecha: Estructura de la netropsina.

- [78] Baraldi, P. G.; Bovero, A.; Fruttarolo, F.; Preti, D.; Tabrizi, M. A.; Pavani, M. G.; Romagnoli, R. *Medicinal Research Reviews* **2004**, *24*, 475.
- [79] Nguyen, B.; Lee, M. P. H.; Hamelberg, D.; Joubert, A.; Bailly, C.; Brun, R.; Neidle, S.; Wilson, D. W. *J. Am. Chem. Soc.* **2002**, *124*, 13680.
- [80] Hamelberg, D.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2001**, *123*, 7745.
- [81] Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A* **1985**, *82*, 1376.

Aunque la distamicina A interacciona con buena afinidad con secuencias ricas en adenina o timina, estudios posteriores demostraron que dependiendo de la secuencia es posible la inserción de dos moléculas de distamicina de forma antiparalela.^[82] Posteriormente también se comprobó que la sustitución de los grupos pirrol (Py) de la distamicina por diferentes heterociclos, tal como el imidazol (Im), facilita la interacción con pares GC (Figura 27).^[83] Basándose en estos hallazgos el grupo de Peter Dervan diseñó sistemas de reconocimiento formados por dos unidades de poliheterociclos unidas covalentemente que pueden reconocer un gran abanico de secuencias de ADN de forma específica.^[84]

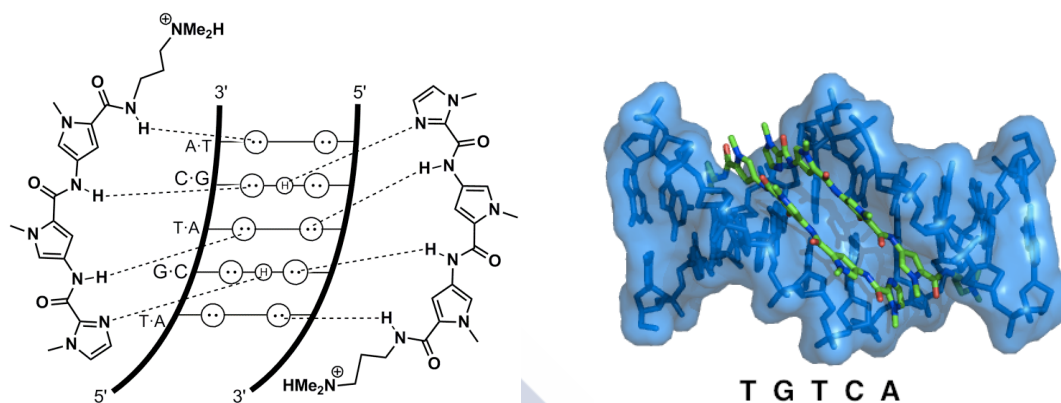


Figura 27. Izquierda: Diagrama de la interacción de dos derivados de la distamicina que se insertan en el surco menor de forma antiparalela. Derecha: Estructura cristalográfica de ambos derivados interaccionado con el ADN (PDB: 365D).

Las mayores limitaciones de estas poliamidas derivan de las dificultades sintéticas y de que dichas moléculas se internalizan en las células de forma deficiente,^[85] lo que complica su aplicación práctica. Nuestro grupo de investigación ha centrado su interés en la familia de las bisbenzamidas, más simples desde el punto de vista estructural y con buenas propiedades de transporte intracelular.^[86] Un ejemplo de esta clase de moléculas es la *propamidina*, que se usa para el tratamiento de varios tipos de infecciones provocadas por protozoos.^[87]

- [82] a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. c) Dwyer, T. J.; Geierstanger, B.H.; Bathid, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911.
- [83] a) Goodsell, D. S.; Dickerson, R. E. *J. Med. Chem.* **1986**, *29*, 727. b) Goodsell, D. S.; Ng, H. L.; Kopka, M. L.; Lown, J. W.; Dickerson, R.E. *Biochemistry* **1995**, *34*, 16654. c) Chen, Y-H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995.
- [84] a) Demeunynck, M.; Bailly, C.; Wilson, W. D. *Small Molecule DNA and RNA Binders* **2003**, WILEY-VCH, Weinheim, cap. 15. b) Dervan, P. B. *Science* **1986**, *232*, 464. c) Bailly, C.; Chaires, J. B. *Bioconj. Chem.* **1998**, *9*, 513. d) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284.
- [85] a) Dervan, P. B. *Top. Curr. Chem.* **2005**, *253*, 1. b) Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M. Foister, S.; Heckel, A.; Dervan, P. B. *Nucl. Acids Res.* **2004**, *32*, 2802. c) Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12063.
- [86] a) Susbielle, G.; Blattes, R.; Brevet, V.; Monod, C.; Kas, E. *Curr. Med. Chem. Anticancer Agents* **2005**, *5*, 409; b) Reddy, B. S. P.; Sondhi, S. M.; Lown, J. W. *Pharm. Ther.* **1999**, *84*, I-III. c) Neidle, S.; Kelland, L. R.; Trent, J. O.; Simpson, I. J.; Boykin, D. W.; Kumar, A.; Wilson, W. D. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1403.
- [87] Perrine, D.; Chenu, J. P.; Georges, P.; Lancelot, J. C.; Saturnino, C.; Robba, M. *Antimicrob. Agents Chemother.* **1995**, *39*, 339.

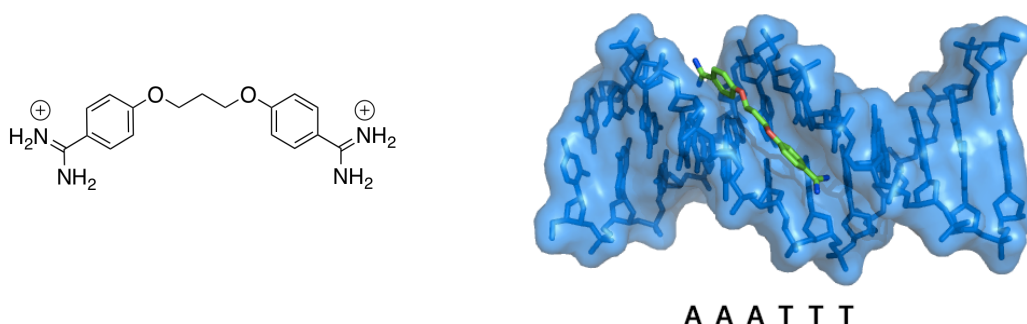


Figura 28. Izquierda: Estructura de la propamidina. Derecha: Estructura de rayos X del complejo propamidina y ADN (PDB: 102D).^[88]

En nuestro grupo se demostró que la sustitución de los átomos de oxígeno de la propamidina por átomos de nitrógeno, da lugar a nuevos agentes capaces de unirse al ADN y que presentan interesantes propiedades fluorescentes. El análogo de la propamidina, denominado BAPPA (*N*¹,*N*³-bis(4-amidinofenil)propano-1,3-diamina) interacciona con alta especificidad con secuencias ricas en A/T y al unirse al ADN experimenta un aumento de más de diez veces en su fluorescencia respecto a la moléculas en disolución acuosa.^[89]

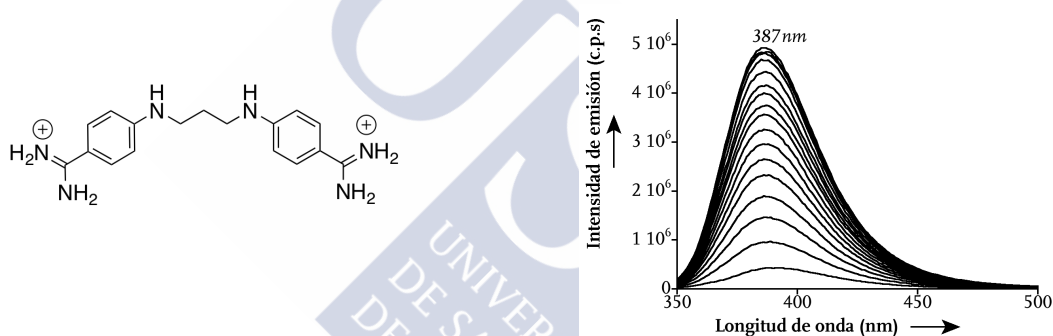


Figura 29. Izquierda: Estructura de BAPPA. Derecha: Espectros de fluorescencia de la valoración de BAPPA (0.5 μ M en Tris•HCl 20 mM pH 7.5, NaCl 100 mM) con una ADN de doble cadena con la secuencia AATTT.^[89]

Además del análogo BAPPA se han sintetizado en nuestro grupo otro tipo de compuestos con diferentes conectores entre las dos unidades de benzamidina. Esta familia de compuestos pueden ser prepararse en una sola etapa a partir de sustancias comerciales y son fácilmente derivatizables (Figura 30).^[90]

[88] Nunn, C. M.; Neidle, S. *J. Med. Chem.* **1995**, *38*, 2317.

[89] Vázquez, O.; Sánchez, M. I.; Martínez-Costas, J.; Vázquez, M. E.; Mascareñas, J. L. *Org Lett.* **2010**, *12*, 216.

[90] a) Sánchez, M. I.; Vázquez, O.; Martínez-Costas, J.; Vázquez, M. E.; Mascareñas, J. L. *Chem. Sci.*, **2012**, *3*, 2383. b) Sánchez, M. I.; Vázquez, O.; Vázquez, M. E.; Mascareñas, J. L. *Chem. Eur. J.* **2013**, *19*, 9923.

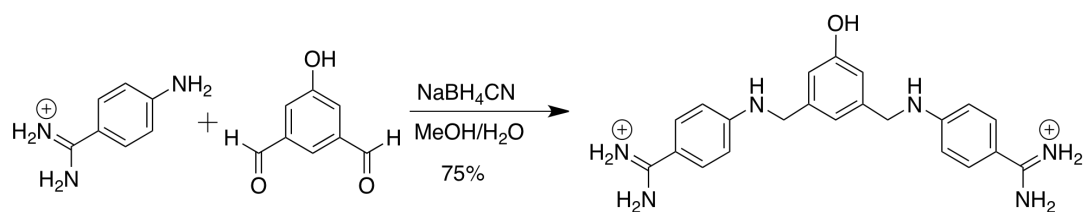


Figura 30. Ejemplo de la síntesis de análogos de bisbenzamidinas en una sola etapa.





6. RECONOCIMIENTO DEL ADN MEDIANTE PÉPTIDOS SINTÉTICOS

En los últimos años ha habido un gran interés en el diseño de sistemas sintéticos no naturales capaces de emular a los factores de transcripción. La mayor parte de este tipo de moléculas consiste en versiones simplificadas de FTs. A pesar de que se han logrado algunos avances, podría decirse que el progreso en este campo ha sido muy lento.^[91]

En la mayor parte de los casos tanto los péptidos diseñados de *novo*, como los obtenidos a partir del truncamiento de secuencias de FTs naturales, suelen mostrar baja afinidad y especificidad por el ADN. El hecho de que no exista un código de reconocimiento entre los factores de transcripción y el ADN, hace extremadamente difícil diseñar alternativas sintéticas. Aunque podría pensarse que la eliminación de residuos de los factores de transcripción naturales alejados del dominio de unión al ADN no debería influir en la interacción, en la mayoría de los casos se ha comprobado que estos son también decisivos para la formación de los complejos con el ADN.^[92] Por lo tanto, para el diseño de análogos sintéticos de FTs es inevitable recurrir a estrategias del tipo ensayo y error. Algunas de las aproximaciones más exitosas obtenidas hasta la fecha se comentan a continuación.

6.1. Dimerización artificial de proteínas bZIP

El grupo de Peter Kim demostró en 1990 que la región denominada cremallera de leucinas de los FTs del tipo bZIP puede ser sustituida por otro tipo de grupos dimerizantes, como por ejemplo un simple enlace disulfuro, sin que se vean alteradas significativamente las propiedades de interacción con el ADN. Así pues, el dímero disulfuro de la región básica del factor de transcripción GCN4, modificada con la secuencia Gly-Gly-Cys como conector C-terminal, reconoce su secuencia consenso de ADN, AP1 o CRE de forma específica, aunque con una afinidad de 10 veces menor que la proteína natural.^[93]



Figura 31. Izquierda: representación esquemática de GCN4. Derecha: modelo hipotético de la interacción del dímero disulfuro diseñado por el grupo de Kim.

[91] Ansari, A. Z.; Mapp, A.K. *Curr. Opin. Chem. Biol.* **2002**, 6, 765.

[92] Pabo, C. O.; Peisach, E.; Grant, R. A. *Annu. Rev. Biochem.* **2001**, 70, 313.

[93] Talanian, R. V.; McKnight, C.J.; Kim, P. S. *Science* **1990**, 249, 769.

El grupo de Kim también realizó un estudio en el que demostró que es incluso posible reducir el tamaño de esta región básica del GCN4 hasta 23 aminoácidos, manteniendo una buena afinidad por el ADN (en torno a 10 nM a 4 °C). En este estudio queda patente que la eliminación de parte de la secuencia en el extremo C-terminal (RMKQ) afecta menos a la estabilidad del complejo final que la eliminación del extremo N-terminal (Tabla 1).^[94] Los residuos N-terminales DPAAL no establecen contactos con el ADN, pero estabilizan la estructura helicoidal y con ello facilitar la interacción.^[95]

| Péptido | Secuencia | AP-1 ATGAcTCAT | CRE ATGAcgTCAT |
|----------------------------------|---------------------------------|-------------------|-------------------|
| GCN4-rb1 (región básica natural) | PESSDPAALKRARNTAAARRSRARKLQRMKQ | + | + |
| GCN4-rb2 | DPAALKRARNTAAARRSRARKLQRMKQ | + | + |
| GCN4-rb3 | AALKRARNTAAARRSRARKLQRMKQ | - | + |
| GCN4-rb4 | DPAALKRARNTAAARRSRARKLQ | + | + |
| GCN4-rb5 | DPAALKRARNTAAARRSRAR | - | + |

Tabla 1. Resumen de la investigación llevada por Kim para crear versiones truncadas de GCN4 (“+” significa que conserva la capacidad de reconocimiento por esa secuencia, “-” significa que no conserva la capacidad de reconocimiento).

Después del descubrimiento de P. Kim, varios grupos de investigación emplearon estrategias similares, demostrando que las regiones básicas de GCN4 podían ser dimerizadas empleando diferentes tipos de conectores:

- *Dimerización mediante complejos metálicos con terpiridinas:* Alanna Schepartz en 1993 publicó dímeros basados en complejos de terpiridina con Fe(II). Comparados a sus análogos dimerizados a través de un disulfuro, este tipo de complejos mantienen la misma afinidad por el sitio de unión CRE, pero presentan una menor capacidad de unión a API.^[96]
- *Dimerización mediante complejos ciclodextrina-adamantano:* el grupo de Takashi Morii publicó en 1995 un ejemplo de dimerización supramolecular de las regiones básicas de GCN4 basada en la complejación de un adamantano y una ciclodextrina en los extremos C-terminales de regiones básicas de proteínas bZIP.^[97] Esta estrategia permitió además hacer heterodimerizaciones entre dos regiones básicas de diferentes FTs del tipo bZIP, concretamente entre la región básica del GCN4 y la del C/EBP.^[98]
- *Dimerización mediante moléculas de ácido peptidonucleicos (APN):* los ácidos peptidonucleicos (APN) son moléculas poliméricas que combinan un esqueleto de unidades de N-(2-aminoetil)-glicina unidas por enlaces peptídicos con bases nitrogenadas como las de los oligonucleótidos. La falta de fosfatos elimina la repulsión electrostática entre las cadenas por lo que la unión entre bases complementarias es más fuerte que en el caso del ADN o ARN. El grupo de Romanelli, en 2009, usó estos análogos de ácidos nucleicos para dimerizar dos

[94] Talanian, R. V.; McKnight, C. J.; Rutkowski, R.; Kim, P. S. *Biochemistry* **1992**, *31*, 6871.

[95] Hollenbeck, J. J.; McClain, D. L.; Oakley, M. G. *Protein Sci.* **2002**, *11*, 2740.

[96] a) Cuenoud, B.; Schepartz, A. *Science* **1993**, *259*, 510. b) Palmer, C. R.; Sloan, S. S.; Adrian, J. C.; Cuenoud, B.; Paoletta, D. N.; Schepartz, A. *J. Am. Chem. Soc.* **1995**, *117*, 8899.

[97] a) Ueno, M.; Murakami, A.; Makino, K.; Morii, T. *J. Am. Chem. Soc.* **1993**, *115*, 12575. b) Morii, T.; Yamane, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* **1996**, *118*, 10011.

[98] Ueno, M.; Sawada, M.; Makino, K.; Morii, T. *J. Am. Chem. Soc.* **1994**, *116*, 11137.

regiones básicas de GCN4. Con esta estrategia se pudo obtener una K_d para CRE de 5 nM a 4 °C.^[99]

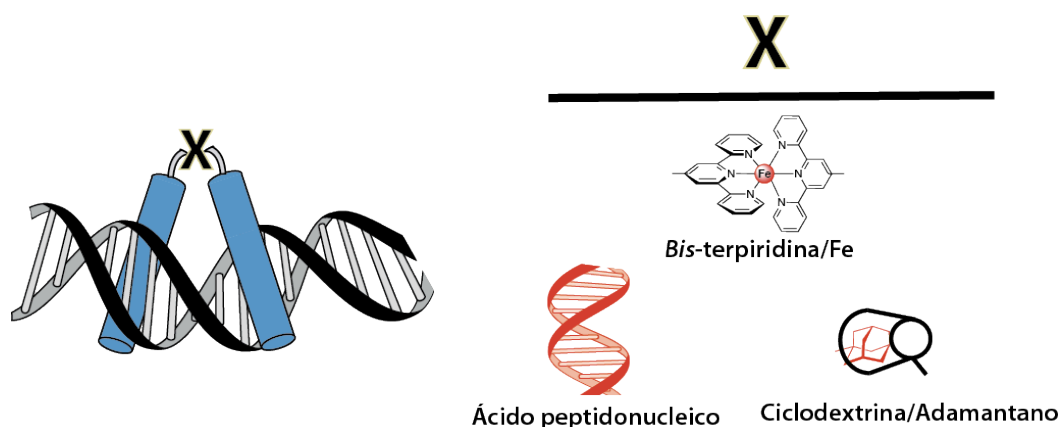


Figura 32. Representación esquemática de las diferentes estrategias de dimerización usadas con las regiones básicas de GCN4. X representa los agentes dimerizantes que se muestran en la parte derecha.

Además de la dimerización a través del extremo *C*-terminal, el grupo de Morii también ha diseñado dímeros covalentes a través del extremo *N*-terminal. A diferencia de los anteriores, estos reconocen secuencias de ADN “inversas” respecto a la secuencia natural de GCN4, tales como (5'-TCAT cg ATGA -3). En estos análogos, Morii obtuvo los mejores resultados con conectores rígidos colocados en la posición δN -terminal (**N6R** y **N6S**).^[100]

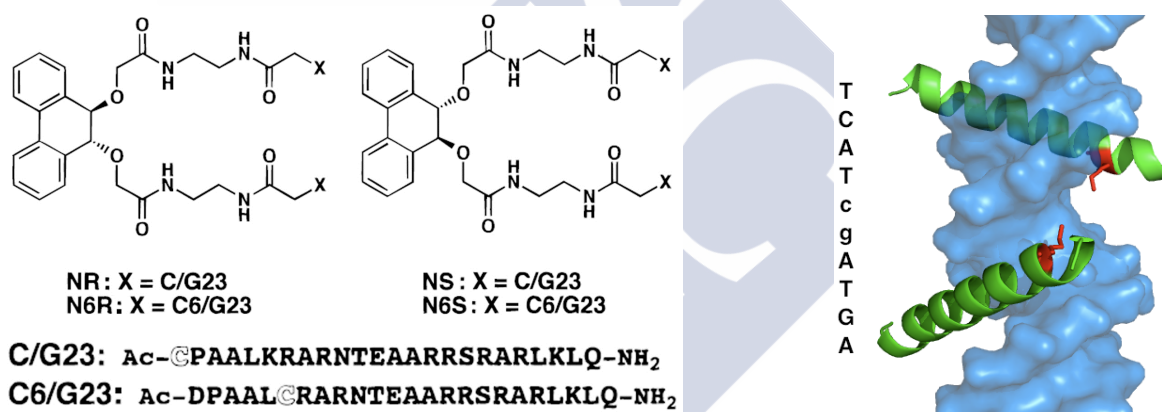


Figura 33. Izquierda: Representación esquemática de dímeros *N*- y δN -terminales de la región básica de GCN4 desarrollados por el grupo de Morii, en los que se observa la secuencia modificada convenientemente con residuos de cisteína, donde se une el agente dimerizador.^[100] Derecha: Modelo de interacción para los dímeros **N6**, la cisteína usada para unir el conector se muestra en rojo.

[99] Pensato, S.; Renda, M.; Leccia, F.; Saviano, M.; D’Andrea, L. D.; Pedone, C.; Pedone, R. V.; Romanelli, A. *Biopolymers* **2009**, 93, 434.

[100] Morii, T.; Saimei, Y.; Okagami, M.; Makino, K.; Sugiura, Y. *J. Am. Chem. Soc.* **1997**, 119, 3649.

6.2. Motivos helicoidales estabilizados

La mayor parte de los FTs interactúan con el ADN a través de una hélice α que se inserta en el surco mayor. Normalmente esta hélice sólo se encuentra plegada cuando interactúa con el ADN y por lo tanto este reconocimiento conlleva un coste entrópico considerable. Con el fin de disminuir dicho coste, varios grupos han desarrollado estrategias para estabilizar dicha hélice y con ello favorecer la unión al ADN de péptidos pequeños.

6.2.1. Uso de estructuras plantilla preorganizadas (grafting de residuos)

Esta aproximación, desarrollada por el grupo de A. Schepartz, se basa en el uso de una miniproteína estructurada como soporte para la estabilización de la hélice de reconocimiento. Concretamente, la proteína aPP (polipéptido pancreático de ave) presenta una hélice α muy estabilizada mediante interacciones hidrofóbicas con una cadena de poliprolinas. Sustituyendo determinados aminoácidos de la hélice por aquellos necesarios para la interacción con el ADN, este grupo desarrolló varios péptidos que muestran muy buena afinidad por secuencias de 4-6 pares de bases. El mayor impedimento para la aplicación de esta técnica, es la dificultad de predecir en que medida la sustitución de ciertos residuos puede llegar a afectar a la funcionalidad de la proteína final.^[101] Esta estrategia se ha aplicado con éxito a dos factores de transcripción, el GCN4 de la familia bZip^[101] y el homeodominio Engrailed.^[102] Especialmente interesante fue el resultado obtenido con la proteína GCN4, ya que el péptido monomérico obtenido **PPrb**, es capaz de interactuar con la secuencia consenso ATGAC con una constante de afinidad de 5 nM a 4 °C sin necesidad de dimerización.



Figura 34. Izquierda: Estructura del dominio de unión del FT GCN4 complejo al sitio AP1 (5'-ATGAcTCAT-3', PDB: 1YSA) y estructura superpuesta de la miniproteína aPP (en naranja, PDB: 1PPT). Arriba a la derecha: Alineamiento de las secuencias de aPP y la región básica de GCN4. Abajo a la derecha: Secuencia de **PPrb**. Se muestran en rojo los aminoácidos implicado en el reconocimiento del ADN y en negrita los aminoácidos sustituidos por alanina para facilitar el plegamiento en hélice α .

[101] a) Zondlo, N. J.; Schepartz, A. *J. Am. Chem. Soc.* **1999**, *121*, 6938. b) Chin, J.; Schepartz, A. *J. Am. Chem. Soc.* **2001**, *123*, 2929.

[102] Montclare, J. K.; Schepartz, *J. Am. Chem. Soc.* **2003**, *125*, 3416.

6.2.2. Restricción conformacional (péptidos grapados)

Esta aproximación se basa en hacer uniones covalentes entre determinados aminoácidos con el fin de facilitar el plegamiento de la cadena peptídica en una hélice α .^[103] Así pues, el grupo de Taylor sintetizó un análogo de la región básica de GCN4 con un puente lactama en el extremo C-terminal entre las cadenas laterales de la Lys25 y el Asp29 ($i, i+4$).^[104] Más tarde el grupo de M. Inouye aplicó esta estrategia al FT Engrailed, en este caso usando como agente entrecruzante una molécula orgánica rígida con dos grupos acetileno, unido a dos residuos de ornitina en posiciones i y $i+11$ de la hélice, obteniendo péptidos con constantes de disociación por el ADN menores de nanomolar y muy resistentes a proteasas.^[105]

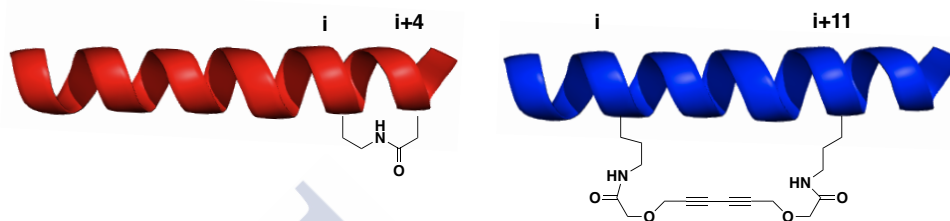


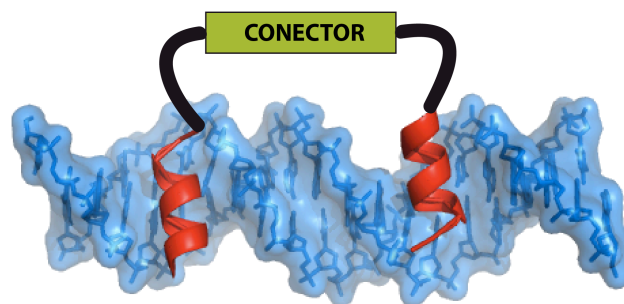
Figura 35. Izquierda: Estrategia usada por el grupo de Taylor para obtener una versión del estabilizada del FT GCN4. Derecha: Estrategia usada por Inouye con el FT Engrailed.

6.2.3. Estabilización mediante aminoácidos inductores de helicidad

Otro modo de estabilización se basa en el uso de aminoácidos no naturales con gran tendencia a plegar en forma de hélice α por sus restricciones conformacionales.^[106]

Como ejemplo de esta aproximación, el grupo de Siddhartha Roy usó el aminoácido ácido 2-aminoisobutírico (Aib) para hacer una versión simplificada de la proteína *Cro* del bacteriófago λ . Este FT interactúa con el ADN como homodímero, por lo que se sintetizó un dímero covalente de dos hélices de reconocimiento, cada una de ellas estabilizada por tres aminoácidos Aib situados en la cara de la hélice no implicada en el reconocimiento del ADN (Figura 36).^[107]

- [103] Araujo, A. D.; Hoang, H. N.; Kok, W. M.; Diness, F.; Gupta, P.; Hill, T. A.; Driver, R. W.; Price, D. A.; Liras, S.; Fairlie, D. P. *Angew. Chem. Int. Ed.* **2014**, *53*, 6965.
- [104] a) Zhang, M.; Wu, B.; Baum, J.; Taylor, J. W. *J. Peptide Res.* **2000**, *55*, 398. b) Zhang, M.; Wu, B.; Zhao, H.; Taylor, J. W. *Peptide Sci.* **2002**, *8*, 125.
- [105] Kajino, M.; Fujimoto, K.; Inouye M. *J. Am. Chem. Soc.* **2011**, *133*, 656.
- [106] Mahalakshmi, R.; Balaram, P. *Methods Mol. Biol.* **2006**, *340*, 71.
- [107] Mazumder, A.; Maiti, A.; Roy, K.; Roy, S. *ACS Chem. Biol.* **2012**, *7*, 1084.



Hélice natural: Y Q S A I N K A I H A G R
Hélice estabilizada: Y Q S A B N K A B H B G R C G

Figura 36. Arriba: Modelo de la versión simplificada de la proteína *Cro* diseñado por Siddhartha Roy basada en la estructura cristalográfica del factor *Cro* (PDB: 4CRO).^[108] La hélice de reconocimiento se muestra en rojo. Abajo: Secuencia de la hélice de reconocimiento de la proteína natural y de la versión estabilizada. La B en rojo representa al aminoácido 2-aminoisobutírico (Aib), la C en azul es el aminoácido al que se une el conector.

6.3. Conjugados entre fragmentos de FTs y moléculas que se unen al surco menor del ADN

En general los dominios de reconocimiento de los FT pierden su capacidad de interacción con el ADN cuando se aíslan del resto de la proteína.^[109] Sin embargo, en nuestro grupo se ha desarrollado una estrategia que permite recuperar la capacidad de unión de fragmentos peptídicos de dominios de reconocimiento, consistente en la conjugación a moléculas sintéticas pequeñas que se unen con selectividad y afinidad moderada al surco menor en sitios adyacentes.

Los primeros ejemplos de esta aproximación hacían uso de la hélice de reconocimiento del GCN4 y polipirroles derivados de la distamicina A (Figura 37). Estos híbridos bivalentes reconocen selectivamente sitios de unión de 9 pares de bases compuestos por parte de la secuencia AP1 de unión del péptido (ATGA) y del polipirrol (AAATT), con una afinidad en el rango de nanomolar. Sin embargo, estos híbridos presentan baja selectividad y también se unen con notable afinidad a secuencias mutadas en la zona de reconocimiento del péptido. Esto puede ser debido al gran número de residuos básicos en la región peptídica, los cuales interactúan electrostáticamente con los fosfatos del ADN mientras el polipirrol se inserta en su secuencia consenso.^[110]

[108] Brennan, R. G.; Roderick, S. L.; Takeda, Y.; Matthews, B. W.; *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8165.

[109] a) Nelson, H. C. *Curr. Opin. Genet. Dev.* **1995**, *5*, 180. b) Moretti, R.; Ansari, A. Z. *Biochimie* **2008**, *90*, 1015. c) Segal, D. J.; Barbas III, C. F. *Curr. Opin. Chem. Biol.* **2000**, *4*, 34.

[110] a) Vázquez, M. E.; Caamaño, A. M.; Castedo, L.; Gramberg, D.; Mascareñas, J. L. *Tetrahedron Letters* **1999**, *40*, 3625. b) Vázquez, M. E.; Caamaño, A. M.; Martínez-Costas, J.; Castedo, L.; Mascareñas, J. L. *Angew. Chem. Int. Ed.* **2001**, *40*, 4723.



Figura 37. Izquierda: Modelo de interacción simultánea en surcos vecinos. Derecha: Estructura general de los híbridos surco mayor-surco menor, basados en regiones básicas de proteínas bZIP.

Esta estrategia también se aplicó a la familia de FTs de los dedos de zinc, concretamente a la proteína GAGA.^[111] En este caso fue necesario usar el núcleo de la proteína, formado por la hélice de reconocimiento y dos láminas beta, para obtener un híbrido con capacidad de reconocimiento de ADN (Figura 38). A pesar de esta menor simplificación, se logró reducir a la mitad el número de aminoácidos que componían la proteína natural. La interacción en este caso resultó ser mucho más específica que en el caso del GCN4 y la constante de disociación obtenida fue 2.5 nM a 20 °C, siendo una de las mejores afinidades logradas hasta la fecha por miméticos artificiales de factores de transcripción.^[112]

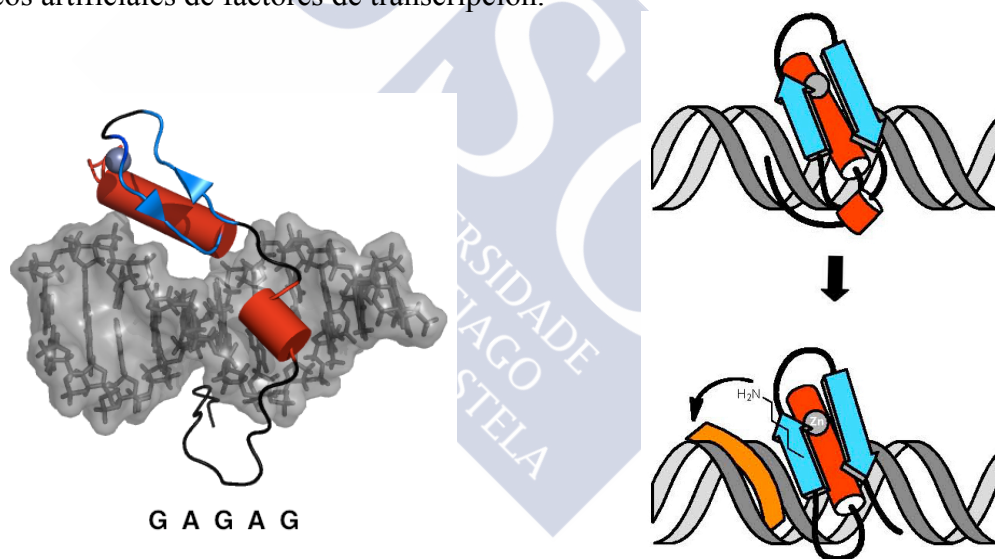


Figura 38. Izquierda: estructura del complejo entre el ADN y la proteína GAGA (PDB: 1YUI). Derecha: Representación esquemática de la estrategia seguida para hacer el conjugado con la sonda del surco menor (mostrada en naranja).

Posteriormente se amplió esta estrategia haciendo uso de otros agentes de unión al surco menor, tales como las bisbenzamidas.^[90] Como ya se comentó, este tipo de moléculas presentan varias ventajas respecto a los polipirroles: su accesibilidad sintética y las

[111] Omichinski, J. G.; Pedone, P. V.; Felsenfeld, G.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1997**, *4*, 122.

[112] Vázquez, O.; Vázquez, M. E.; Blanco, J. B.; Castedo, L.; Mascareñas, J. L. *Angew Chem, Int. Ed. Engl.* **2007**, *46*, 6886.

propiedades de aumento de emisión al interactuar con el ADN, que permiten estudiar los conjugados y sus complejos con el ADN en disolución por técnicas de fluorescencia. Los híbridos sintetizados con estas nuevas moléculas mostraron un comportamiento muy parecido a los previamente obtenidos con los polipirroles. El híbrido basado en GCN4 mostró otra vez baja especificidad, mientras que el basado en GAGA fue mucho más selectivo.

Una etapa crítica en el diseño de estos conjugados es la longitud y naturaleza del conector. Se ha comprobado que conectores suficientemente largos pueden evitar tensiones desestabilizantes, ayudando a que ambos módulos se unan a sus respectivos sitios en la secuencia de unión híbrida. Aunque conectores demasiado largos y flexibles pueden perjudicar a la interacción debido al consiguiente aumento del coste entrópico del proceso de unión. Además, otro factor importante es la presencia de grupos amino en el conector, que al estar protonados en medio fisiológico puedan interactuar electrostáticamente con la cadena de fosfatos y así aumentar la constante de asociación. Esto se demostró mediante la síntesis de un híbrido en que el grupo amino del conector se reemplazó por un grupo éter, la constante de interacción de este híbrido resultó ser un orden de magnitud menor (Figura 39).^[113]

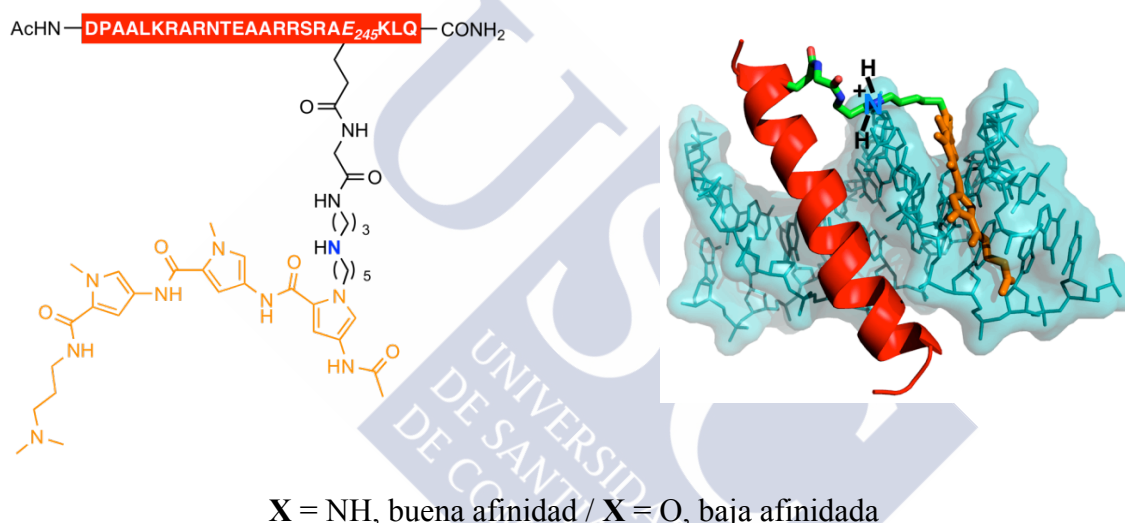


Figura 39. Izquierda: estructura del híbrido GCN4-polipirrol usado para valorar la importancia del grupo amino en el conector. Derecha: modelo de la interacción del conjugado con el ADN. El grupo amino sustituido se muestra en azul en ambas figuras.

6.4. Conjugados entre intercalantes y fragmentos de factores de transcripción

Además del diseño de híbridos entre sondas que se unen al surco menor y fragmentos de FTs, algunos grupos extendieron este concepto al uso de agentes intercalantes. La mayor limitación de esta estrategia respecto a la anterior es la falta de especificidad de este tipo de moléculas que se insertan entre las bases del ADN. Por lo tanto los híbridos sintetizados mediante esta técnica sólo presentan la especificidad que viene otorgada por la secuencia peptídica.

[113] Blanco, J. B.; Vázquez, O.; Martínez-Costas, J.; Castedo, L.; Mascareñas, J. L. *Chem. Eur. J.* **2005**, *11*, 4178.

El primer ejemplo de este tipo fue publicado por el grupo de N. W. Woodbury, que diseñó un híbrido entre el agente intercalante *Thiazole Orange* (TO)^[114] y una versión truncada del dedo de zinc del dominio de unión del receptor glucocorticoides (GR). Mientras la proteína entera presenta 66 aminoácidos y una constante de disociación de 10 nM, el híbrido con sólo 25 residuos conjugado al TO muestra una constante de 14 nM. Además debido a las propiedades del Thiazole Orange para cortar ADN cuando se irradia con luz visible, puede usarse como un análogo de nucleasa específica de secuencia y activable con luz visible.^[115]

Otro ejemplo del uso de intercalantes, lo describió el grupo de Jackeline Barton; en este caso los agentes intercalantes elegidos fueron complejos metálicos de rodio, concretamente $[\text{Rh}(\text{phi})_2(\text{bpy})]^{+3}$ y $[\text{Rh}(\text{phi})_2(\text{phen})]^{+3}$ (phi = 9,10-fenantrenoquinona diamina, phen = 1,10-fenantrolina). Estos dos intercalantes son capaces de introducir un anillo de phi entre las bases del ADN mientras el resto de complejo se sitúa en el surco mayor. La constante de disociación para este tipo de interacción es menor a 1 μM .^[116] Ambos complejos metálicos se conjugaron con la hélice de reconocimiento del represor fago 434 de la familia HTH.^[117] Al igual que en el caso del thiazole orange, estos híbridos también puede ser usado como nucleasa cuando se irradia con luz UV.

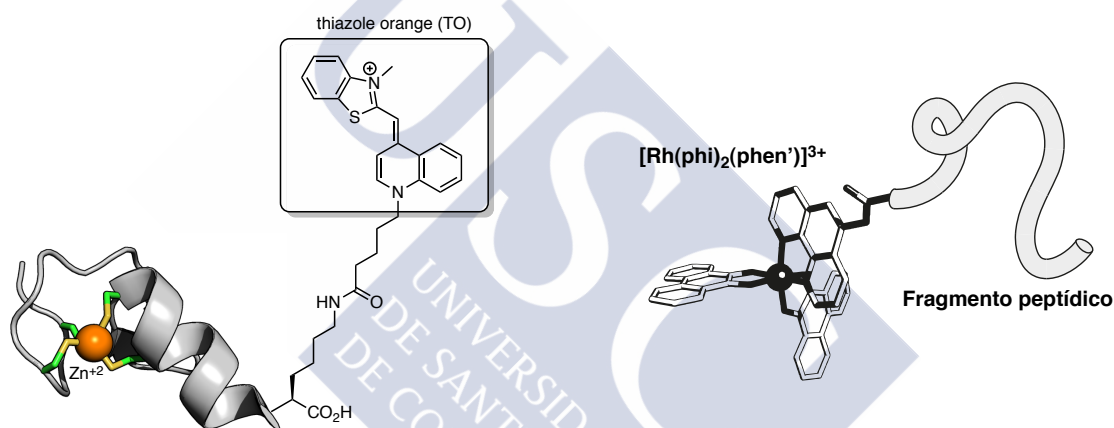


Figura 40. Izquierda: Representación esquemática del conjugado entre el thiazole orange y una versión truncada del dedo de zinc GR. Derecha: Híbrido entre un agente intercalante de Rh y hélice de reconocimiento del represor fago 434.

6.5. Estudio de la interacción entre péptidos y el ADN

Una de las técnicas más usadas para el estudio de la interacción de péptidos con el ADN es la *electroforesis*. Esta es una técnica que permite la separación de sustancias con diferente carga que se encuentran en una mezcla, tras someterlas a la acción de un campo eléctrico. La diferencia de tamaño y de carga existente entre el ADN libre y el ADN complejado con un péptido induce un retardo en la migración del complejo a través de un gel polimérico respecto al ADN libre, de tal forma que se pueden visualizar y analizar por separado en el gel ambas

[114] Nygren, J.; Svanvik, N.; Kubista, M. *Biopolymers* **1998**, *46*, 39.

[115] Thompson, M.; Woodbury, N. W. *Biochemistry* **2000**, *39*, 4327.

[116] Pyle, A. M.; Long, E. C.; Barton, J. K. *J. Am. Chem. Soc.* **1989**, *111*, 4520.

[117] Sardesai, N.; Barton, J. K. *J. Biol. Inorg. Chem.*, **1997**, *2*, 762.

moléculas.^[118] La visualización de las bandas se puede llevar a cabo mediante marcaje radiactivo o empleando reveladores fluorogénicos que aumentan su emisión al unirse al ADN (como es el caso del SYBR. Gold).^[119] El marcaje radiactivo del ADN permite trabajar con concentraciones de ADN del orden de picomolar, mientras que para el revelado fluorescente se debe usar una concentración mínima de 50 nM de ADN. La cuantificación de la relación entre el ADN libre y el complejo en cada calle permite la obtención de la constante de afinidad de un determinado híbrido por un ADN.

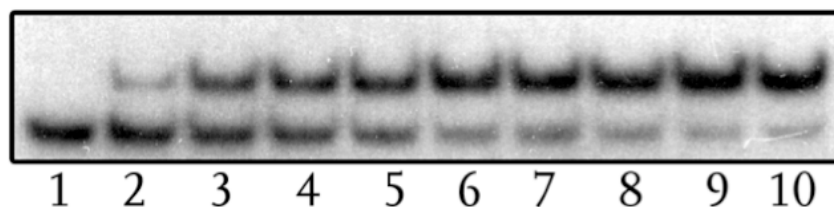


Figura 41. Electroforesis revelada radiactivamente para el híbrido previamente descrito de la proteína GAGA, con un ADN que contiene su secuencia consenso. La banda superior corresponde a la banda del complejo y la inferior a la del ADN libre. La concentración de ADN es siempre constante (45 pM) las concentraciones del híbrido en las calles 1 a la 10 son: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 nM.

Otra técnica para el estudio de la interacción con el ADN es la *anisotropía de fluorescencia*, basada en el cambio del ángulo de polarización de la radiación emitida por un fluoróforo al excitarlo con luz polarizada.^[120] Esta técnica se usa en algunos casos para investigar interacciones de proteínas con ADN utilizando oligonucleótidos de doble hélice marcados con un fluoróforo en un extremo de la cadena nucleotídica. Cuando un péptido grande se une al ADN, el tamaño y masa del complejo formado es mayor que el del ADN aislado, de manera que experimenta en disolución una rotación más lenta, lo cual se ve reflejado en un aumento del valor de anisotropía.

Finalmente existe otra técnica muy útil para el estudio de péptidos que interactúan con el ADN, denominada *dicroísmo circular* (DC).^[121] A diferencia de las anteriores, esta técnica aporta datos estructurales sobre el plegamiento peptídico y por lo tanto es especialmente útil cuando el péptido sufre un plegamiento al interactuar con el ADN como es el caso de las regiones básicas de la familia bZip.

Cada tipo de estructura secundaria presenta un espectro de DC característico y en ausencia de contribuciones por parte de cromóforos aromáticos, el espectro de DC de un péptido o una proteína puede ser considerado como la suma de las contribuciones de los espectros de DC correspondientes a los distintos motivos de estructura secundaria. Por ejemplo el espectro de DC de una hélice α está caracterizado por una banda negativa a 222 nm y otra a 208 nm, su intensidad está directamente relacionada con el 100 % de hélice

[118] Van Holde, K. E.; Jonson, W. C.; Ho, P. S. *Principles of Physical Biochemistry* **1998**, Prentice Hall, Upper Saffle River, N. J.; cap. 5.

[119] a) Zheng, D.; Zou, R.; Xinhui, L. *Anal. Chem.* **2012**, *84*, 3554. b) Kirsanov, K. I.; Lesovaya, E. A.; Yakubovskaya, M. G.; Belitsky, G. A. *Mutat. Res.* **2010**, *699*, 1.

[120] Lakowicz, J.R. *Principles of fluorescence spectroscopy* **2006**, Springer, Berlin.

[121] a) Fasman, G. D., *Circular Dichroism and the Conformational Analysis of Biomolecules* **1990**; Plenum Press: New York. b) Berova, N.; Nakanishi, K.; Woody, R. W. *Circular Dichroism* **2000**; Wiley-VCH: New York.

formada. Se ha descrito que el valor de la elipticidad molar por residuo correspondiente para el 100% de hélice α varía entre -33000 y -40000 .^[122]

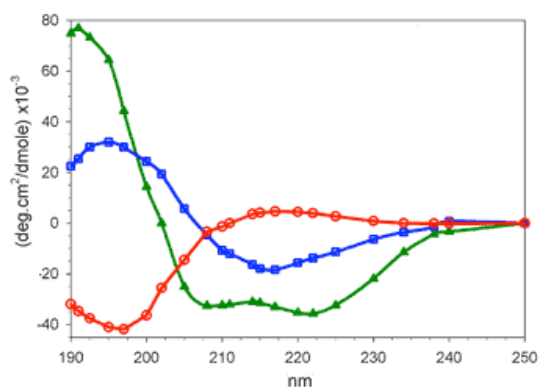


Figura 42. Espectro de DC asociado a diferentes tipos de estructuras secundarias en proteínas: hélice α (verde), lámina β (azul), estructura desordenada (rojo).



[122] Chen, Y. H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350. b) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R.L. *Nature* **1990**, *344*, 268. c) Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413.



7. PÉPTIDOS CUYA INTERACCIÓN CON EL ADN PUEDE MODULARSE

Como ya se ha comentado, los procesos biológicos y en particular los procesos de activación génica están estrictamente regulados de forma espacio-temporal por lo que sólo se producen en respuesta a determinadas señales celulares. Sin duda, disponer de moléculas sintéticas cuya interacción con el ADN pueda ser regulada a voluntad mediante estímulos externos podría abrir nuevas oportunidades en biomedicina y biología celular. Aunque es un campo poco desarrollado, existen ya algunos precedentes basados en el uso de luz y metales.

7.1. Derivados conmutables mediante el uso de luz

La luz representa un elemento de control externo ideal para lograr la regulación de interacciones proteína/ADN de forma artificial puesto que puede ser controlada de forma temporal y espacial de una manera sencilla, no invasiva y la irradiación con longitudes de onda altas no perturban el funcionamiento celular.^[123] Hasta la fecha se han descrito dos metodologías para el control de procesos moleculares mediante el uso de luz: el uso de grupos fotoisomerizables y los grupos protectores fotolábiles.

7.1.1. Grupos fotoisomerizables

Este tipo de moléculas presentan dos conformaciones o estructuras químicas cinéticamente estables, las cuales se pueden interconvertir por la acción de la luz. Se conocen varias familias de compuestos con estas propiedades: azobencenos, fúlgidos, espiropiranos y diariletanos.^[124] Aunque este tipo de moléculas presentan un gran potencial debido a que el proceso es reversible, es difícil el diseño de sistemas conmutables basados en ellas porque normalmente es complicado el desplazamiento completo del equilibrio fotoestacionario hacia uno de los dos estados.^[125]

Nuestro grupo publicó en el año 2000, el primer ejemplo de un péptido cuya interacción con el ADN puede ser modulada por la luz, basado en la utilización de una unidad de *azobenceno* como agente dimerizante de regiones básicas de GCN4. El azobenceno presenta la peculiaridad de ser isomerizable entre las formas *trans* y *cis* usando luz UV, luz visible o calentando,^[126] produciendo así cambios significativos en la geometría de los péptidos y de su interacción con el ADN, de manera que en nuestro caso el isómero *cis* interacciona 60 veces mejor con el ADN consensado que el isómero *trans*.^[127]

[123] Young, D. D.; Deiters, A. *Org. Biomol. Chem.* **2007**, *5*, 999.

[124] a) Mayer, G.; Heckel, A. *Angew. Chem. Int. Ed.* **2006**, *45*, 4900. b) Brieke, C.; Rohrbach, F.; Gottschalk, A.; Mayer, G.; Heckel, A. *Angew. Chem. Int. Ed.* **2012**, *51*, 8446.

[125] Beharry, A. A.; Woolley, G. A. *Chem. Soc. Rev.* **2011**, *40*, 4422.

[126] Bandarab, H. M. D.; Burdette, S. C. *Chem. Soc. Rev.* **2012**, *41*, 1809.

[127] Caamaño, A. M.; Vázquez, M. E.; Martínez-Costas, J.; Castedo, L.; Mascareñas, J.L. *Angew. Chem. Int. Ed.* **2000**, *39*, 3104.

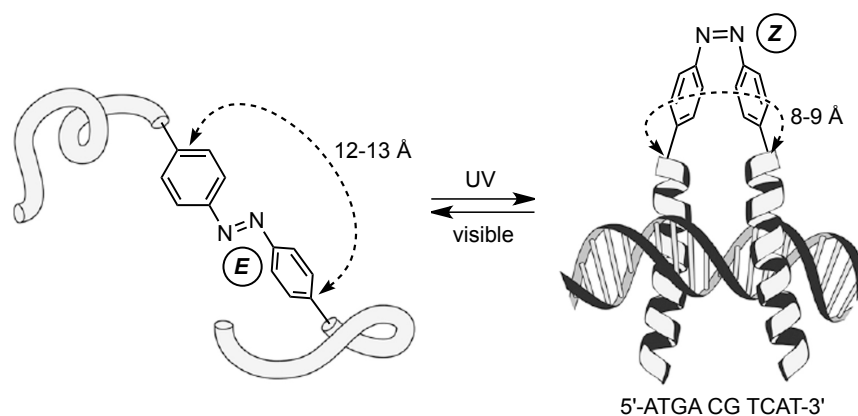


Figura 43. Isomerización del grupo azobenceno en el dímero de regiones básicas de GCN4.

En otra versión de esta estrategia, el grupo de Woolley usó un azobenceno para controlar la helicidad de un péptido. En este caso el grupo azobenceno conecta dos residuos en las posiciones i y $i+7$ de un péptido (de manera análoga a los péptidos de Taylor descritos previamente),^[104] de tal forma que el isómero *trans* estabiliza mejor la conformación de hélice α que el *cis* (Figura 44). Esta metodología se aplicó a péptidos derivados de las familias bZIP^[128] y homeodominio.^[129] Por ejemplo, en el caso de la proteína Engrailed el péptido obtenido fue capaz reconocer la secuencia QRE con una constante de disociación de 7.5 nM a 4 °C cuando el azobenceno está en *trans*, mientras que en la configuración *cis* lo hace con una constante de 140 nM.

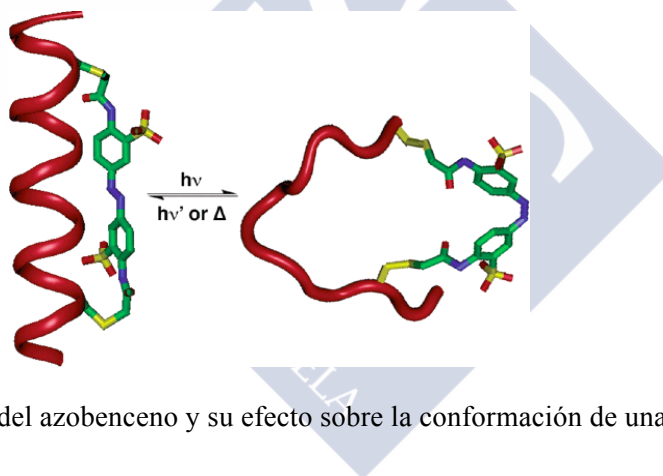


Figura 44. Fotoisomerización del azobenceno y su efecto sobre la conformación de una hélice α .^[129]

7.1.2. Grupos protectores fotolábiles

Hasta la fecha una de las estrategias más usadas para la regulación de procesos biológicos mediante el uso de luz, es el uso de grupos protectores fotolábiles (GPFs) que pueden ser liberados de forma irreversible usando luz y por lo tanto permiten pasar de una forma inactiva de una biomolécula, a una activa.

El único ejemplo del uso de este tipo de grupos para el control de la interacción de un péptido con el ADN, lo ha desarrollado nuestro grupo. En este trabajo se demuestra que es

[128] Woolley, G. A.; Jaikaran, A. S.; Berezovski, M.; Calarco, J.P.; Krylov, S. N.; Smart, O. S.; Kumita, J. R. *Biochemistry* **2006**, *45*, 6075.

[129] Guerrero, L.; Smart, O. S.; Woolley, G. A.; Allemann, R. K. *J. Am. Chem. Soc.* **2005**, *127*, 15624.

posible inhibir la capacidad de unión al ADN de un péptido dimérico de las regiones básicas de un motivo bZIP. Esto se ha logrado por la incorporación de cadenas de ácidos glutámicos en el extremo *N*-terminal, lo que genera una repulsión electrostática entre los glutámicos cargados negativamente y el ADN. Estas cadenas se unieron a las regiones básicas a través de un grupo fotolábil del tipo *orto*-nitrobencilo,^[130] que puede fragmentarse por acción de la luz, con lo que el péptido recupera la interacción con el ADN.^[131]

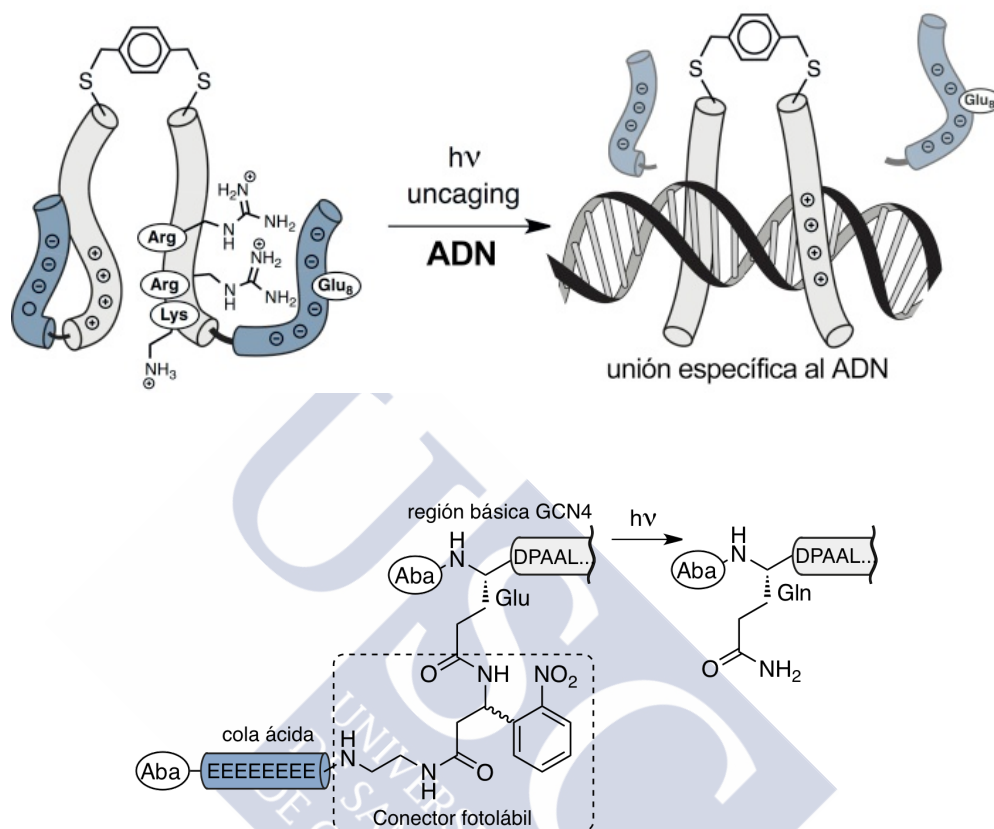


Figura 45. Arriba: Representación esquemática de la activación del dímero bZip por liberación de las colas ácidas mediante el uso de luz. Abajo: Reactividad del conector fotolábil basado en el grupo *orto*-nitrobencilo.

Los derivados del grupo *orto*-nitrobencilo, usado en el ejemplo anterior, son los GPFs más utilizados para el diseño de moléculas fotolábiles.^[132] Esto se debe básicamente a que son fáciles de hacer y tienen altos rendimientos cuánticos. Sin embargo, este tipo de grupos presentan una importante desventaja, derivada de la necesidad de usar luz ultravioleta para su fragmentación—alrededor de 350 nm—la cual es problemática en sistemas biológicos.^[133]

[130] Mayer, G.; Heckel, A. *Angew. Chem. Int. Ed.* **2006**, *45*, 4900.

[131] Jiménez-Balsa, A.; Pazos, E.; Martínez-Albarreda, B.; Mascareñas, J.L.; Vázquez, M.E. *Angew. Chem. Int. Ed.* **2012**, *51*, 8825.

[132] a) Yu, H.; Li, J.; Wu, D.; Qiu, Z.; Zhang, Y. *Chem. Soc. Rev.* **2010**, *39*, 464. b) Humphrey, D.; Rajfur, Z.; Vázquez, M. E.; Scheswohl, D.; Schaller, M. D.; Jacobson, K.; Imperiali, B. *J. Biol. Chem.* **2005**, *280*, 22091. c) Sánchez, M. I.; Martínez-Costas, J.; Gonzalez, F.; Bermudez, M. A.; Vázquez, M. E.; Mascareñas, J. L. *ACS Chem. Biol.* **2012**, *7*, 1276. d) Ciesienki, K. L.; Hyman, L. M.; Yang, D. T.; Haas, K. L.; Dickens, M. G.; Holbrook, R. J.; Franz, K. J. *Eur. J. Inorg. Chem.* **2010**, *15*, 2224.

[133] Pelliccioli, A. P.; Wirz, J. *Photochem. Photobiol. Sci.* **2002**, *1*, 441.

En la actualidad existen algunas variantes de GPFs que pueden ser liberados con luz visible, la mayor parte de ellos basados en complejos metálicos. Un ejemplo reciente es la cobalamina, en la cual un átomo de Co(II) está coordinado a un anillo de corrina y a un pseudonucleótido, el 5,6-dimetilbencimidazol (Figura 46). Este complejo, desarrollado como grupo fotolábil por el grupo de David S. Lawrence,^[134] presenta una posición de coordinación libre a la que se pueden unir diferentes funcionalidades que son liberadas cuando el complejo se irradia con longitudes de onda entre 500 y 570 nm.^[135]

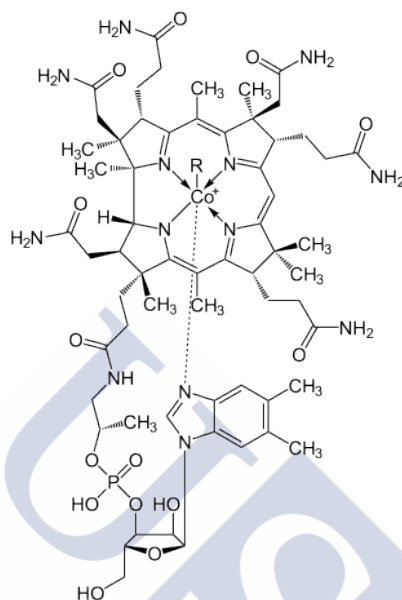


Figura 46. Estructura de la cobalamina. *R* representa el grupo que puede liberarse con la irradiación.

Otro ejemplo de este tipo de GPFs lo publicó en el año 2003 el grupo de R. Etchenique, y se basa en el uso de complejos de rutenio-bipiridina. Este complejo de Ru(II) enlazado a dos bipiridinas y a dos moléculas 4-aminopiridina (coordinadas al rutenio por el nitrógeno de la piridina) es capaz de liberar una de las moléculas 4-aminopiridina al ser irradiado con luz verde, aproximadamente 480 nm (Figura 47).^[136]

Posteriormente el mismo grupo demostró que si este rutenio con dos bipiridinas se coordina a una fosfina y en la otra posición a una amina o una piridina, al irradiar con luz se libera selectivamente la amina o piridina mientras que la fosfina permanece coordinada.^[137]

[134] a) Priestman, M. A.; Shell, T. A.; Sun, L.; Lee, H.; Lawrence, D. S. *Angew. Chem. Int. Ed.* **2012**, *51*, 7684. b) Shell, R. A.; Shell, J. R.; Rodgers, Z. L.; Lee, J.; Lawrence, D. S. *Angew. Chem. Int. Ed.* **2014**, *53*, 875.

[135] Taylor, R. T.; Smucker, L.; Hanna, M. L.; Gill, J.; *Arch. Biochem. Biophys.* **1973**, *156*, 521.

[136] Zayat, L.; Calero, C.; Alborés, P.; Baraldo, L.; Etchenique, R. *J. Am. Chem. Soc.*, **2003**, *125*, 882.

[137] Salierno, M.; Marceca, E.; Peterka, D. S.; Yuste, R.; Etchenique, R. *J. of Inorg. Biochem.* **2010**, *104*, 418.

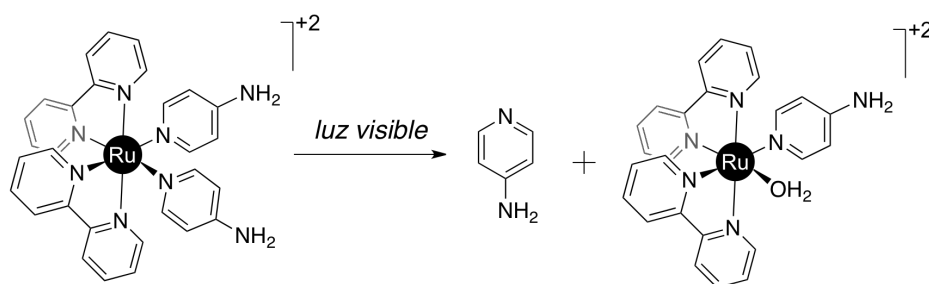


Figura 47. Fotodesprotección de la molécula 4-aminopiridina inducida por el uso de luz visible.

7.2. Derivados conmutables por la acción de metales

Al igual que los metales son usados por la naturaleza para estabilizar estructuras secundarias, también en la bibliografía existen ejemplos de péptidos sintéticos que cambian su conformación por la acción de metales.^[138] Sin embargo hasta la fecha existen muy pocos ejemplos del uso de metales para la modulación de la interacción de péptidos con el ADN. El grupo de S. Futaki describió derivados del motivo bZip cuya cremallera de leucinas se derivatizó con residuos de cisteína funcionalizados con ácido iminodiacético (Ida). Este grupo demostró que cuando dichas cadenas se coordinan a iones Co(II) se dificulta la unión específica al ADN.^[139]

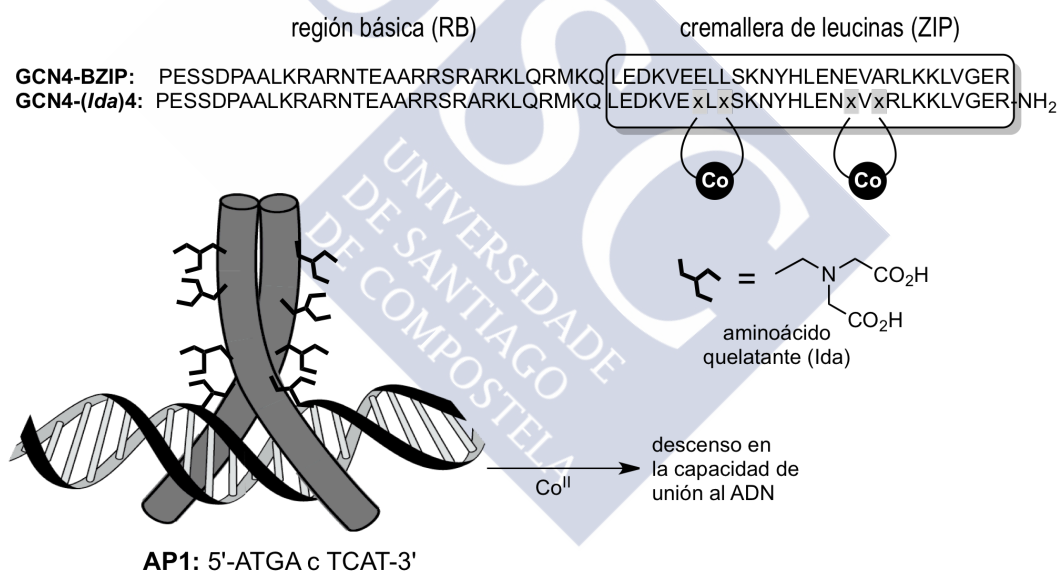


Figura 48. Proteína bZIP con elementos quelatantes de metales Ida en posiciones *i*, *i*+2 de la cremallera de leucinas que permiten la desactivación de la unión al ADN.

El grupo T. Yamamura diseñó un híbrido entre un FT del tipo dedo zinc (zif268) y la región de unión a iones calcio de la Troponina C cuya actividad pueda ser modulada por la presencia de iones calcio. El conjugado resultante presenta una *K_d* de 5.8 nM en ausencia de calcio y 13 nM en presencia de 100 equivalentes de calcio. La diferencia en la constante de interacción se

[138] a) Ghadiril, M. R.; Choib, C. *J. Am. Chem. Soc.* **1990**, *112*, 1630. b) Ma, M. T.; Hoang, H. N.; Scully, C. G.; Appleton, T. G.; Fairlie, D. P. *J. Am. Chem. Soc.* **2009**, *131*, 4505.

[139] Azuma, Y.; Imanishi, M.; Yoshimura, T.; Kawabata, T.; Futaki, S. *Angew. Chem. Int. Ed.* **2009**, *48*, 6853.

debe a que el motivo de unión a calcio presenta una estructura flexible en ausencia del metal dicatiónico, sin embargo en su presencia tiene una estructura rígida que dificulta la unión al ADN del motivo dedo de zinc.^[140]

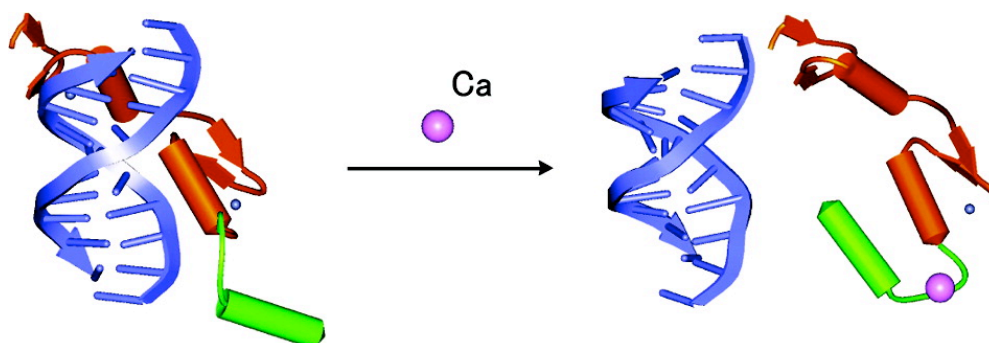


Figura 49. Híbrido entre un FT del tipo Dedo Zinc (zif268) y la región de unión a Ca^{+2} de la Troponina C diseñado por Takeshi Yamamura.



[140] Onoda, A.; Arai, N.; Shimazu, N.; Yamamoto, H.; Yamamura, T. *J. Am. Chem. Soc.* **2005**, *127*, 16535.

OBJETIVOS





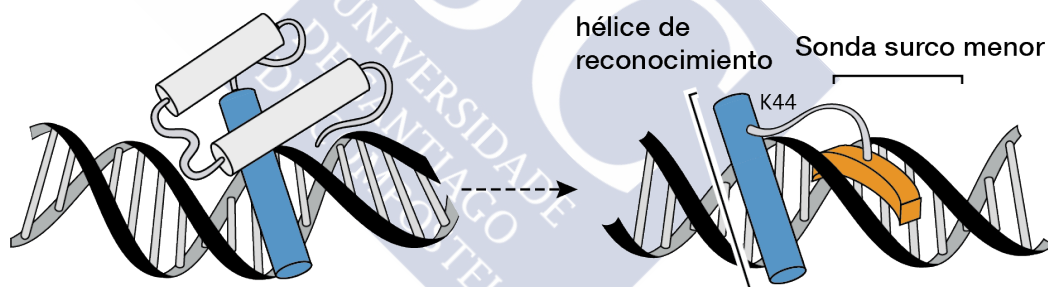
Como se ha comentado en la introducción, la interacción de FTs con el ADN es un aspecto clave en la expresión génica. En muchos casos dicha interacción está perfectamente regulada en tiempo y en espacio.

El papel crucial del ADN como centro de información genética ha llevado a desarrollar diferentes tipos de moléculas sintéticas capaces de unirse a determinadas secuencias de ADN con buena afinidad. En todo caso, el progreso en el desarrollo de sistemas miméticos de factores de transcripción ha sido bastante escaso, siendo incluso más lento el desarrollo de sistemas sintéticos cuya capacidad para unirse al ADN pueda regularse utilizando señales externas, de forma paralela a como lo hacen los FTs.

En este contexto se fijaron los siguientes objetivos para esta tesis doctoral:

- **Desarrollo de versiones simplificadas de FTs del tipo homeodominio mediante el uso de bisbenzamidinas.** En nuestro grupo se ha logrado la obtención de versiones simplificadas de FT de las familias bZIP y dedos de zinc que presentan capacidad de reconocimiento del ADN con alta afinidad y especificidad. Este objetivo se llevó a cabo mediante la conjugación del núcleo peptídico de estos FT a sondas moleculares que se unen al surco menor. Sin embargo, hasta la fecha no se ha podido extender esta estrategia a los FT de la familia homeodominio, a pesar de la gran importancia de estas proteínas en la diferenciación celular de los mamíferos.

Por lo tanto se plantea como objetivo el diseño de híbridos entre sondas de surco menor del tipo bisbenzamidina y la hélice de reconocimiento de FT homeodominio con el fin de obtener conjugados que tengan afinidades y selectividades comparables a las de las proteínas naturales.



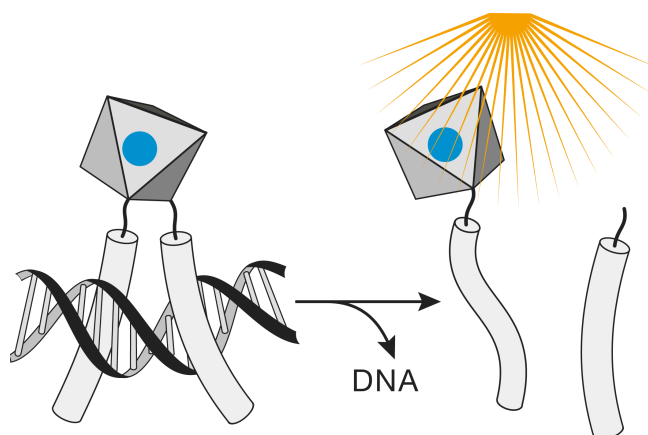
Publicación asociada con este objetivo:

Selective DNA-binding by designed Bisbenzamidine-Homeodomain Chimeras, Mosquera, J.; Rodríguez, J.; Vázquez, M. E.; Mascareñas, J. L. *ChemBioChem* **2014**, *15*, 1092 (DOI: 10.1002/cbic.201400079).

- **Diseño de péptidos sintéticos cuya unión al ADN pueda ser desactivada por la acción de luz visible.** Hasta la fecha existen dos formas generales de modular la actividad de péptidos o proteínas mediante la luz. La primera de ellas se basa en el uso de grupos fotoisomerizables del tipo azobenceno, la otra es el uso de grupos fotoprotectores, como es el caso del nitrobeneno. Ambas estrategias presentan limitaciones; para el caso de la fotoisomerización la mayor desventaja es la pequeña diferencia de actividad que se suele encontrar entre los dos isómeros y la dificultad de desplazar de forma adecuada los equilibrios; para el caso de los grupos fotoprotectores es la necesidad de utilizar luz ultravioleta, que es incompatible con medios biológicos y la dificultad de aplicar esta metodología para producir procesos

de desactivación en lugar de activación.

Por ello nos planteamos el diseño de péptidos diméricos basados en FTs bZIP cuya unión al ADN pueda desactivarse mediante irradiación con luz visible. En concreto nos planteamos usar como unidad dimerizante un complejo de rutenio que experimenta la disociación de uno de sus ligandos cuando se irradia con luz visible.

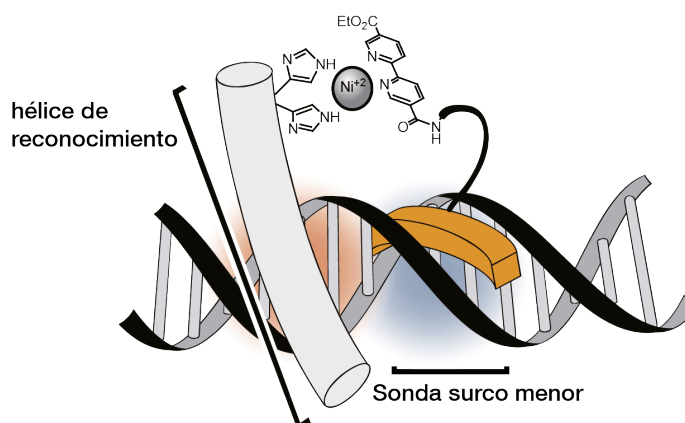


Publicación asociada con este objetivo:

Ruthenium Bipyridyl Complexes as Photocleavable Dimerizers: Deactivation of DNA-Binding Peptides Using Visible Light, Mosquera, J.; Sánchez, M. I.; Vázquez, M. E.; Mascareñas, J. L. *Chem. Commun.* **2014**, *50*, 10975 (DOI: 10.1039/C4CC04512B)

- **Desarrollo de sistemas reversibles para el reconocimiento específico de secuencias de ADN que responden a la presencia de determinados metales.** Como se ha comentado, algunas proteínas naturales, como los dedos de zinc, necesitan metales para ser activas. Inspirándonos en este hecho, nos planteamos el uso de metales como estímulos químicos externos que permitan regular la unión al ADN.

En concreto nos planteamos el diseño de un sistema multicomponente que reconozca secuencias de ADN de forma específica. Este sistema consta de dos elementos (una bisbenzamidina que une al surco menor y un péptido derivado de la región básica de una proteína bZIP) derivatizados con grupos coordinantes, lo que les va a permitir cooperar entre ellos para interactuar con el ADN, sólo en presencia del metal adecuado.

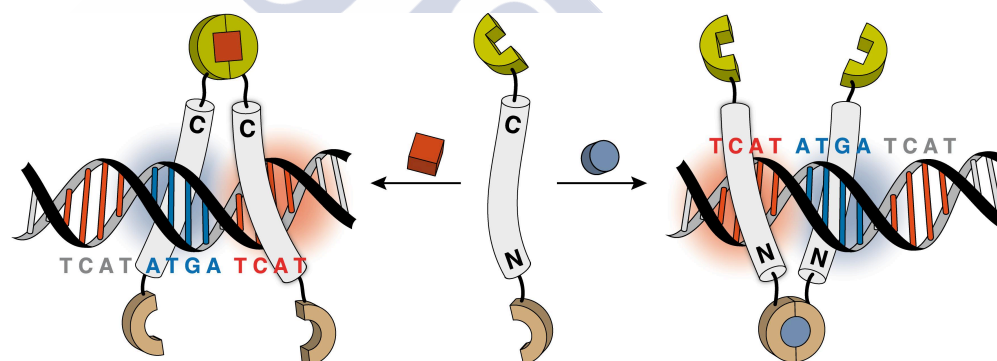


Publicación asociada con este objetivo:

Reversible Supramolecular Assembly at Specific DNA Sites: Nickel-Promoted Bivalent DNA Binding with Designed Peptide and Bipyridyl–Bis(benzamidine) Components, Sánchez, M. I.; Mosquera, J.; Vázquez, M. E.; Mascareñas, J. L. *Angew. Chem. Int. Ed.* **2014**, *53*, 9917. (DOI: 10.1002/anie.201405726)

- **Diseño de sistemas peptídicos cuya selectividad de unión al ADN puede regularse mediante estímulos externos.** Hasta la fecha la mayor parte de las moléculas sintéticas que unen ADN son monoméricas. Sin embargo, en la naturaleza el reconocimiento del ADN suele ocurrir mediante la interacción cooperativa entre diferentes módulos proteicos. Esto permite reconocer un gran número de secuencias diferentes usando un pequeño número de proteínas (control combinatorio) y facilita su activación o desactivación (véase página 38 para el caso de proteínas bZIP).

Inspirados por este tipo de control combinatorio, nos planteamos el diseño de un FTs artificiales cuya especificidad de secuencia pueda ser regulada a voluntad mediante la adición de determinados agentes externos. En concreto, nos proponemos controlar la dimerización de regiones básicas de proteínas bZIP a través de los extremos C- o N-terminales, con el fin de dirigirlos a diferentes secuencias de ADN. Específicamente, ATGA•TCAT para el caso del péptido C-terminal o TCAT•ATGA para el N-terminal.



Publicación asociada con este objetivo:

Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides, Mosquera J.; Jiménez-Balsa, A.; Doderio, V. I.; Vázquez, M. E.; Mascareñas, J. L. *Nature Communications* **2013**, *4*, 1874 (DOI: 10.1038/ncomms2825).



CAPÍTULO 1

***Desarrollo de versiones
simplificadas de FTs del tipo
homeodominio mediante el
uso de bisbenzamidinas***





***Selective DNA-Binding by Designed Bisbenzamidine-Homeodomain
Chimeras***





CAPÍTULO 2

***Diseño de péptidos sintéticos
cuya unión al ADN pueda ser
desactivada por la acción de
luz visible***





***Ruthenium bipyridyl complexes as photocleavable dimerizers:
deactivation of DNA-binding peptides using visible light***





CAPÍTULO 3

Desarrollo de sistemas reversibles para el reconocimiento específico de secuencias de ADN que responden a la presencia de determinados metales



Reversible Supramolecular Assembly at Specific DNA Sites: Nickel-Promoted Bivalent DNA Binding with Designed Peptide and Bipyridyl–Bis(benzamidine) Components





CAPÍTULO 4

***Diseño de sistemas peptídicos
cuya selectividad de unión al
ADN puede regularse
mediante estímulos externos***

UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA



Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides





CONCLUSIONS





In this thesis were studied different strategies to achieve emulated the DNA binding properties of natural transcription factors. The main conclusions for each chapter are shown below.

Chapter 1. Development of simplified homeodomain transcription factors conjugated to bisbenzamidines.

- Three conjugates between variants of the helix 3 region of a Q50K Engrailed homeodomain and bisbenzamidine minor-groove DNA binders were prepared.
- It was proved the hybrid featuring the sequence of the native protein failed to bind to DNA; however, modifications that increased the α -helical folding propensity of the peptide allowed specific DNA binding by a bipartite (major/minor groove) interaction.

Chapter 2. Design of light-switchable DNA-binding peptides.

- A photolabile biselectrophilic Ru(II) complex, which can be used for homo- or heterodimerization of cysteine-containing peptides, was developed. This complex can be efficiently disassembled by visible light.
- The Ru(II) complex was used to prepare a homodimer of GCN4 basic regions and a heterodimer between the basic regions of both GCN4 and C/EBP.
- It was proved that the interaction with the DNA for both dimers can be suppressed at will by irradiation with visible light due to the generation of monomeric species that can not interact with DNA.

Chapter 3. Development of a reversible supramolecular assembly at specific DNA sites: Ni-promoted, bivalent DNA binding with designed peptide and bipyridyl-bisbenzamidine components.

- A quaternary supramolecular system was designed, which is made up of a Ni(II) atom, a bis(histidine)-modified peptide that is derived from a bZIP transcription factor, a bis(benzamidine) unit that is equipped with a bipyridine and a double stranded DNA with the target sequence.
- It was proved that the system dynamically assembles independently of the order in which the components are mixed, and all components are required to obtain the supramolecular structure.
- It was demonstrated the resulting multicomponent supramolecular complex can be dismantled by the addition of an external nickel chelator.

Chapter 4. Switching between sites: Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides.

- A peptide based on GCN4 basic region was synthesized. It can be dimerized through the C-terminal region using metals and through the N-terminal using oxidating conditions.
- It was proved that the monomeric peptide is non-functional. However, a C-terminal metallo-dimer recognizes the natural ATF/CREB-binding site (5'-ATGA

cg TCAT-3'), and a *N*-terminal disulphide dimer binds preferentially to the swapped sequence (5'-TCAT cg ATGA-3').

- The peptide can be driven to either DNA target site by using specific external signals that code its dimerization mode. It is even possible to induce a reversible 'hopping between sites' in a stimuli-responsive manner by using oxidizing or reducing agents.



RESUMEN





Capítulo 1: Desarrollo de versiones simplificadas de FTs del tipo homeodominio mediante el uso de bisbenzamidinas.

Este capítulo se ha centrado en el diseño de análogos de factores de transcripción de la familia homeodominio (HD). Como ya se ha comentado en la introducción, esta familia tiene una gran importancia en el desarrollo del organismo y especialmente en la diferenciación celular del sistema nervioso. Por lo tanto, el disponer de moléculas sintéticas que puedan interferir en su unión al ADN podría dar lugar a nuevas herramientas para el control celular.

Como proteína de referencia de esta familia elegimos el FT *Engrailed*, concretamente el Engrailed Q50K que tiene una alta afinidad por la secuencia QRE (*quaking response element*: TAATCC). Este FT, al igual que el resto de la familia, consta de tres hélice α de las cuales solamente una (la hélice de reconocimiento) interacciona con el ADN a través del surco mayor, mientras que una cola en el extremo *N*-terminal se inserta en el surco menor del ADN. Para el diseño de los análogos nos planteamos hacer una simplificación estructural consistente en eliminar las hélices de soporte, no implicadas directamente en el reconocimiento del ADN y sustituir la cola *N*-terminal por una sonda del tipo bisbenzamidina en una estrategia análoga a la previamente usada para la obtención de híbridos del FTs del tipo bZIP y dedos de zinc.

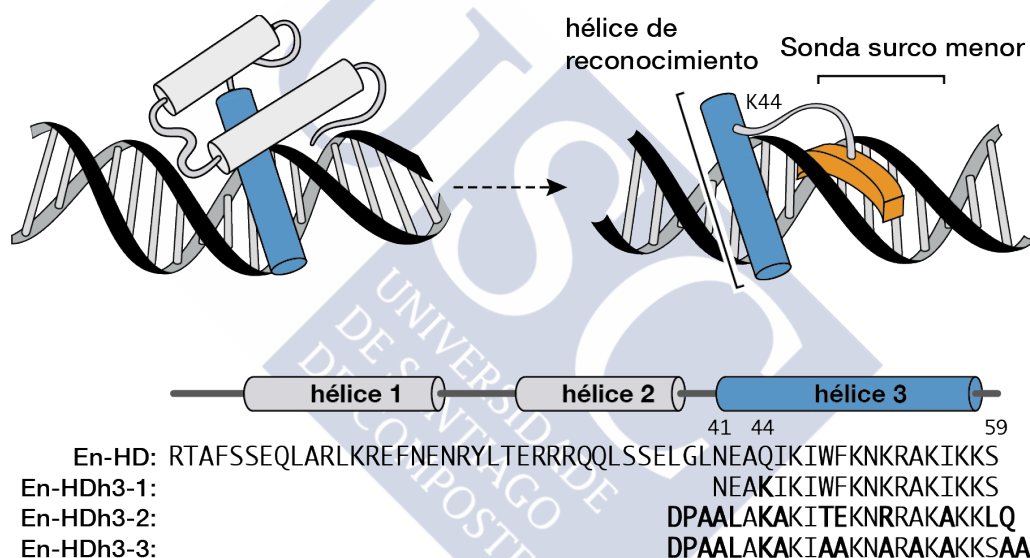


Figura 1. Arriba: Representación de la estrategia para el diseño de versiones simplificadas del factor de transcripción *Engrailed*. Abajo: secuencia del Engrailed Q50K (**En-HD**) y secuencias de la región peptídica de los tres híbridos que se conjugaron a una bisbenzamidina a través del residuo de Lys44. Las letras en negrita indican los aminoácidos mutados respecto a la secuencia de la proteína modelo.

La primera aproximación hacia el diseño de híbridos basados en el FT Engrailed fue la unión de una bisbenzamidina a un péptido con la secuencia de la hélice de reconocimiento del Engrailed Q50K (Híbrido **En-HDh3-1**). El estudio de la interacción de este híbrido con el ADN fue llevada a cabo mediante la técnica de EMSA (*electrophoretic mobility shift assay*) y DC (*circular dichroism*). Los experimentos de EMSA no mostraron ninguna interacción entre este híbrido y un ADN de doble cadena con la secuencia diana compuesta por los sitios de unión del péptido y de la bisbenzamidina, además en los experimentos de DC tampoco se observó diferencia entre la helicidad del péptido en presencia o ausencia de ADN.

La experiencia previa de nuestro grupo en el diseño de híbridos entre fragmentos de FTs y sondas del surco menor nos ha demostrado que una parte fundamental en el diseño de estos conjugados es la longitud del conector que une ambas entidades. Especialmente importante es

que el conector sea suficientemente largo para evitar tensiones desestabilizantes cuando se produce la interacción simultánea a través de los surcos mayor y menor. Para aclarar si este problema estaba afectando al diseño original, se sintetizó otro híbrido con un conector de mayor tamaño. Desafortunadamente, este conjugado continuó sin mostrar capacidad para interactuar con el ADN. De este trabajo se concluyó que la hélice de reconocimiento del Engrailed no tiene suficiente tendencia para adoptar el plegamiento helicoidal en ausencia del resto de la cadena peptídica, por lo que no es capaz de interactuar con el ADN. Para intentar solucionar este problema se recurrió a una estrategia de *grafting* (introducción página 50) con la región básica del GCN4. De esta forma se sintetizó el híbrido **En-HDh3-2** en que los aminoácidos no esenciales de la hélice de reconocimiento del Engrailed se sustituyeron por los de la región básica del GCN4. Además, se llevo a cabo la síntesis de otro híbrido (**En-HDh3-3**) en el que esos mismos residuos no esenciales se mutaron por alaninas, debido a la gran tendencia de este aminoácido para plegar en forma de hélice α , y se le añadió el *N*-Cap del GCN4 (introducción página 23).

Al igual que en el primer caso, estudiamos la interacción de ambos híbridos con el ADN mediante EMSA y DC. En ambos casos se observó interacción con el ADN de doble cadena con la secuencia consenso del péptido y la sonda (TTTAATCC); concretamente, se comprobó que el híbrido **En-HDh3-2** se une al ADN con una constante de disociación de 188 nM a 4 °C y el **En-HDh3-3** con una afinidad ligeramente superior de 131 nM. También se demostró que ambos híbridos poseen buena especificidad de secuencia, ya que no interactúan con ADNs con mutaciones en la región de reconocimiento del péptido. Finalmente, los experimentos de DC mostraron un aumento de helicidad del péptido en presencia del ADN con la secuencia consenso, consistente con el plegamiento de la hélice de reconocimiento.

Estos resultados demuestran que combinando la técnica del *grafting* con la unión a una sonda del surco menor, se puede inducir que las hélices de reconocimiento de las proteínas de la familia homeodominio unan a secuencias específicas de ADN con buena afinidad y selectividad. Estos conjugados simulan el reconocimiento del ADN del homeodominio original, uniendo a secuencias análogas e interactuando simultáneamente por el surco mayor y menor.

Capítulo 2: Diseño de péptidos sintéticos cuya unión al ADN pueda ser desactivada por la acción de luz visible.

Las células son capaces de responder a estímulos externos y modular una gran número de procesos para adaptarse a sus necesidades. Una de las etapas claves en las que la célula llevan a cabo esta regulación es en la *transcripción*, a través principalmente del control de la actividad de los factores de transcripción. Como se demostró en la introducción hay un gran interés en el diseño de análogos de factores de transcripción que, al igual que los FT naturales, puedan ser activados o inactivados en momentos puntuales.

Uno de los estímulos externos ideales para el control de la actividad de biomoléculas es el uso de la luz visible, debido a su compatibilidad con el entorno celular. Hasta ahora la luz se ha usado como estímulo para la activación de diferentes moléculas, sin embargo no existen ejemplos del su uso para la desactivación. En este contexto diseñamos un conector sensible a la luz visible que puede ser usado para la dimerización de péptidos o proteínas a través de residuos de cisteína. Este conector, basado en un complejo de Ru(II) con dos bipyridinas y dos ligandos amino, experimenta la liberación de una de las aminas al ser irradiado con luz visible.

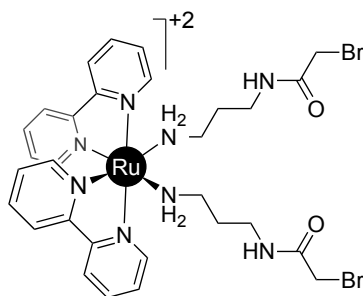


Figura 2. Conector basado en un complejo de rutenio capaz de liberar uno de sus ligandos amino al ser irradiado con luz visible.

El conector de rutenio se usó para dimerizar dos regiones básicas del FT GCN4 de la familia bZIP, derivatizadas con dos residuos de cisteína (**Ru(brC)₂**). La reacción entre el conector y los péptidos ocurre a través de un ataque nucleófilo del átomo de azufre, esta reacción se lleva a cabo en tampón fosfato a pH 7.8 con el péptido completamente desprotegido y a temperatura ambiente.

Por otra parte, estudiamos la interacción con el ADN del homodímero **Ru(brC)₂** mediante ensayos de EMSA, DC y anisotropía fluorescente. Los experimentos de anisotropía fluorescente permitieron obtener la constante de disociación de este dímero con su secuencia consenso (ATGA cg TCAT), que resultó ser 12 nM a 20 °C. También se estudió la fotólisis del dímero **Ru(brC)₂** con dos fuentes diferentes, observándose que la fotólisis se completa en 20 minutos al irradiar con longitudes de onda entre 530–550 nm, pero puede llevarse a cabo sólo en 2 minutos cuando se irradia la muestra con una fuente LED a 455 nm. Además este proceso se puede llevar a cabo en presencia del ADN y los monómeros resultantes no tienen capacidad de interacción con el ADN. Por otro lado, para garantizar que no ocurre ninguna oxidación del ADN al irradiar por parte del complejo del rutenio, análisis por MALDI demostraron que la masa del oligonucleótido no cambia durante la fotólisis.

Finalmente, se realizó la síntesis de un heterodímero entre la región básica del GCN4 y la región básica del FT C/EBP, también de la familia bZIP. La síntesis en este caso se hizo de forma secuencial ajustando los equivalentes de péptido. Este heterodímero fue capaz reconocer la secuencia (ATTG cg TCAT) y al igual que en el caso del **Ru(brC)₂** puede ser fotolizado en presencia de su ADN consenso.

Capítulo 3: Desarrollo de sistemas reversibles para el reconocimiento específico de secuencias de ADN que responden a la presencia de determinados metales.

En muchos casos el proceso de reconocimiento del ADN por los FTs naturales está acoplado al plegamiento de su dominio de unión al ADN, normalmente una hélice α que se inserta en el surco mayor. En otros casos, como en el de la familia de los dedos de Zn, el plegamiento es dependiente de la coordinación de un ión metálico a las cadenas laterales de residuos de histidina o cisteína.

En este capítulo se presenta una nueva estrategia para el reconocimiento específico de secuencias de ADN que combina muchos de los atributos de los factores de transcripción naturales: bivalencia, plegamiento condicional, respuesta a estímulos externos y reversibilidad. Dicha estrategia consiste en el doble papel del níquel (+2) como agente promotor del plegamiento de un péptido derivado del FT GCN4 y como agente dimerizador entre este péptido y una sonda del surco menor.

El péptido usado para la construcción de este sistema deriva de la región básica del FT GCN4 en el que los aminoácidos Leu230 y Arg234, cercanos al extremo *N*-terminal, fueron sustituidos por histidinas (**brHis₂**). Por lo tanto, la adición de Ni(II) induce la formación de la hélice α por las restricciones conformacionales derivadas de la interacción entre las histidinas y el ión metálico. Al mismo tiempo, el Ni(II) proporciona un sitio de coordinación para una bisbenzamidina equipada con un ligando bipyridina. Dicha bisbenzamidina unida selectivamente al surco menor del ADN permite reclutar el péptido **brHis₂** al surco mayor adyacente que contiene la secuencia consenso de este péptido.

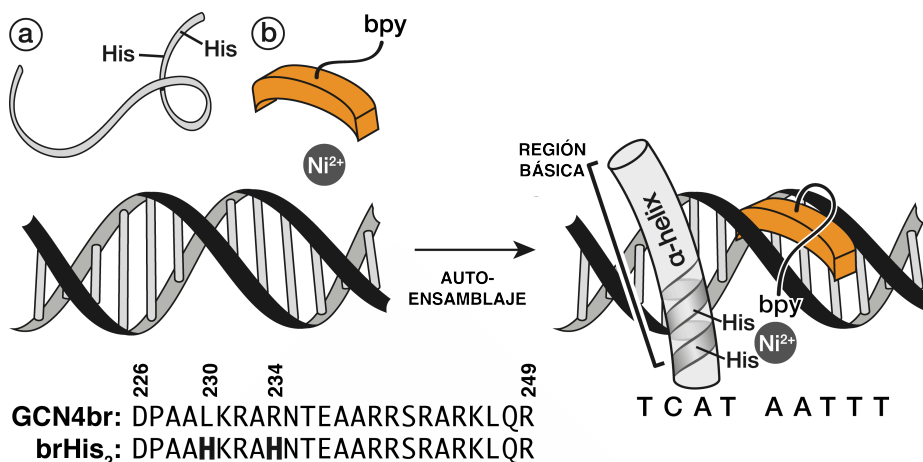


Figura 3. Representación esquemática de la estrategia de autoensamblaje. a) Péptido **brHis₂** que pasa de estar desplegado a plegarse en forma de hélice α cuando forma parte del complejo (derecha). b) Bisbenzamidina que se inserta en el surco menor del ADN. Bpy: bipyridina unida a la bisbenzamidina a través de un conector.

Las propiedades de unión al ADN de este sistema fueron estudiadas por EMSA. Como era de esperar la incubación del péptido **brHis₂** y la bisbenzamidina con un oligonucleótido de doble cadena que contiene la secuencia del péptido y la sonda del surco menor, no indujo formación de bandas de retardo en el gel. Sin embargo, cuando a este mezcla se le añade una sal de Ni(II) se observa la aparición de una banda en el gel consistente con la formación del complejo mostrado en la figura 3. Experimentos adicionales desvelaron que todos los componentes son necesarios para que el péptido pueda interactuar con el ADN, además debido a la reversibilidad del autoensamblaje el orden de adición de los componentes no es relevante. Como cabría esperar los datos de DC presentaron un bajo contenido de helicidad para el péptido **brHis₂** incluso en presencia de ADN, solamente con la adición de todos los componentes el péptido es capaz de plegar en forma de hélice.

Una valoración mediante EMSA usando cantidades equimoleculares del péptido y sonda, y con cinco equivalentes de níquel permitieron calcular una constante de disociación aproximada de 450 nM a 4 °C para el complejo.

Seguidamente realizamos experimentos de selectividad de secuencia, usando ADNs con una sola mutación en la región de reconocimiento del péptido o de la sonda. En ninguno de los dos casos se observó interacción con estas secuencias de ADN, poniendo de manifiesto una gran selectividad derivada del reconocimiento cooperativo de este sistema multicomponente. También investigamos la selectividad de iones metálicos, para ello se hicieron EMSAs con sales divalentes de Co, Ni, Cd, Zn, Cu, Hg y Fe. Usando condiciones idénticas sólo fueron observadas bandas de retardo con las sales de níquel, por lo tanto se concluyó que sólo el níquel posee las propiedades de coordinación necesarias para permitir el ensamblaje

heterodimérico en presencia del ADN. Debido al papel clave de los iones de Ni(II) como componente esencial para permitir el autoensamblaje del sistema, estudiamos la posibilidad del uso de moléculas quelatantes de níquel como el AEDT para la obtención de un sistema reversible; así pues, cuando este agente se añade a una mezcla de todos los componentes se produce una gran disminución en la helicidad del sistema y una desaparición de la banda del complejo en los geles de EMSA, demostrándose así que la presencia de níquel es crítica para la interacción con el ADN.

Capítulo 4: Diseño de sistemas peptídicos cuya selectividad de unión al ADN puede regularse mediante estímulos externos.

El patrón de la expresión génica en un organismo multicelular viene determinado por la acción de señales extracelulares que aseguran la producción de determinadas proteínas sólo cuando y dónde sean necesarias. En la mayor parte de los casos estas señales afectan a la actividad de los factores de transcripción. Por otra parte, se sabe que muchos FTs no son capaces de unirse al ADN por sí mismos y necesitan formar complejos cooperativos para lograrlo, de esta forma la interacción combinatoria permite el reconocimiento selectivo de un gran número de secuencias de ADN usando un pequeño número de FTs monoméricos.

A pesar de los avances en el diseño de sistemas cuya interacción con el ADN pueda ser regulada, hasta la fecha no se habían descrito sistemas capaces de unirse a diferentes secuencias de ADN en respuesta a estímulos externos. En este capítulo se demuestra por primera vez el concepto de “*site switchability*” en la unión a ADN de sistemas sintéticos. En los últimos años ha habido varios ejemplos de diferentes tipos de conectores sintéticos, que pueden dimerizar péptidos de reconocimiento sustituyendo así a la unidad cremallera de leucinas característica de la familia de FT bZIP. Además, el grupo T. Morii diseñó un conector que permite dimerizar al FT GCN4 por el extremo *N*-terminal y por lo tanto el dímero resultante en vez de reconocer la secuencia ATF/CREB (ATGA cg TCAT) reconoce la secuencia inversa TCAT cg ATGA. En base a los anteriores antecedentes, diseñamos una región básica derivada del FT GCN4 equipada con dos unidades dimerizantes diferentes en cada extremo: una terpiridina en el extremo *C*-terminal, que permite la dimerización de forma reversible por la adición iones metálicos como Ni(II) o Fe(II), y un residuo de Cys en el extremo *N*-terminal, que puede ser fácilmente oxidado a disulfuro o reducido a tiol por la adición de agentes oxidantes o reductores de forma ortogonal a la dimerización *C*-terminal mediada por metales.

Como primera aproximación se sintetizó un péptido denominado **3a** que incorpora las dos funcionalidades, y su interacción con las diferentes secuencias de ADN fue estudiada mediante EMSA. Se comprobó que el dímero metálico de níquel derivado de este péptido (**3a**)₂Ni, se une a la secuencia de ADN directa (ATGA cg TCAT), pero no a la inversa (TCAT cg ATGA). Sin embargo cuando los mismos ensayos se repitieron con el derivado dimerizado por el extremo *N*-terminal a través de un disulfuro (**3a**)₂SS se obtuvo que este dímero interaccionaba con las dos secuencias de ADN. Aunque la interacción con su secuencia consenso, la inversa, parece mas fuerte debido a la formación de bandas más intensas en el gel, en presencia de la secuencia directa se observaban también bandas difusas. Estudios con péptidos modelo que no tienen terpiridinas libres en el extremo *C*-terminal y con secuencia de ADN mitad (que incluye solamente la secuencia para una región básica), indican que la interacción con el ADN de secuencia directa probablemente se debe a la intercalación de la terpiridina entre las bases del ADN. De esta forma, una de las regiones básicas interacciona con su secuencia consenso a través del surco menor, al tiempo que la otra hace contactos electrostáticos con el ADN y favorece la intercalación de la terpiridina.

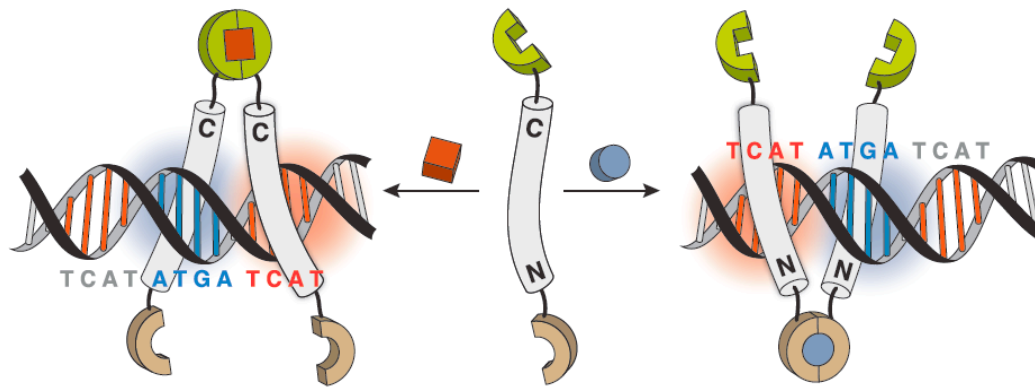


Figura 4. Representación de la estrategia de dimerización dependiente de estímulos para la unión a dos regiones del ADN. El péptido derivado de la región básica de GCN4 contiene dos funcionalidades en las dos regiones terminales que le permiten dimerizar por uno u otro extremo en función del estímulo recibido y de esta forma unir a dos secuencias de ADN diferentes: 5'-ATGA cg TCAT-3' para el dímero C/C terminal natural (izquierda), o la invertida 5'-TCAT cg ATGA-3' para el dímero no natural N/N-terminal (derecha).

Con el objetivo de obtener un dímero disulfuro con mayor especificidad de secuencia se sintetizó el péptido **4**. A diferencia del anterior, en este caso la terpiridina unida al extremo C-terminal presenta dos grupos metilos para dificultar la intercalación entre las bases. Los estudios de EMSA para los dímeros derivados de este péptido, el $(\mathbf{4})_2\text{Ni}$ y el $(\mathbf{4})_2\text{SS}$, probaron que ahora ambos dímeros sólo interaccionan con su secuencia consenso. Ensayos de anisotropía fluorescente demostraron que este nuevo dímero disulfuro une diez veces mejor con la secuencia inversa que con la directa.

Una vez confirmado que el péptido **4** posee la ortogonalidad de unión al ADN deseada, se exploró su reconocimiento en una situación en que ambas secuencias se encuentran simultáneamente accesibles para la unión. Este experimento se llevó a cabo preparando una disolución que contiene el péptido **4** y dos ADNs de diferente tamaño, uno con la secuencia inversa y el otro la directa. De esta forma, mientras el peptídico monomérico no interacciona con el ADN, una vez que se le añade un agente oxidante (reactivo de Ellman) y Ni(II) se genera el dímero $(\mathbf{4})_2\text{SS}$ que muestra sólo bandas de retardo con el ADN inverso. A continuación si a este dímero se le añade un agente reductor (TCEP), se genera el dímero $(\mathbf{4})_2\text{Ni}$, que sólo une a la secuencia directa. Una nueva oxidación regenera el dímero $(\mathbf{4})_2\text{SS}$ y por lo tanto desplaza su interacción al ADN inverso. A pesar del exceso de reactivos es posible todavía inducir un nuevo intercambio por la adición de TCEP. Este resultado representa la primera prueba de concepto de la viabilidad de hacer moléculas que pueden hacer saltos entre diferentes sitios de unión al ADN, en respuesta a estímulos externos.

SUMMARY





residues not involved in DNA binding were replaced by alanines, while retaining the short *N*-terminal capping motif of GCN4 (DPAAL).

EMSA experiments incubating conjugates **En-HDh3-2** and **En-HDh3-3** with the consensus ds-oligonucleotide (TTTAATCC) showed a new slower-migrating band, consistent with the formation of specific complexes between the peptide conjugates and the dsDNA. Isothermal titrations revealed K_d of 188 nM at 4 °C for **En-HDh3-2** and K_d of 131 nM for **En-HDh3-3**. Importantly, incubation of these hybrids with an oligonucleotide containing a double mutation in its peptide-binding site did not elicit stable complexes, thus confirming highly selective interactions with the target DNA. Furthermore, CD experiments showed that upon addition of the double-stranded oligonucleotide containing the target sequence, both peptides adopted significantly increased helical content.

These results have shown that by combining an appropriate amino acid grafting technique with the tethering of a minor groove binder, helix 3 regions of homeodomain proteins can be induced to bind specific DNA sites with good affinity and very good selectivity. The resulting conjugates simulate the DNA recognition of native homeodomains by binding to similar sequences with a bipartite (major and minor groove) interaction.

Chapter 2: Design of light-switchable DNA-binding peptides.

Cells respond to external signals that modulate a wide variety of processes to adapt themselves to the changing conditions in their environment. One of the key steps where cells carry out this regulation is the transcription, mainly through the control of TF activity. As it was discussed in the introduction, the design of TF analogs that can be activated or deactivated like the natural ones is a relevant issue.

Visible light is considered as an ideal external stimuli to control biomolecules activity because it is not harmful to the cell. So far, several light-activated agents have been developed, however the inverse approach involving light-induced deactivation has been scarcely studied. Herein, we describe a light-sensitive bis-electrophilic dimerizing linker based on a Ru(II) bipyridyl complex containing two thiol-reactive bromoacetyl units. This complex also contains with two amine moieties, which can be released one of them for the action of visible light.

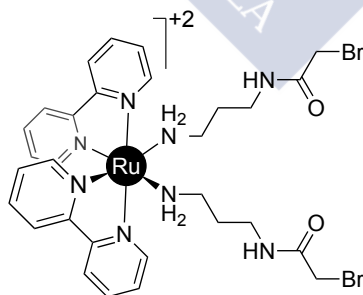


Figure 2. Bis-electrophilic dimerizing linker based on a Ruthenium(II) bipyridyl complex which can be released one amine moiety for the action of visible light.

The artificial photocleavable linker was used to connect two basic regions of the TF GCN4 equipped with *C*-terminal nucleophilic Cys (**Ru(brC)₂**). The reaction between the linker and the peptides takes place by nucleophilic attack of the cysteine sulphur atom to the dibromide complex with unprotected peptide at room temperature in 100 mM aqueous phosphate buffer (pH 7.8). The DNA binding of **Ru(brC)₂** was first studied by EMSA. Thus, when a double

stranded oligonucleotide containing the target ATF/CREB site (ATGA cg TCAT) is incubated with increasing concentrations of **Ru(brC)₂**, we observe new retarded bands consistent with the formation of a specific DNA/Ru(brC)₂ complex. Fluorescence anisotropy titrations using a TAMRA-labeled ATF/CREB confirmed the high affinity interaction of the peptide dimer with the DNA, with an apparent K_d of 12 nM at 20 °C. On the other hand, irradiation of a solution of the complex **Ru(brC)₂** for 20 min with a long wavelength light source (530–550 nm) or for 2 min with a 455 nm LED source led to the almost complete photolysis of the starting complex. Importantly, the ruthenium dimer can also be efficiently photolyzed in the presence of the target DNA without causing damage to the DNA. Therefore EMSA analysis of the mixture of **Ru(brC)₂** and the oligo ATF/CREB after 30 min of irradiation did not show retarded bands corresponding to the Ru(brC)₂–DNA complexes.

Since many biological interactions rely on the formation of heterodimeric assemblies, a heterodimer between two different basic regions of bZIP TFs, GCN4 and C/EBP, was synthesized. The dimer was assembled in a two-step protocol involving an initial reaction between the dibromide linker and a C-terminal Cys derivative of the C/EBP basic region, followed by a second coupling with the GCN4 basic region. This heterodimer shown also slow-migrating bands with its consensus DNA (ATTG cg TCAT) by EMSA and could be effectively disassembled by irradiation with long-wavelength light.

Chapter 3: Development of a reversible supramolecular assembly at specific DNA sites: Ni-promoted, bivalent DNA binding with designed peptide and bipyridyl-bisbenzamidine components.

The DNA recognition process by many TFs is coupled to the folding of their DNA binding domains into well-defined secondary structures, typically α -helices, which create a complementary contacting surface with the DNA major groove. In some cases, such as in the zinc-finger family, the α -helical folding is mediated by coordination of a metal ion (Zn⁺²) to amino acid side chains, typically Cys and His residues.

In this chapter, it is presented a new approach for the specific recognition of DNA sequences, which combines many of the attributes of these previous designs in a single system: bivalence, conditional folding, responsiveness to external stimuli (metal ions), and even reversibility. The strategy, which relies on a planned, dual role for nickel(II) species as both a dimerizing agent and a folding promoter, involves a programmed self-assembly of dissimilar components.

The peptide used for the construction of this system is derived from the GCN4 DNA binding domain, but with the residues Leu230 and Arg234 replaced by His (**brHis₂**). The addition of Ni(II) induces the formation of an α -helix by nucleating the N-terminal turns, while simultaneously providing a coordination site for bis(benzamidine) derivatives equipped with a bipyridine ligand. Given that the bis(benzamidine) binds to the minor groove of adenine/thymine-rich (A/T-rich) DNA sites, the bipyridine unit might be able to recruit the modified GCN4 basic region (**brHis₂**) into the adjacent major groove that features a consensus DNA binding sequence.

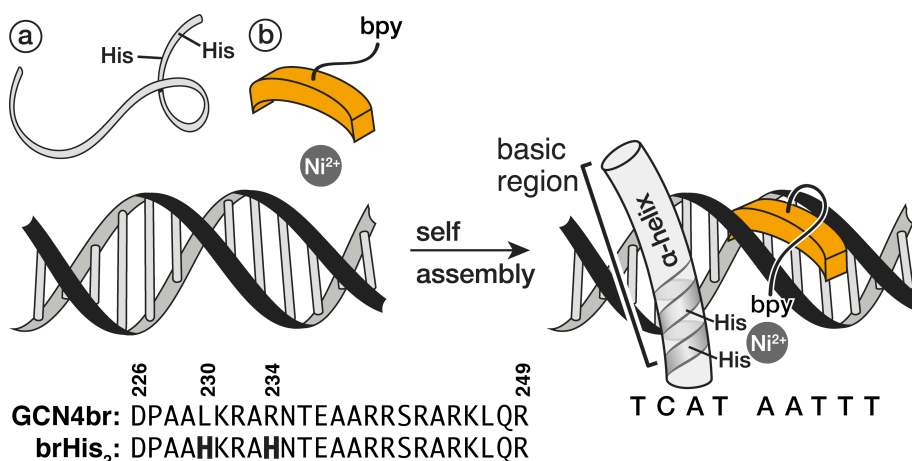


Figure 3. Outline of the cooperative assembly strategy. a) Unfolded **brHis₂**, which will fold into an α -helix in the complex (right). b) Bis(benzamidine) minor-groove binder. bpy: bipyridine chelator tethered to the bis(benzamidine).

The DNA binding properties of the synthetic constructs were studied by EMSA. As expected, incubation of the peptide **brHis₂** in the presence of bis(benzamidines) with the double-stranded oligonucleotide, which contains the peptide target sequence next to an A/T-rich site, does not induce the formation of retarded bands in the gel. However, the addition of increasing concentrations of Ni(II) to the previous mixtures generates retarded bands due to the complex shown in the figure 3. Additional experiments confirms the requirement of all components for the specific DNA binding of the peptide moiety and the system dynamically assembles independently of the order in which the components are mixed. We next performed a forward titration experiment by adding premixed combinations of equimolecular amounts of peptide and bis(benzamidine), and Ni²⁺ (5 equiv) to a ³²P-labeled dsDNA oligonucleotide. Considering the [(brHis₂)(bis(benzamidine)Ni)]²⁺ complex as a single species binding to the DNA, we could calculate an approximate dissociation constant of 0.45 mM at 4 °C. As expected for a poorly structured peptide, the circular dichroism spectrum of **brHis₂** presents low helix content, even in the presence of its cognate DNA, only the addition of all components to the mixture promoted a considerable increase in the peptide helicity.

The sequence selectivity of the system was studied by using other dsDNAs that contain only one specific mutations in the recognition sequence of the peptide or in the sequence of the bis(benzamidine), in both cases no interaction were observed with the DNA. The above observations highlight the high specificity that is attainable through cooperative recognition in noncovalent multicomponent systems. Furthermore, we also studied the selectivity of the system with regard to the metal ions (divalent Co, Ni, Cd, Zn, Cu, Hg, and Fe salts). Using comparable conditions, we only observed the formation of more slowly migrating bands in the presence of Ni(II) cations. Therefore, it seems that nickel has the right coordination properties to promote an effective heterodimeric assembly in the presence of the cognate DNA.

Due to the key role of the nickel ion as an adaptor component that mediates the assembly of the system, the possibility of dismantling the DNA complex by using an external nickel chelator was examined. As chelator agent was used EDTA, the addition of EDTA to the supramolecular complex that resulted from mixing all components promoted a drastic decrease in the helicity of the peptide, which correlates with the disruption of the DNA complex, as demonstrated by EMSA experiments. Therefore, whereas the presence of nickel is critical for the DNA interaction, the resulting multicomponent supramolecular complex can be dismantled by the addition of an external nickel chelator.

Chapter 4: Switching between sites: Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides.

The pattern of gene expression in a multicellular organism is determined by the action of defined extracellular signals that ensure the production of specific proteins only when and where needed. In most of the cases, these signals work by switching the activity of transcription factors (TFs). It is well known that many TFs are unable to bind to their DNA targets by themselves and need to form cooperative multimeric complexes with other partners. This is a clever regulatory mechanism used by nature that allows the combinatorial interaction with a large number of sites using a relatively small set of monomeric transcription regulators.

Despite the advances in switchable DNA binders, designed systems capable of interacting with alternative DNA sites in an externally regulated manner are unknown. Herein we demonstrate for the first time the concept of ‘*site switchability*’ in the DNA binding of synthetic systems, and demonstrate how a single bZIP-based peptide can be directed towards two different DNA sequences depending on its dimerization arrangement. In the last years there have been several examples of synthetic linkers that can be used to replace the C-terminal dimerization unit (the leucine zipper) of the bZIP proteins. In addition, Morii *et. al.* described a synthetic linker to dimerize through the N-terminal position of GCN4, this dimer has selectivity towards the swapped-binding site (TCAT cg ATGA) instead of the normal binding site (ATGA cg TCAT). On these grounds, we designed a GCN4 basic region equipped with orthogonally responsive elements so that either dimerization can be triggered in response to different external stimuli. As C-terminal dimerization unit we chose a terpyridine chelating ligand, which allows promoting the dimerization by addition of a metal ion and formation of a stable metal complex with two terpyridine-tagged basic regions, while providing for reversibility. For the N-terminal dimerization, we selected a disulphide bond, as it could be easily implemented by introduction of a N-terminal cysteine residue. The disulphide bridge is sensitive to oxidizing or reducing conditions, and should be chemically orthogonal to the metal-mediated C-terminal dimerization.

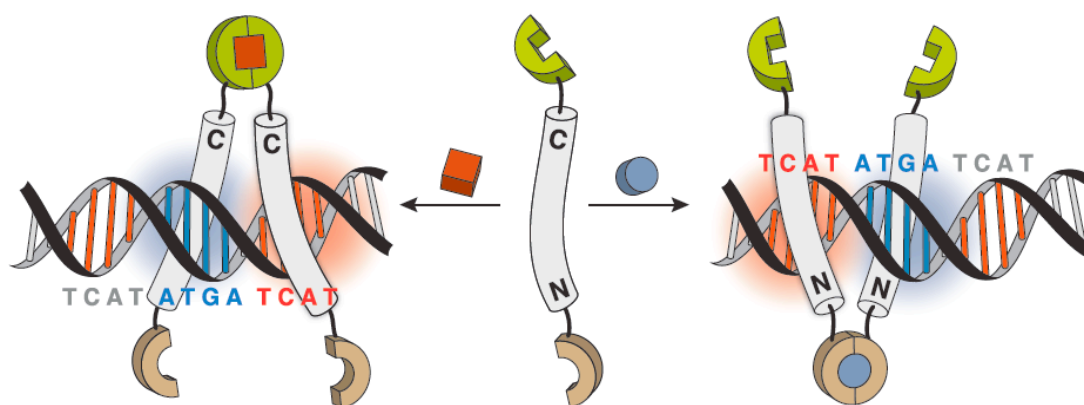


Figure 4. Representation of a stimuli-responsive dimerization strategy for specific binding to two DNA sites. A peptide contains orthogonal functionalities at both termini that in response to different stimuli (represented as a cube or a cylinder) result in alternative dimerization states with different DNA-binding preferences: 5'- ATGA cg TCAT -30 for the natural C/C-terminal dimer (left), or reversed 5'-TCAT cg ATGA-3' for the non-natural N/N-terminal dimer (right).

As a first approximation the peptide **3a**, which have the two dimerization units, was synthesized and its DNA-binding properties were studied by EMSA. As expected, while monomeric peptide **3a** does not bind to short double-stranded (ds)DNA containing the direct ATF/CREB consensus DNA-binding sequence (ATGA cg TCAT), addition of 5 equiv. of

Ni^{2+} results in a new slow-migrating band, consistent with the formation of the specific complex between the $(\mathbf{3a})_2\text{Ni}$ dimer and the DNA. Importantly, this dimer with the inverted target site (TCAT cg ATGA) did not lead to any new bands in the EMSA assays. However, for the *N*-terminal disulphide dimer $(\mathbf{3a})_2\text{SS}$ was observed the formation of a DNA complex with its target inverted DNA but, curiously, it also produces retarded bands with the other oligos with direct sequence and with a half sequence (ATGA). These bands with the non-consensus dsDNAs are more diffuse than those observed with the direct DNA. Further experiments shown that the interaction probably arise from a relatively loose complex in which one of the basic regions is interacting specifically in the major groove of the consensus site while the other makes nonspecific contacts and the terpyridine intercalates between DNA bases.

With the aim of obtaining a disulphide dimer more specific peptide **4** was synthesized. This peptide is equipped with a terpyridine unit featuring two methyl substituents to hinder the intercalation. In this case both dimers $(\mathbf{4})_2\text{Ni}$ y el $(\mathbf{4})_2\text{SS}$ shown only interaction with its consensus DNA. Fluorescence anisotropy titrations with rhodamine-labelled oligonucleotides displayed $(\mathbf{4})_2\text{SS}$ binds the inverted sequence ten times better than the natural ATF/CREB-binding site. Once we confirmed that peptide **4** exhibits the desired orthogonal DNA binding, we explored the sequence-selective recognition in a situation where both target sites are simultaneously available for binding. Thus, a stock solution containing peptide **4** was added to an equimolar mixture of both dsDNAs, containing the direct and the inverse sequences. Note that in this case we use ds-oligonucleotides of different lengths in order to distinguish the binding to either of them. As anticipated, the monomer by itself is unable to elicit retarded shifts; addition of 1 equiv. of Ellman's reagent and 5 equiv. of $\text{Ni}(\text{ClO}_4)_2$ to that solution resulted and in the appearance of a new band consistent with the formation of a specific complex with the inverse DNA. Interestingly, addition of TCEP to the mixture led to the disappearance of this complex and the formation of a new retarded band involving the oligo with the direct sequece, consistent with the DNA binding of the *in situ* assembled dimer $(\mathbf{4})_2\text{Ni}$. A new addition of Ellman's reagent dismounts this complex and regenerates binding to the inverse DNA. Despite the excess of reagents in the mixture, it is yet possible to induce a new switch and regenerate binding to other DNA. These results represent the first proof of concept on the viability of making molecules that can be induced to jump between two different DNA sites, at will, in response to specific external stimuli.

PARTE EXPERIMENTAL





Parte experimental del Capítulo 1

Selective DNA-binding by designed Bisbenzamidine-Homeodomain Chimeras

Jesús Mosquera, Jéssica Rodríguez, M. Eugenio Vázquez, José L. Mascareñas
ChemBioChem **2014**, *15*, 1092.

Supporting material





General

All reagents were acquired from commercial sources: DMF and TFA were purchased from *Scharlau*, CH₂Cl₂ from *Panreac* and CH₃CN from *Merck*, the rest of reagents were acquired from *Sigma-Aldrich*. All peptide synthesis reagents and amino acid derivatives were purchased from GL Biochem (Shanghai) Ltd. and Novabiochem. Amino acids were purchased as protected Fmoc amino acids with the standard side chain protecting scheme, except for the orthogonally protected Fmoc-Lys(Alloc)-OH.

Reactions were followed by analytical RP-HPLC with an *Agilent 1100* series LC/MS using an *Eclipse XDB-C18* analytical column (4.6 x 150 mm, 5 μm). Electrospray Ionization Mass Spectrometry (ESI/MS) was performed with an *Agilent 1100* Series LC/MS model in positive scan mode using direct injection of the purified peptide solution into the MS. Standard conditions for analytical RP-HPLC consisted on an isocratic regime during the first 5 min at 5%, followed by a linear gradient from 5% to 75% of solvent B for 30 min at a flow rate of 1 mL/min (A: water with 0.1% TFA, B: acetonitrile with 0.1% TFA). Compounds were detected by UV absorption at 220, 270, 304 and 330 nm. Purifications were performed by semipreparative RP-HPLC with an *Agilent 1100* series LC using a *Luna 5u C18(2) 100A* (5 μm, 10 × 250 mm) reverse-phase column from *Phenomenex*. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI/MS) was performed with a *Bruker* Autoflex MALDI/TOF model in positive scan mode by direct irradiation of the matrix-absorbed peptide. Concentrations were measured using the listed extinction coefficients.

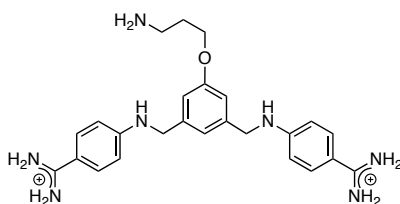
Oligonucleotides were purchased from *Thermo Fisher Scientific GmbH* on a 0.2 mmol scale as freeze-dried solids. After solving in H₂O *milliQ* their concentrations were measured by UV absorption at 260 nm with a *BioRad SmartSpec Plus Spectrophotometer*. Absorbance was measured twice and concentrations were calculated applying Lambert-Beer's equation. The molar extinction coefficients of single strand oligonucleotides were calculated by using the following formula.^[S1]

$$\varepsilon_{(260 \text{ nm})} = \{(8.8 \times \#T) + (7.3 \times \#C) + (11.7 \times \#G) + (15.4 \times \#A)\} \times 0.9 \cdot 10^3 \text{ M}^{-1} \text{ cm}^{-1}$$

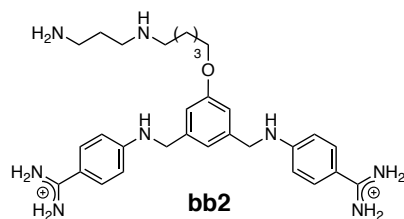
Where #A, #T, #C, #G stand for the number of each type of bases in the DNA strand. Oligonucleotides were hybridized by mixing complementary sequences at equal molar concentration (50 μM), heating at 90 °C for 10 min and then slowly cooling the mixture to rt over 2h.

Minor groove binder synthesis and characterization

Synthesis of **bb1**



bb1 was synthesized according to the procedure reported in the literature.^[S2]

Synthesis of bb2

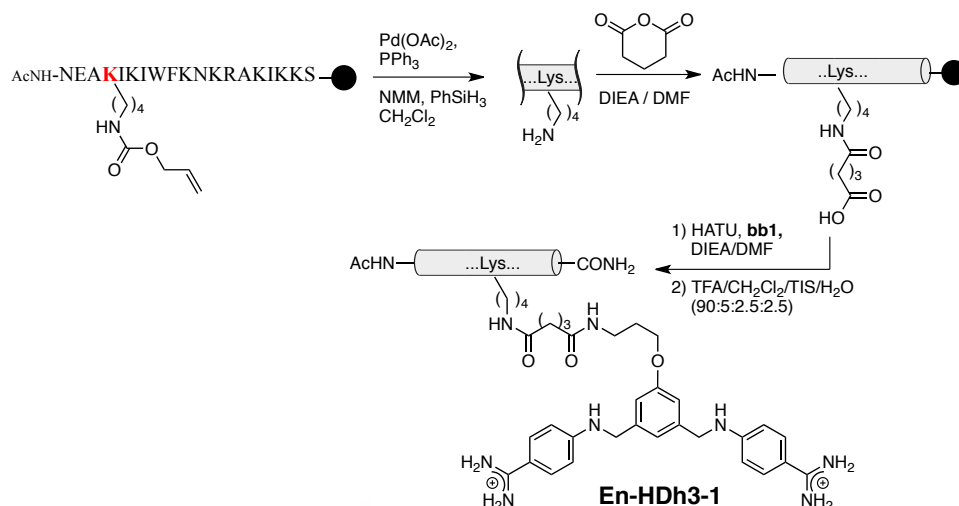
bb2 was synthesized according to the procedure reported in the literature.^[S3]

Peptide Synthesis

C-terminal amide peptides were synthesized following standard peptide protocols (Fmoc/tBu strategy) on a 0.1 mmol scale using a 0.19 mmol/g loading Fmoc-PAL-PEG-PS resin from Applied Biosystems, employing a PS3 automatic peptide synthesizer from Protein Technologies. The amino acids were coupled in 4-fold excess using HBTU as activating agent. Each amino acid was activated for 30 seconds in DMF before being added onto the resin, and couplings were conducted for 30 min. Deprotection of the temporal Fmoc protecting group was performed by treating the resin with 20% piperidine in DMF for 10 min. Manual couplings were monitored using the TNBS (trinitrobenzene sulfonate) test.^[S4]

The resin cleavage-deprotection was accomplished by shaking the resin-bound peptides for 2 h (approx. 0.025 mmol), in 3 mL of the cleavage cocktail (75 μ L of H₂O, 75 μ L of TIS (triisopropylsilane), 150 μ L of DCM and 2700 μ L of TFA). The resin is filtered, and the TFA filtrate is added to ice-cold diethyl ether (30 mL). After 10 min, the precipitate is centrifuged and washed again with 20 mL of ice-cold ether. The solid residue is dried under argon, dissolved in acetonitrile/water 1:1 (2 mL) and purified by semipreparative RP-HPLC. The collected fractions are lyophilized and stored at -20 °C.

Deprotection of Lys(Alloc) side chain. The resin-bound peptide (approx 0.05 mmol) was treated overnight at rt with a deoxygenated mixture of methylmorpholine (55 μ L, 10 eq.), triphenylphosphine (20 mg, 1.5 equiv), phenylsilane (62 μ L, 10 equiv) and Pd(OAc)₂ (3.4 mg, 0.3 equiv) in CH₂Cl₂ (2.7 mL). The resin was filtered and washed with DMF (2 \times 1.5 mL \times 2 min), diethyldithiocarbamate (DEDTC, 25 mg in 5 mL of DMF, 2 \times 1.5 mL \times 5 min), DMF (2 \times 1.5 mL \times 2 min) and CH₂Cl₂ (2 \times 5 mL \times 2 min).

Synthesis of the conjugate En-HDh3-1

The peptide was elongated by the previously described SPPS procedures, and the lysine at position 44 was inserted as Lys(alloc). Once introduced the acetyl group at the N terminus, the protecting alloc group was removed (see Deprotection of Lys(Alloc) side chain). The resin-bound peptide (20 mg, aprox. 2.2 μmol) was suspended in DMF (500 μL), shaken for 15 min and then filtered. Glutaric anhydride (115 μL , 0.15 M in DMF, 18 μmol , 8 equiv), DIEA (35 mL, 0.5 M in DMF, 18 mmol, 8 equiv) and DMAP (26 μL , 0.164 M in DMF, 4.4 mmol, 2 equiv) were added, and the resulting resin suspension was shaken for 1 h at RT. The resin was then filtered and washed with DMF (3 x 0.6 mL). HATU (122 μL , 53 mM in DMF, 6.5 μmol , 3 equiv) and DIEA (35 μL , 0.5 M in DMF, 18 mmol, 8 equiv) were added, and the resulting mixture was shaken for 5 min and filtered. A solution of the bis-benzamidine bb1 (5.2 mg in 215 μL of DMF, 6.6 mmol, 3 equiv) and DIEA (36 μL , 0.5 M in DMF, 18 mmol, 8 equiv) was added. The mixture was shaken for 2 h after which the resin was filtered and washed with DMF (3 x 1 mL, 5 min) and Et₂O (2 x 1 mL, 5 min). The cleavage step was performed following standard conditions, and the desired peptide conjugate was purified by RP-HPLC. After lyophilization the product was obtained as white powders in an approximated 28% yield. Analytical data of the purified products: MS: ESI⁺-MS [M+H⁺]: calcd for C₁₄₀H₂₂₃N₄₀O₂₈ : 2914.0; found: 2913.2; t_R=18.2 min.

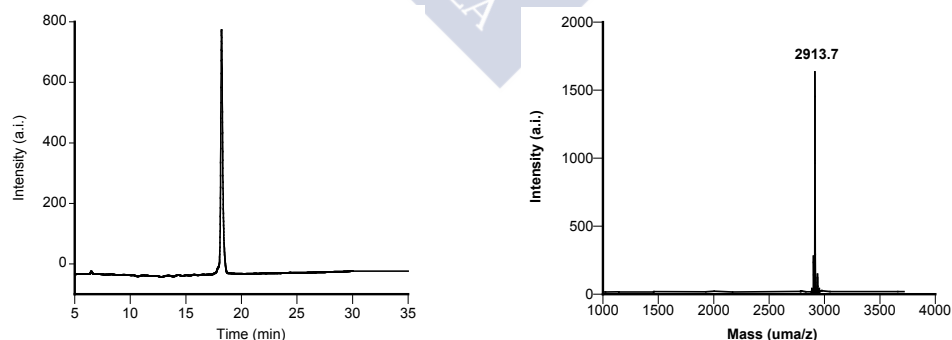
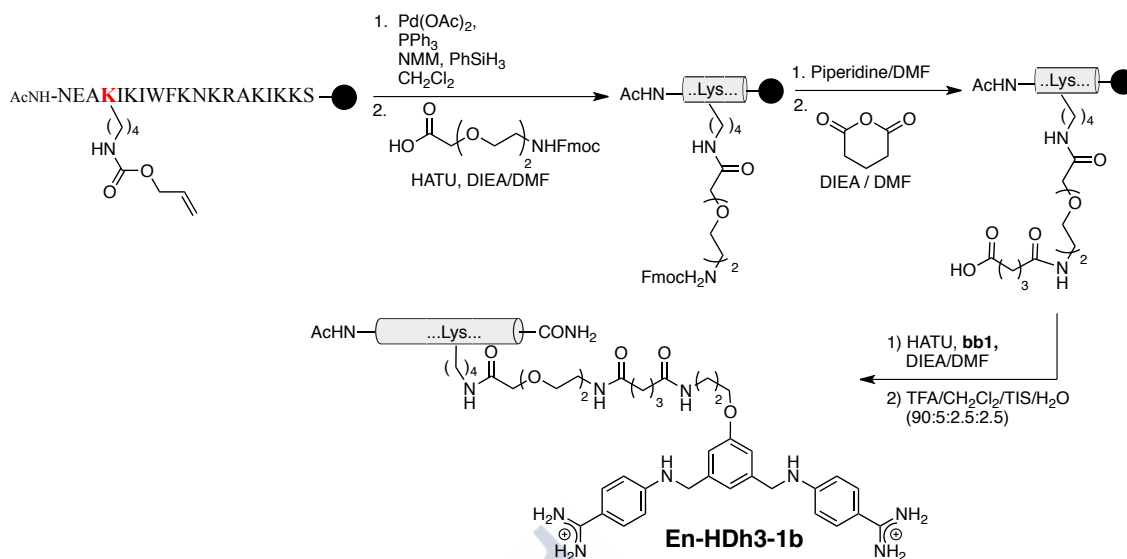


Figure S1. Left: HPLC chromatogram of the purified peptide En-HDh3-1. Right: ESI⁺-MALDI of En-HDh3-1.

Synthesis of the conjugate En-HDh3-1b

The peptide was elongated by the previously described SPPS procedures, and the lysine at position 44 was inserted as Lys(alloc). Once introduced the acetyl group at the N terminus, the protecting alloc group was removed (see Deprotection of Lys(Alloc) side chain). The resin-bound peptide (20 mg, aprox. 2.2 μ mol) was suspended in DMF (500 μ L), shaken for 15 min and then filtered. The amino acid Fmoc-O₂Oc-OH (2.54 mg, 6.6 μ mol, 3 equiv) was coupled using HATU (122 μ L, 53 mM in DMF, 6.6 μ mol, 3 equiv) and DIEA (35 μ L, 0.5 M in DMF, 18 mmol, 8 equiv). After 30 min, the resin was washed three times with DMF and the Fmoc was deprotected with piperidine/DMF (1:5) and washed again with DMF.

A mixture of glutaric anhydride (115 μ L, 0.15 M in DMF, 18 μ mol, 8 equiv), DIEA (35 μ L, 0.5M in DMF, 18 mmol, 8 equiv) and DMAP (26 μ L, 0.164 M in DMF, 4.4 mmol, 2 equiv) was added, and the resulting resin suspension was shaken for 1 h at RT. The resin was then filtered and washed with DMF (3x0.6 mL). HATU (122 μ L, 53 mM in DMF, 6.6 μ mol, 3 equiv) and DIEA (35 μ L, 0.5 M in DMF, 18 mmol, 8 equiv) were added and the mixture was shaken for 5 min and filtered. A solution of the bis-benzamidine **bb1** (5.2 mg in 215 μ L of DMF, 6.6 mmol, 3 equiv) and DIEA (36 μ L, 0.5 M in DMF, 18 mmol, 8 equiv) was added. The mixture was shaken for 2 h after which the resin was filtered and washed with DMF (3 x 1 mL, 5 min) and Et₂O (2 x 1 mL, 5 min). The cleavage step was performed following standard conditions, and the desired peptide conjugate was purified by RP-HPLC. After lyophilization the product was obtained as white powders in an approximated 24% yield. Analytical data of the purified products: MS: ESI⁺-MS [M+H⁺]: calcd for C₁₄₆H₂₃₄N₄₁O₃₁ : 3059.67; found: 3059.4; t_R=17.5 min.

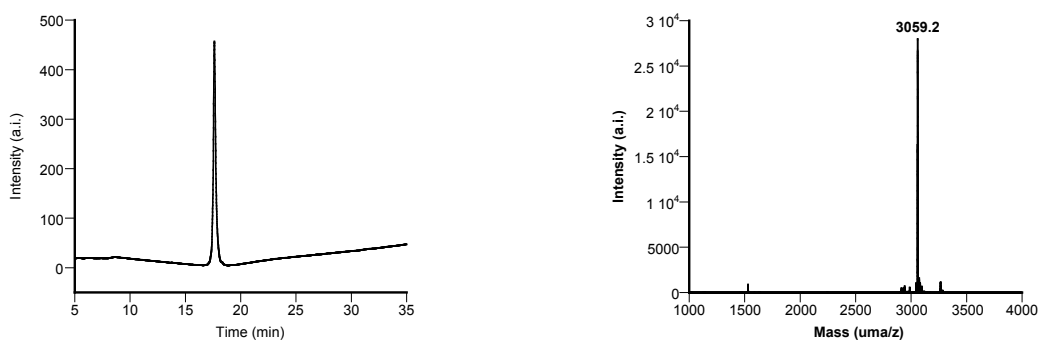
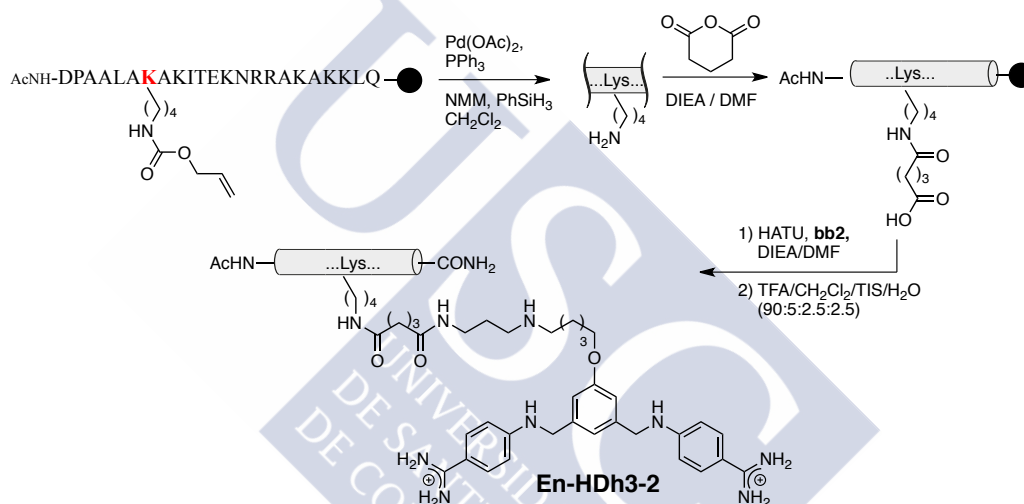


Figure S2. Left: HPLC chromatogram of the purified peptide En-HDh3-1b. Right: ESI⁺-MALDI of En-HDh3-1b.

Synthesis of the conjugate *En-HDh3-2*



The peptide was elongated by the previously described SPPS procedures, and the lysine at position 44 was inserted as Lys(alloc). Once introduced the acetyl group at the N terminus, the protecting alloc group was removed (see Deprotection of Lys(Alloc) side chain). The resin-bound peptide (20 mg, approx. 2.2 μmol) was suspended in DMF (500 μL), shaken for 15 min and then filtered. A mixture of glutaric anhydride (115 μL , 0.15 M in DMF, 17 μmol , 8 equiv), DIEA (35 μL , 0.5 M in DMF, 18 mmol, 8 equiv) and DMAP (26 μL , 0.164 M in DMF, 4.4 mmol, 2 equiv) was added, and the resulting resin suspension was shaken for 1 h at RT. The resin was then filtered and washed with DMF (3x0.6 mL). HATU (122 μL , 53 mM in DMF, 6.6 μmol , 3 equiv) and DIEA (35 μL , 0.5 M in DMF, 18 mmol, 8 equiv) were added, and the resulting mixture was shaken for 5 min and filtered. A solution of the bis-benzamidine bb2 (5.8 mg in 215 μL of DMF, 6.6 mmol, 3 equiv) and DIEA (36 μL , 0.5 M in DMF, 18 mmol, 8 equiv) was added. The mixture was shaken for 2 h after which the resin was filtered and washed with DMF (3 x 1 mL, 5 min) and Et₂O (2 x 1 mL, 5 min). The cleavage step was performed following standard conditions, and the desired peptide conjugate was purified by RP-HPLC. After lyophilization the product was obtained as white powders in an approximated 17% yield. Analytical data of the purified products: MS: ESI⁺-MS [M+H⁺]: calcd for C₁₄₈H₂₅₁N₄₆O₃₄ : 3218.8; found: 3219.1; t_R=17.2 min.

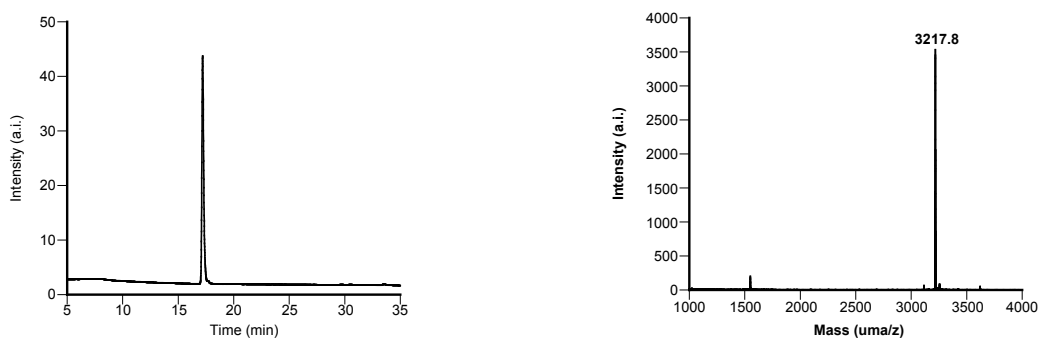
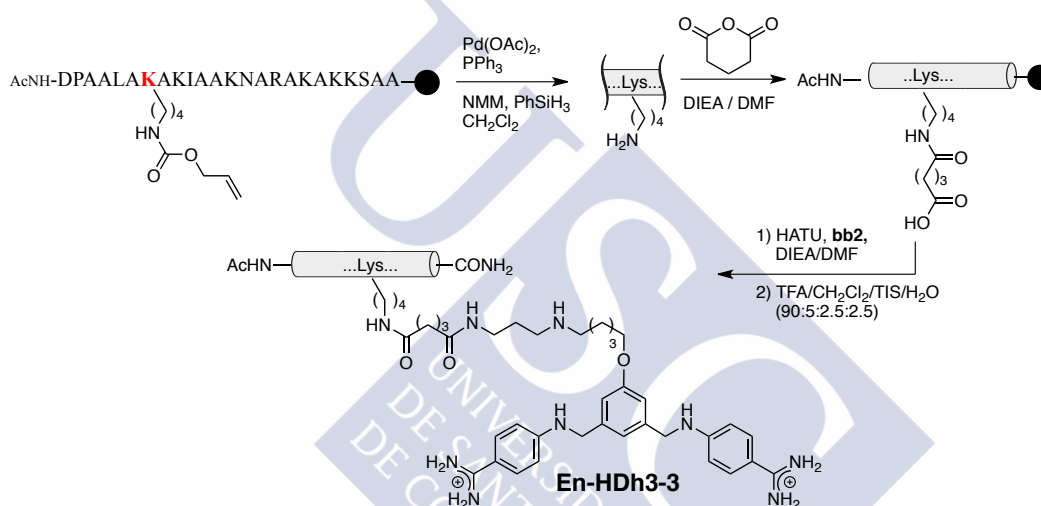


Figure S3. Left: HPLC chromatogram of the purified peptide En-HDh3-2. Right: ESI⁺-MALDI of En-HDh3-2.

Synthesis of the conjugate *En-HDh3-3*



The peptide was elongated by the previously described SPPS procedures, and the lysine at position 44 was inserted as Lys(alloc). Once introduced the acetyl group at the N terminus, the protecting alloc group was removed (see Deprotection of Lys(Alloc) side chain). The resin-bound peptide (20 mg, aprox. 2.2 μmol) was suspended in DMF (500 μL), shaken for 15 min and then filtered. A mixture of glutaric anhydride (115 μL , 0.15 M in DMF, 17 μmol , 8 equiv), DIEA (35 μL , 0.5 M in DMF, 18 mmol, 8 equiv) and DMAP (26 μL , 0.164 M in DMF, 4.6 mmol, 2 equiv) was added, and the resulting resin suspension was shaken for 1 h at RT. The resin was then filtered and washed with DMF (3x0.6 mL). HATU (122 μL , 53 mM in DMF, 6.5 μmol , 3 equiv) and DIEA (35 μL , 0.5 M in DMF, 18 mmol, 8 equiv) were added, and the resulting mixture was shaken for 5 min and filtered. A solution of the bis-benzamidine bb2 (5.8 mg in 215 μL of DMF, 6.6 mmol, 3 equiv) and DIEA (36 μL , 0.5 M in DMF, 18 mmol, 8 equiv) was added. The mixture was shaken for 2 h after which the resin was filtered and washed with DMF (3 x 1 mL, 5 min) and Et₂O (2 x 1 mL, 5 min). The cleavage step was performed following standard conditions, and the desired peptide conjugate was purified by RP-HPLC. After lyophilization the product was obtained as white powders in an approximated 19 % yield. Analytical data of the purified products: MS: ESI⁺-MS [M+H⁺]: calcd for C₁₄₀H₂₃₆N₄₃O₃₂ : 3033.6; found: 3033.0; t_R=15.3 min.

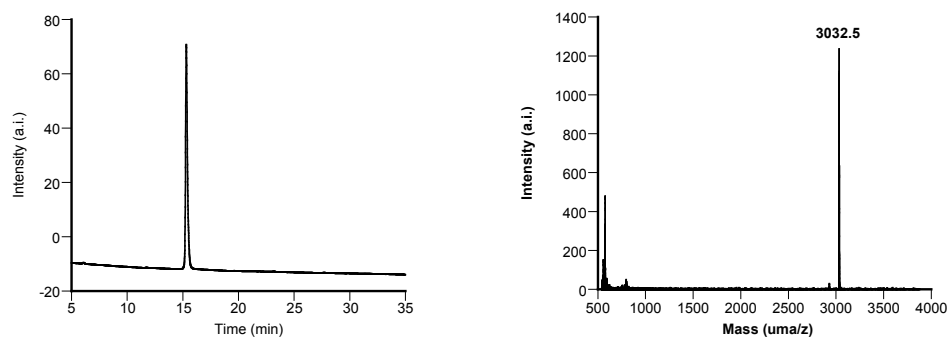


Figure S4. Left: HPLC chromatogram of the purified peptide En-HDh3-3. Right: ESI⁺-MALDI of En-HDh3-3.

Binding studies

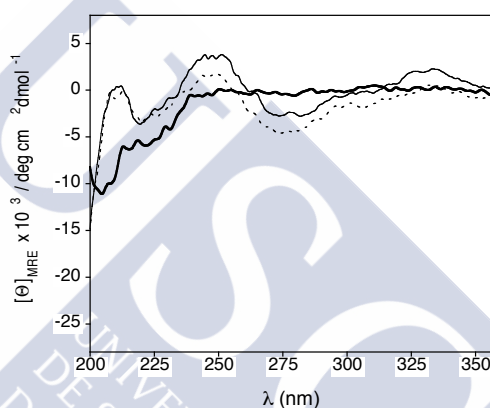
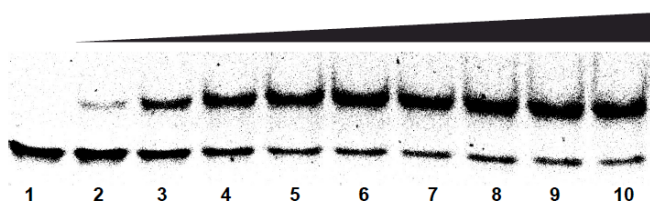


Figure S5. Circular dichroism spectra of a 5 μM solution of HDh3-1b (black solid line); in the presence of 1 equiv of QRE-A/T (solid line); same solution in the presence of 5 μM MUT-A/T DNA (dashed line). The CD spectra of the peptides (when measured in the presence of DNA) were calculated as the difference between the spectrum of the peptide/DNA mixture and the measured spectrum of a sample of the DNA oligonucleotide sequences (only one strand shown): QRE-A/T: 5'-CGTGC GGATTAAA AGCTGCG-3'; MUT-A/T: 5'-CGTGC GACTTAAA GCTGCG-3'.

a)



b)

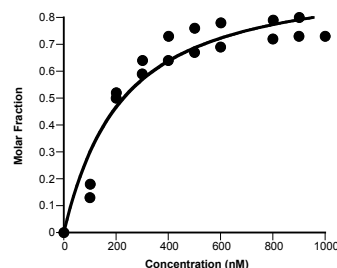


Figure S6. EMSA titration for conjugate En-HDh3-2 with A•T-ERE dsDNA. a) Lanes 1-10: A•T-ERE dsDNA (50 nM); Lanes 1-10: 0, 100, 200, 300, 400, 500, 600, 800, 900 and 1000 nM of peptide En-HDh3-2. b) binding isotherm resulting from densitometry analysis of two different gels. The curve represents the best fit to the data using nonlinear analysis with the DynaFit program^[S5] (Biokib Software) to the equation derived using a 1:1 model ($K_d=188 \pm 45$ nM at 4 $^{\circ}\text{C}$).

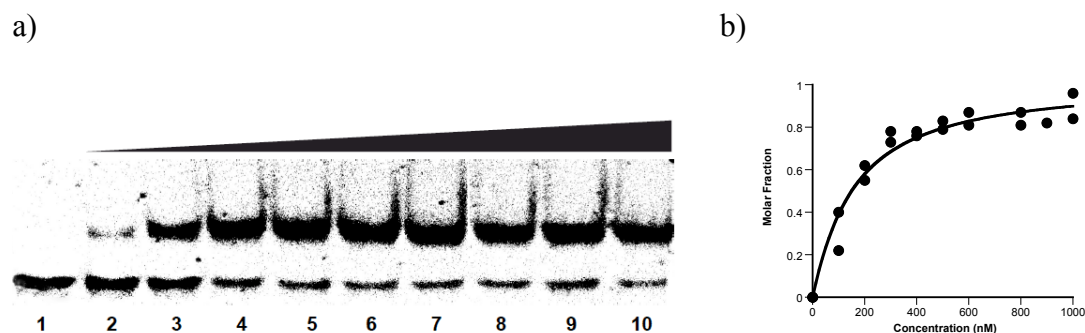


Figure S7. EMSA titration for conjugate En-HDh3-3 with A•T-ERE dsDNA. a) Lanes 1-10: A•T-ERE dsDNA (50 nM); Lanes 1-10: 0, 100, 200, 300, 400, 500, 600, 800, 900 and 1000 nM of peptide En-HDh3-3. b) binding isotherm resulting from densitometry analysis of two different gels. The curve represents the best fit to the data using nonlinear analysis with the DynaFit program^[S5] (Biokib Software) to the equation derived using a 1:1 model ($K_d = 131 \pm 45$ nM at 4 °C).

References

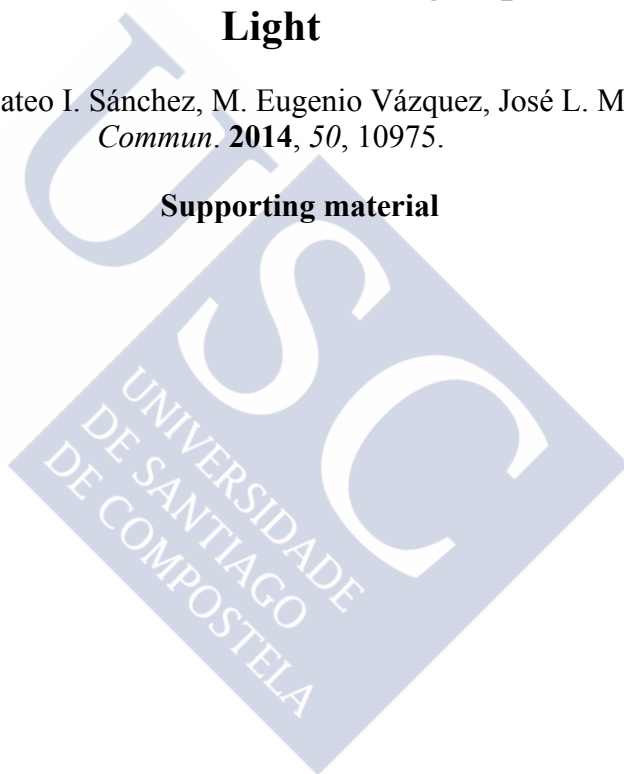
- [S1] K. C. Engman, P. Sandin, S. Osborne, T. Brown, M. Billeter, P. Lincoln, B. Nordén, B. Albinsson, L. M. Wilhelmsson. *Nucl. Acids Res.* 2004, *17*, 5087-5095. b) G. Kallansrud, B. Ward. *Anal Biochem.* 1996, *236*, 134-138.
- [S2] J. Bordello, M. I. Sánchez, M. E. Vázquez, J. L. Mascareñas, W. Al-Soufi, M. Novo. *Angew. Chem. Int. Ed.* 2012, *51*, 7541-7544.
- [S3] M. I. Sánchez, O. Vázquez, M. E. Vázquez, J. L. Mascareñas, *Chem. Eur. J.* 2013, *19*, 9923-9929.
- [S4] W. Hancock, J. Battersby, *Anal. Biochem* 1976, *71*, 260–264.
- [S5] P. Kuzmic, *Proteinase Anal. Biochem.* 1996, *237*, 260-273.

Parte experimental del Capítulo 2

Ruthenium Bipyridyl Complexes as Photocleavable Dimerizers: Deactivation of DNA-Binding Peptides Using Visible Light

Jesús Mosquera, Mateo I. Sánchez, M. Eugenio Vázquez, José L. Mascareñas, *Chem. Commun.* **2014**, 50, 10975.

Supporting material





General

All reagents were acquired from commercial sources: DMF and TFA were purchased from *Scharlau*, CH₂Cl₂ from *Panreac*, CH₃CN from *Merck*, Ruthenium (III) Chloride was purchased from Johnson Matthey. The rest of reagents were acquired from *Sigma-Aldrich*. All peptide synthesis reagents and amino acid derivatives were purchased from *GL Biochem* (Shanghai) Ltd. and *NovaBiochem*. Amino acids were purchased as protected Fmoc amino acids with the standard side chain-protecting groups.

Reactions were followed by analytical RP-HPLC with an *Agilent 1100* series LC/MS using an *Eclipse XDB-C18* analytical column (4.6 x 150 mm, 5 μm). Electrospray Ionization Mass Spectrometry (ESI/MS) was performed with an *Agilent 1100* Series LC/MSD model in positive scan mode using direct injection of the purified peptide solution into the MS. Standard conditions for analytical RP-HPLC consisted on an isocratic regime during the first 5 min, followed by a linear gradient from 5% to 75% of solvent B for 30 min at a flow rate of 1 mL/min (A: water with 0.1% TFA, B: acetonitrile with 0.1% TFA), compounds **1** and **2** were analyzed in the same equipment but with a linear gradient 5 to 95% of B en 30 min.. Compounds were detected by UV absorption at 220, 270, 304 and 330 nm. Purification were performed by semipreparative RP-HPLC with an *Agilent 1100* series LC using a *Luna 5u C18(2) 100A* (5 μm, 10 × 250 mm) reverse-phase column from *Phenomenex*. Concentrations were measured using the listed extinction coefficients.

Oligonucleotides were purchased from *Thermo Fisher Scientific GmbH* on a 0.2 mmol scale as freeze-dried solids. After solving in H₂O *milliQ* their concentrations were measured by UV absorption at 260 nm with a *BioRad SmartSpec Plus Spectrophotometer*. Absorbance was measured twice and concentrations were calculated applying Lambert-Beer's equation. The molar extinction coefficients of single strand oligonucleotides were calculated by using the following formula,^[S1]

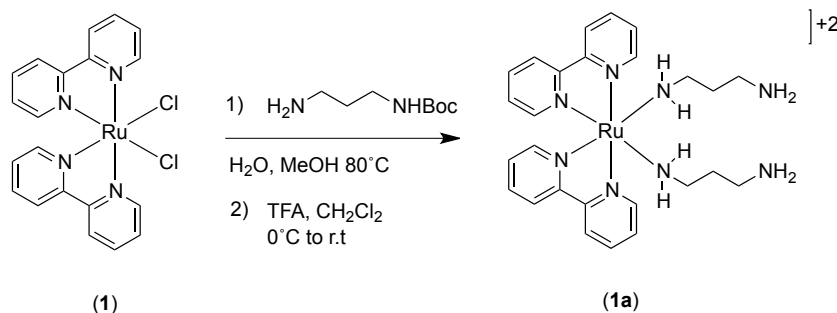
$$\varepsilon_{(260\text{ nm})} = \{(8.8 \times \#T) + (7.3 \times \#C) + (11.7 \times \#G) + (15.4 \times \#A)\} \times 0.9 \cdot 10^3 \text{ M}^{-1} \text{ cm}^{-1}$$

Where #A, #T, #C, #G stand for the number of each type of bases in the DNA strand. Oligonucleotides were hybridized by mixing complementary sequences at equal molar concentration, heating at 90 °C for 10 min and then slowly cooling the mixture to rt over 1h.

Compound **1** was synthesized following known protocols.^[S2] Both diastereoisomers are employed.

Synthesis of the ruthenium-based linker

Bis(2,2'-bipyridine) bis(N-(propane-1,3-diamine) ruthenium(II) (1a)



A solution of *cis*-dichlorobis(2,2'-bipyridine)ruthenium(II) (500 mg, 1.03 mmol) was suspended in 50 mL of distilled water; N₂ was bubbled for 15 min, and the suspension was heated at 80 °C until the precursor was completely dissolved. *tert*-Butyl *N*-(3-aminopropyl)carbamate (1.650 g, 9.5 mmol 10.3 mmol) was dissolved in a small amount of EtOH, and was added onto the ruthenium solution. The mixture was stirred at 80 °C overnight. The reaction was filtered to remove any insoluble particles and, after cooling, precipitated with a NH₄PF₆ saturated solution. The brown solid was dissolved in DMSO and purified by semipreparative RP-HPLC (gradient: 5% B 5 min, 5% to 65 % B 40 min). The combined fractions were concentrated and freeze-dried to obtain the desired product as a brown trifluoroacetate salt (509 mg, 0.51 mmol, 56% yield).^[S3]

The intermediate (100 mg, 0.101 mmol) was dissolved in CH₂Cl₂ (5 mL) in a 25 mL round bottom flask and the mixture was cooled to 0 °C. TFA (5 mL) was slowly added and the resulting mixture was stirred at 0 °C for 5 minutes, and 35 min at rt. The solvent was removed and co-evaporated with toluene to give **1a** as a brown solid.

¹H NMR (400 MHz, MeOD-*d*₄ δ): 1.76-1.96 (m, 8H), 2.74 (t, *J* = 7.3 Hz, 4H), 4.26 (td, *J* = 21.3, 11.3 Hz, 2H), 7.20 (ddd, *J* = 7.1, 5.8, 1.1 Hz, 2H), 7.49 (d, *J* = 5.7 Hz, 2H), 7.80-7.87 (m, 4H) 8.23 (dt, *J* = 7.9, 1.2 Hz, 2H), 8.46 (d, *J* = 8.2 Hz, 2H), 8.65 (d, *J* = 8.2 Hz, 2H), 9.24 (d, *J* = 5.4 Hz, 2H).

¹³C NMR (MeOD-*d*₄ δ): 30.93 (CH₂, broad peak), 37.8 (CH₂), 43.3 (CH₂, broad peak), 124.8 (CH), 125.2 (CH), 127.8 (CH), 128.6 (CH), 137.3 (CH), 138.4 (CH), 153.3 (CH), 153.9 (CH), 159.7 (C), 159.9 (C).

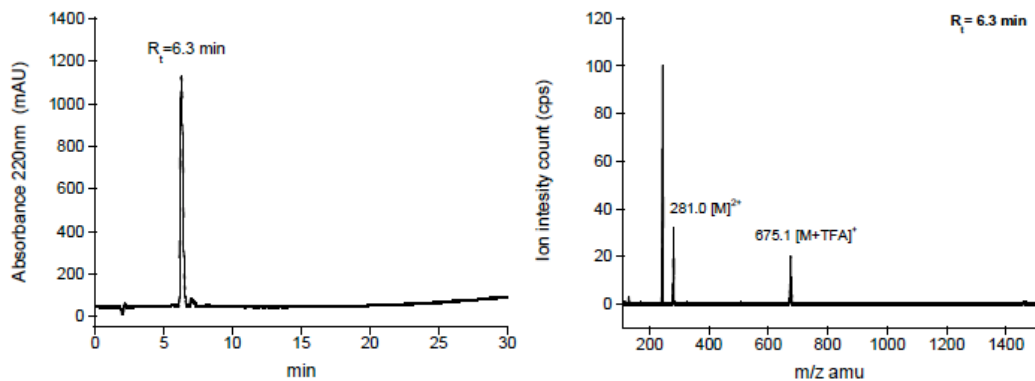
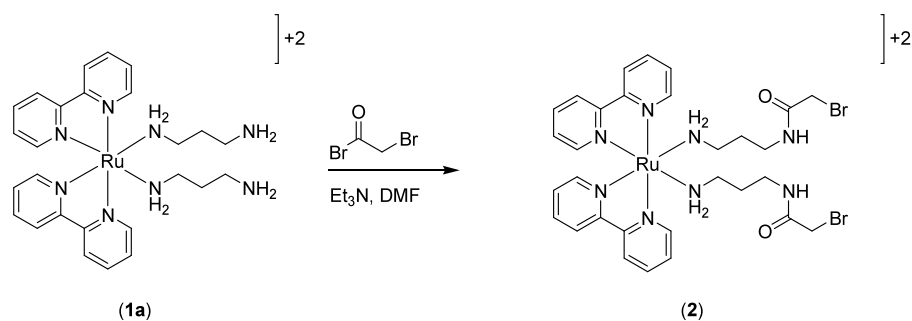


Figure S1. Left: HPLC chromatogram of the crude product (**1a**) after deprotection (220 nm). Right: ESI⁺-MS profile. Calculated ESI-MS: 675.6 calcd. for [M+TFA]⁺, 281.3 calcd. for [M]²⁺, 243.3 calcd. for [M-aminopropyl]²⁺.

Bis(2,2'-bipyridine) bis(N-(3-aminopropyl)-2-bromoacetamide) ruthenium(II) (2)

Bromoacetyl bromide (20.5 mg, 0.101 mmol) was added to a solution of **1a** (38 mg, 0.048 mmol), Et₃N (19.5 mg, 0.192 mmol) in 0.5 ml of DMF. After stirring for 1 h at room temperature, the crude was diluted in water (1:5) and purified by semipreparative RP-HPLC (gradient: 5% B 5 min, 5% to 50 % B 40 min). The combined fractions were concentrated and freeze-dried to obtain the desired product **2** as a brown trifluoroacetic salt (28 mg, 0.028 mmol, 58% yield).

¹H NMR (400 MHz, MeOD-*d*₄ δ): 1.57-1.62 (m, 4H), 1.71-1.76 (m, 4H), 3.13-3.19 (m, 2H), 3.23-3.27 (m, 2H), 4.05-4.11 (m, 2H), 4.16-4.23 (m, 2H), 7.21 (t, *J* = 6.6 Hz, 2H), 7.52 (d, *J* = 5.6 Hz, 2H), 7.82 (t, *J* = 7.8 Hz, 2H), 7.91 (t, *J* = 6.5 Hz, 2H), 8.25 (t, *J* = 7.8 Hz, 2H), 8.46 (d, *J* = 8.1 Hz, 2H), 8.65 (d, *J* = 8.1 Hz, 2H), 9.23 (d, *J* = 5.5 Hz, 2H).

¹³C NMR (MeOD-*d*₄ δ): 28.8 (CH₂), 32.8 (CH₂), 36.7 (CH₂), 42.8 (CH₂), 124.7 (CH), 125.2 (CH), 127.7 (CH), 128.8 (CH), 137.1 (CH), 138.3 (CH), 153.4 (CH), 153.5 (CH), 159.7 (CH), 159.8 (CH), 170.2 (C).

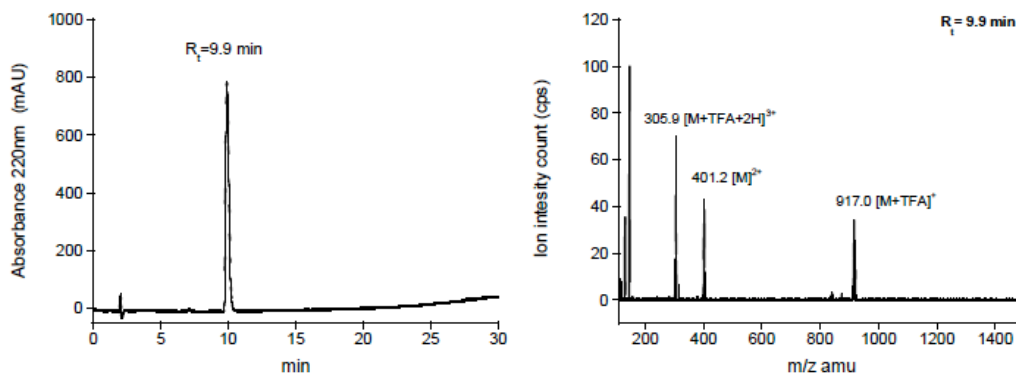


Figure S2. Left: HPLC chromatogram of the pure compound **2** after purification (220 nm). Right: ESI⁺-MS profile. Calculated ESI-MS: 917.0 calcd. for [M+TFA]⁺, 402.1 calcd. for [M]²⁺, 306.2 calcd. for [M+TFA+2H]²⁺.

Peptide Synthesis

C-terminal amide peptides were synthesized following standard peptide protocols (Fmoc/tBu strategy) on a 0.1 mmol scale using a 0.19 mmol/g loading Fmoc-PAL-PEG-PS resin from Applied Biosystems, using a PS3 automatic peptide synthesizer from Protein Technologies. The amino acids were coupled in 4-fold excess using HBTU as activating agent. Each amino acid was activated for 30 seconds in DMF before being added onto the resin, and couplings were conducted for 30 min. Deprotection of the temporal Fmoc protecting group was performed by treating the resin with 20% piperidine in DMF for 10 min. Manual couplings were monitored using the TNBS (trinitrobenzene sulfonate) test.^[S4]

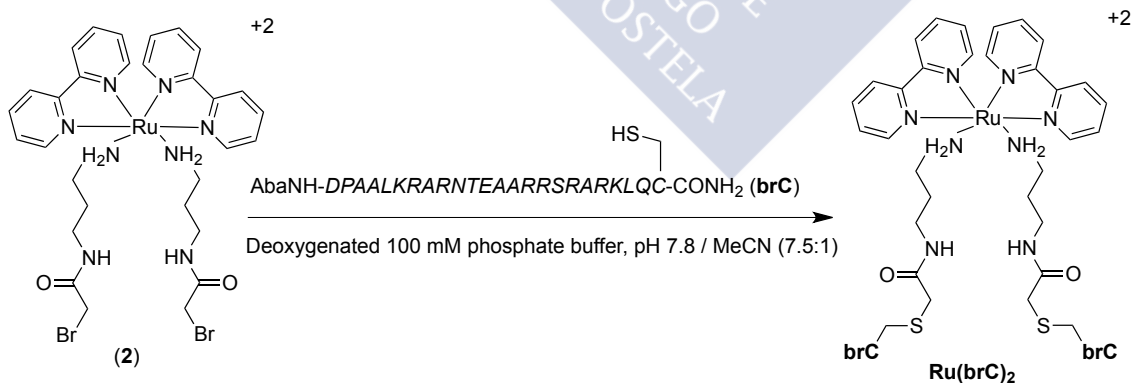
The resin cleavage-deprotection was accomplished by shaking the resin-bound peptides for 2 h (approx. 0.025 mmol), in 3 mL of the cleavage cocktail (25 μ L of EDT (1,2-ethanedithiol), 25 μ L of H₂O, 10 μ L of TIS (triisopropylsilane) and 940 μ L of TFA). The resin is filtered, and the TFA filtrate is added to ice-cold diethyl ether (30 mL). After 10 min, the precipitate is centrifuged and washed again with 20 mL of ice-cold ether. The solid residue is dried under argon, dissolved in acetonitrile/water 1:1 (2 mL) and purified by semipreparative RP-HPLC. The collected fractions are lyophilized and stored at -20 °C.

After lyophilization, the peptides were obtained as white powders. **brC**: 28.1 mg, 7.6 μ mol, approx. 30% yield, and **ebpC**: 28.2 mg, 6.7 μ mol, approx. 27%.

The synthesis of **(brGGC)₂SS** and **(brGGC)SH** peptides has been previously reported.^[S5]

Synthesis of the ruthenium-peptide conjugates

Synthesis of Ru(**brC**)₂



The purified peptide **brC** (6 mg, approx. 1.6 μ mol) and the ruthenium linker **2** (0.6 mg, 0.54 μ mol) were dissolved in a mixture of deoxygenated 100 mM phosphate buffer, pH 7.8 (150 μ L) and acetonitrile (20 μ L). The resulting mixture was stirred for 1 h at rt and the crude purified by RP-HPLC. The combined HPLC fractions containing the product were freeze-dried to give the peptide conjugate as an orange powder (1.4 mg, 0.16 μ mol, 30 %).

ESI-MS [MH]⁺ calcd. for C₂₇₀H₄₅₂N₁₀₄O₇₀RuS₂ = 6440.3, found = 6440.0

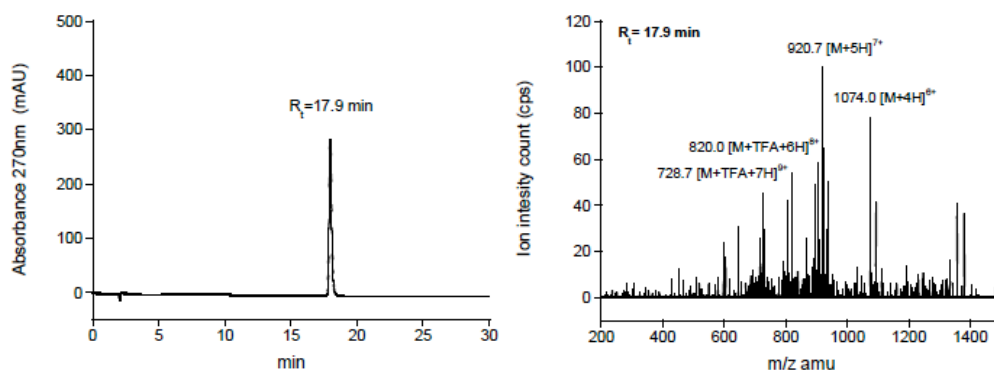
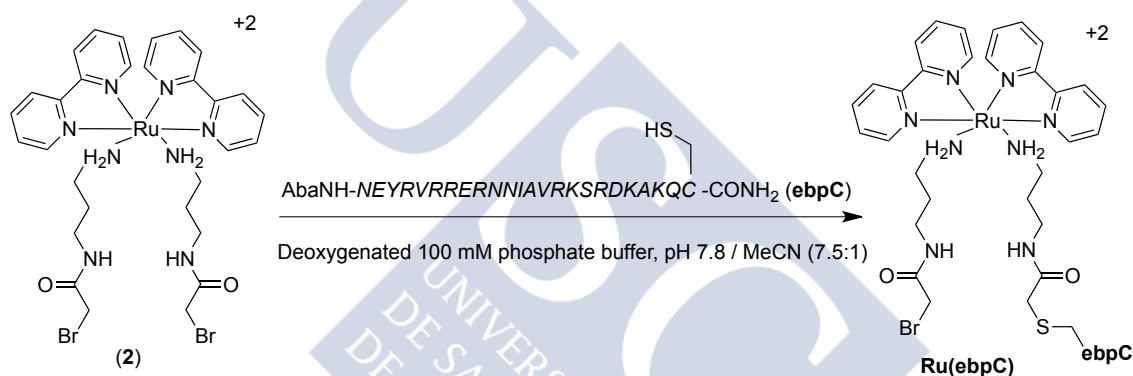


Figure S3. Left: HPLC chromatogram of the purified peptide **Ru(brC)₂** (270 nm). Right: ESI⁺-MS of peptide **Ru(brC)₂**. Calculated ESI-MS: 1074.1 calcd. for [M+4H]⁶⁺, 920.6 calcd. for [M+5H]⁷⁺, 820.0 calcd. for [M+TFA+6H]⁸⁺, 729.0 calcd. for [M+5H]⁷⁺. Only the largest peaks are assigned for clarity. Most of the smaller peaks are consistent with product ions with different numbers of TFAs.

Synthesis of Ru(ebpC)



The purified peptide **ebpC** (3 mg, approx. 0.72 μmol) was dissolved in a deoxygenated 100 mM phosphate buffer, pH 7.8 (150 μL). The ruthenium linker **2** (1.5 mg, 0.54 μmol) was dissolved in acetonitrile (20 μL) and then added to the **ebpC** solution. The mixture was stirred for 1 h at rt and the resulting crude purified by RP-HPLC. After lyophilization the product was obtained as an orange powder (0.72 mg, 0.14 μmol , 26 %).

ESI-MS [MH]⁺ calcd. for C₁₆₁H₂₅₇BrN₅₈O₄₁RuS = 3875.1, found = 3874.5.

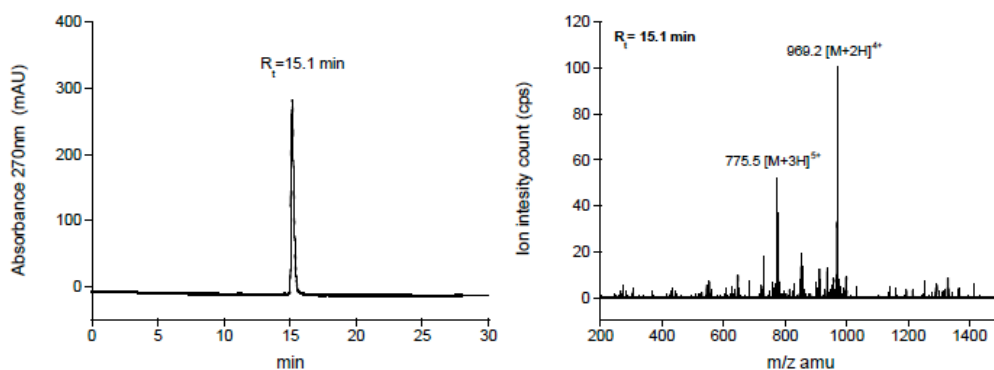
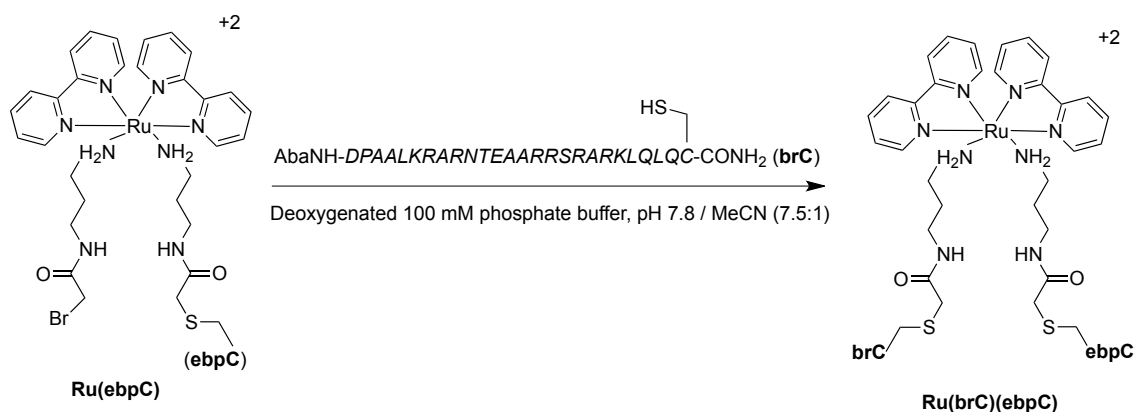
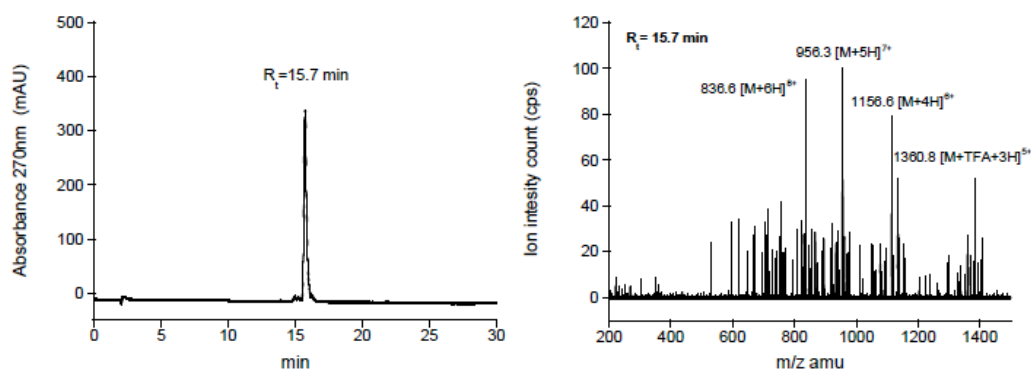


Figure S4. Left: HPLC chromatogram of the purified peptide **Ru(ebpC)** (270 nm). Right: ESI⁺-MS of peptide **Ru(ebpC)**. Calculated ESI-MS: 969.0 calcd. for [M+4H]⁶⁺, 775.4 calcd. for [M+5H]⁵⁺.

Synthesis of **Ru(brC)(ebpC)**

The purified peptides **Ru(ebpC)** (1.2 mg, approx. 0.24 μ mol) and **brC** (2.2 mg, approx. 0.6 μ mol) were dissolved in a deoxygenated 100 mM phosphate buffer, pH 7.8 (200 μ L). The mixture was stirred for 1 h at rt and the resulting crude purified by RP-HPLC. After lyophilization the product was obtained as an orange powder (3.1 mg, 0.36 μ mol, approx. 60 %).

ESI-MS [MH]⁺ calcd. for C₂₈₁H₄₆₄N₁₀₆O₇₅RuS₂ = 6692.5, found = 6691.3



Photolysis experiments

The photolysis experiments were carried out using a fluorescence microscope with a 12 V/100 W halogen bulb lamp, by irradiation through the 530-550 nm excitation filter (green channel).

A solution of **Ru(brC)₂** (60 μ M in milliQ water) was exposed to green light, after 30 min the sample was analyzed by HPLC-MS.

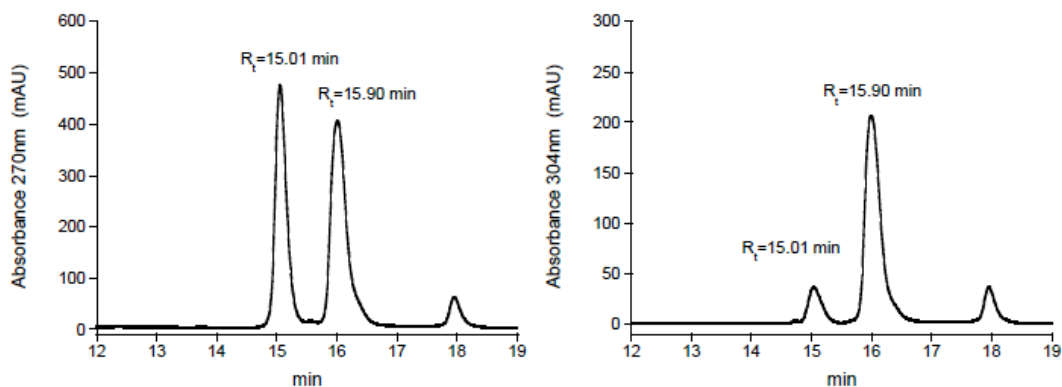


Figure S6. Chromatogram of the irradiated sample, observed at 270 nm (left) and 304 nm (right).

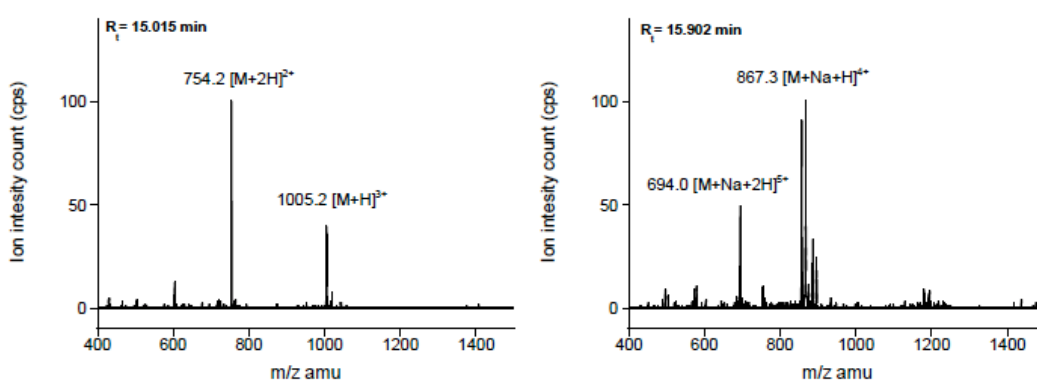


Figure S7. Ion trace of the peak at the indicated retention time: left (**brC-NH₂**), right (**brC-Ru(bipy)₂OH₂**).

A solution of **Ru(brC)(ebpC)** (10 μ M in milliQ water) was exposed to green light, after 30 min the sample is analyzed by HPLC-MS.

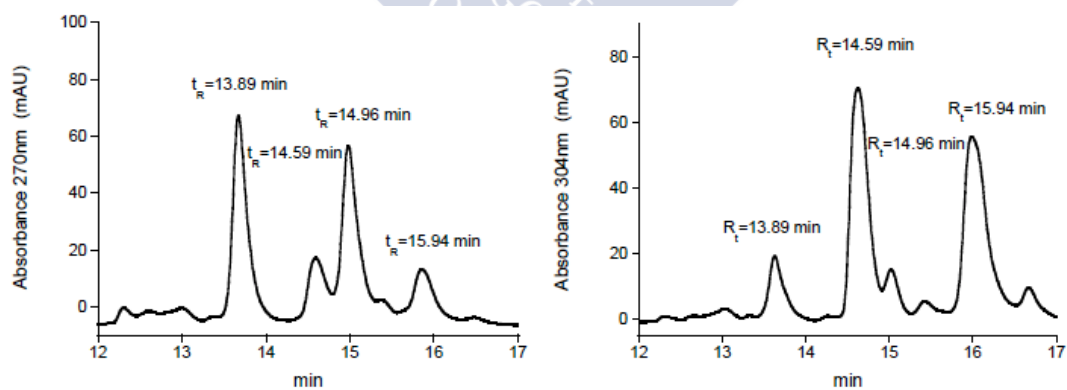


Figure S8. Chromatogram of the irradiated sample, observed at 270 nm (left) and 304 nm (right).

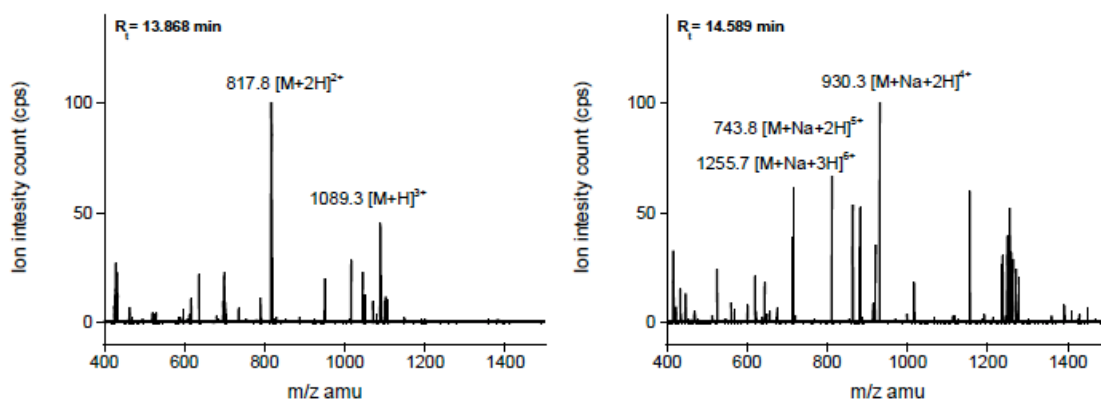


Figure S9. Ion trace of the peak at the indicated retention time: left (**ebpC-NH₂**), right (**ebpC-Ru(bipy)₂OH₂**).

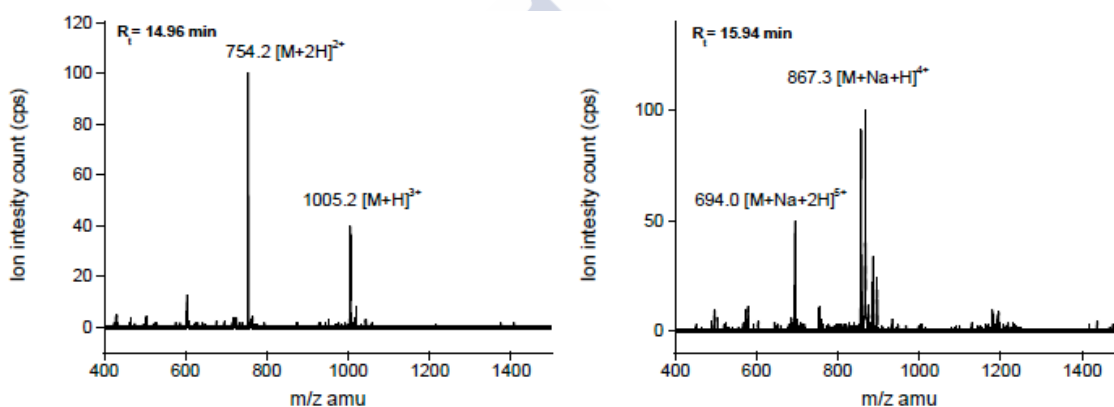


Figure S10. Ion trace of the peak at the indicated retention time: left (**brC-NH₂**), right (**brC-Ru(bipy)₂OH₂**).

The photolysis can be carried much faster using a LED source at 455 nm (900 mW). In this way, the photolysis is completed in only 2 minutes as shown in **figure S11**. This is in part a consequence of the maximum absorption for the bipyridine ruthenium complexes is about 460 nm.^[S6]

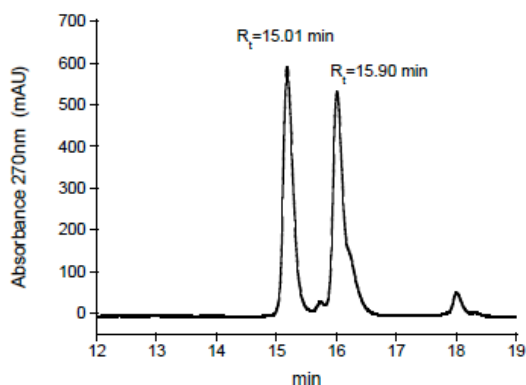


Figure S11. Chromatogram of the irradiated **Ru(brC)₂** using a 455 nm LED source for 2 minutes, observed at 270 nm.

DNA integrity after photolysis

MALDI experiments:

A solution of **Ru(brC)₂** (100 μ M) and the ds-oligonucleotide **ATF/CREB** (100 μ M) in Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl was exposed to green light. After 30 min the sample was analyzed by MALDI-TOF on a Bruker Ultraflex-III TOF/TOF. A ZipTip treatment was carried out before the MALDI experiments.

The masses of the oligonucleotide strands are similar to those observed for the untreated sample.

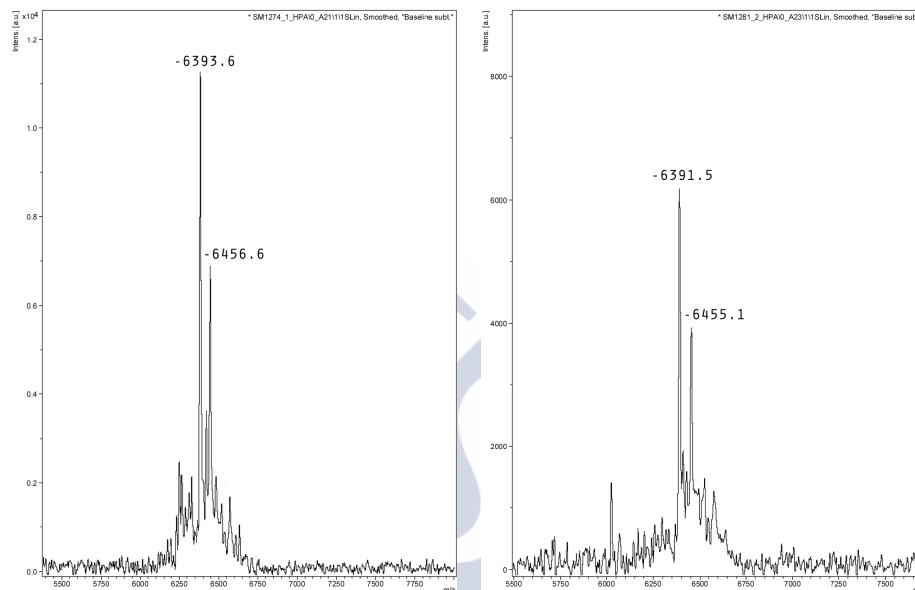


Figure S12. Right: MALDI-TOF of a solution of **ATF/CREB**. Left: MALDI-TOF of a solution of **Ru(brC)₂** (100 μ M) and **ATF/CREB** ds-DNA (100 μ M) which was exposed to green light for 30 min. Calculated $[M-H]^{-1}$ for strand A (CGGATGACGTCATTTTTTTC): 6393.1, calculated $[M-H]^{-1}$ for strand B (GAAAAAAATGACGTCATCCG): 6456.3.

Additional control experiment

The plasmid pGL3-Luc (bought in Promega Company) was digested with restriction enzymes KpnI and HindIII to generate compatible ends and ligated to the insert 5'- C atgacgtcat aatct atgacgtcat aatct atgacgtcat aatct atgacgtcat aatct atgacgtcat A- 3', complementary strand: 5'- AGCTT atgacgtcat agatt atgacgtcat agatt atgacgtcat agatt atgacgtcat agatt atgacgtcat GGTAC- 3' (Bases in upper case are the target sites for restriction enzymes, the insert contains 5 ATF/CREB binding sites).

The resulting plasmid (pGL3-CRE-Luc) was used to test the possibility of base oxidation upon photoirradiation in the presence of Ru(II)-peptide complex. The experiment consisted on the incubation of 1 μ M of **Ru(brC)₂** with 27.6 μ g/mL of the plasmid in 10 mM Na-cacodylate buffer pH 7.5, followed by irradiation of the mixture for 20 min with a long-wavelength light source (530-550 nm). Then, the mixture was treated under standard conditions to detect oxidative damaged in plasmids (0.1 M piperidine at 65°C for 5 min)^[S7] and analysed by 1% agarose gel electrophoresis.

As shown in Figure S13, the treatment doesn't generate new bands in the gel (lane 3), that

there is not any oxidative damage during the photoirradiation, in contrast to that observed when the plasmid is treated with Fe^{+2} in the presence of oxygen (lane 4).^[S8]

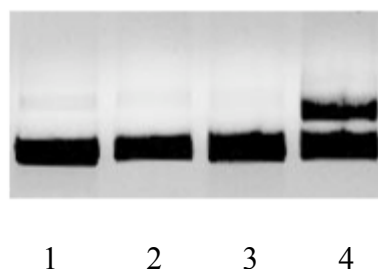


Figure S13. 1% Agarose gel electrophoresis. Lanes 1-4: 27.6 $\mu\text{g}/\text{mL}$ of the pGL3-CRE-Luc plasmid; lane 1: Only plasmid; lane 2: Plasmid and 1 μM of $\text{Ru}(\text{brC})_2$ after being treated with 0.1 M piperidine for 5 min at 65 $^\circ\text{C}$; lane 3: 1 μM of $\text{Ru}(\text{brC})_2$ after being irradiated (20 min with green light) and treated with 0.1 M piperidine for 5 min at 65 $^\circ\text{C}$; lane 4: plasmid with 50 μM of Fe^{+2} .

Fluorescence Anisotropy

Measurements were made with a Jobin-Yvon Fluoromax-3, (DataMax 2.20) coupled to a Wavelength Electronics LFI-3751 temperature controller, using the following settings: integration time: 1.0 s; excitation slit width: 3.0 nm; emission slit width: 6.0 nm; excitation wavelength 559 nm; emission wavelength 585 nm.

TAMRA-ATF/CREB (TAMRA-5'-CGG ATGACGTCAT TTTTTTC-3', one strand shown)^[S9] (5 μL , 5 μM) was added to 995 μL of Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl, and the anisotropy was measured. Aliquots of a stock solution in the same buffer of the dimer $\text{Ru}(\text{brC})_2$ (12.5 μM) were successively added to this solution, and the anisotropic value was recorded after each addition.

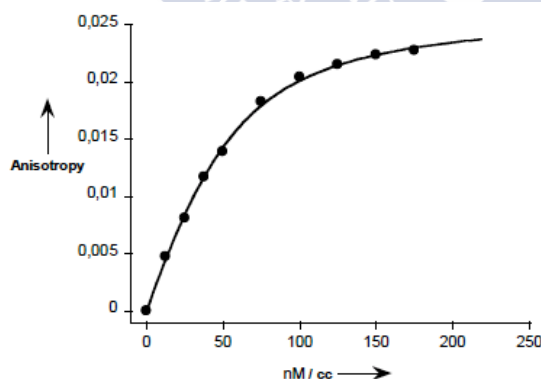


Figure S14. Plot of the anisotropic values of **TAMRA-ATF/CREB** recorded at 585 nm against the total concentration of $\text{Ru}(\text{brC})_2$, and best-fit curve to a 1:1 binding mode with an estimated K_d of 12.0 \pm 0.3 nM. Average of three replicates.

TAMRA-ATF/CREB (5 μL , 5 μM) was added to 995 μL of Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl, and calf thymus DNA (50 μM in base pairs), and the anisotropy was measured. Aliquots the stock solution of the dimer $\text{Ru}(\text{brC})_2$ were successively added to this solution, and the anisotropic value was recorded after each addition.

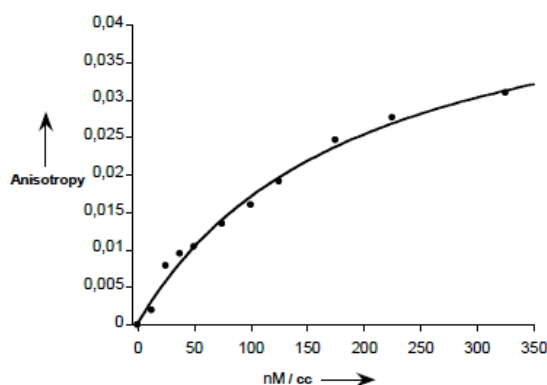


Figure S15. Plot of the anisotropic values of TAMRA-ATF/CREB in the presence of calf thymus DNA recorded at 585 nm against the total concentration of $\text{Ru}(\text{brC})_2$, and best-fit curve to a 1:1 binding mode with an estimated K_D of 191 ± 39 nM. Average of three replicates.

Circular Dichroism (CD)

Measurements were made with a Jasco-715 coupled with a thermostat Nestlab RTE-111. The settings used were: Acquisition range: 300-195nm; band width: 2.0 nm; resolution: 0.2 nm; accumulation: 2 scans; sensitivity 10 mdeg; response time: 0.25 s, speed: 100 nm/min. CD measurements were made in a 2 mm cell at 20°C. Samples contained 5 μM of corresponding dsDNA (when present) and 5 μM of peptides in 10 mM phosphate buffer (pH 7.5) and 100 mM of NaCl. The CD spectra of the peptides (when measured in the presence of DNA) were calculated as the difference between the spectrum of the peptide/DNA mixture and the measured spectrum of a sample of the DNA oligonucleotide.

Control CD experiments

We have done comparative CD assays using $\text{Ru}(\text{brC})_2$, the disulphide dimer $(\text{brGGC})_2\text{SS}$ and the monomer $(\text{brGGC})\text{SH}$. Results are shown in figures S13 and S14. The observed helicity after the photolysis is consistent with the formation of monomeric peptides.

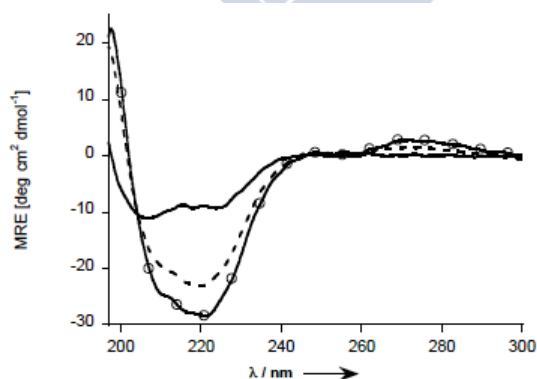


Figure S16. Circular dichroism analysis of DNA interaction of $\text{Ru}(\text{brC})_2$: Circular dichroism spectra of a 5 μM solution of $\text{Ru}(\text{brC})_2$ (solid line); same solution in the presence of 5 μM of ATF/CREB (white circles); same solution after the photolysis (dashed line).

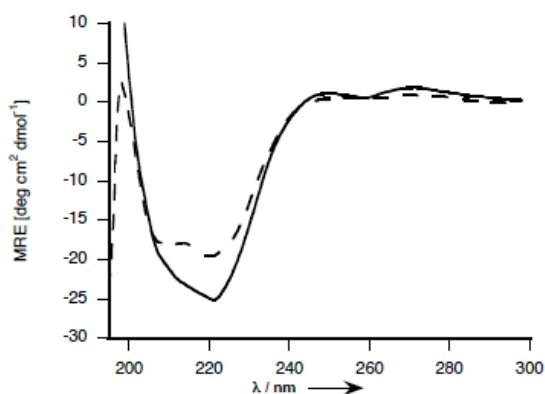


Figure S17. Circular dichroism analysis of DNA interaction of control peptides. Circular dichroism spectra of a 5 μM solution of $(\text{brGGC})_2\text{SS}$ (solid line) and $(\text{brGGC})\text{SH}$ (dashed line) with one equivalent of ATF/CREB.

EMSA experiments

EMSA was performed with a *BioRad Mini Protean* gel system, powered by an electrophoresis power supplies PowerPac Basic model, maximum power 150 V, frequency 50.60 Hz at 140 V (constant V). Binding reactions were performed in 18 mM Tris-HCl pH 7.5, 90 mM KCl, 1.8 mM MgCl_2 , 1.8 mM EDTA, 9% glycerol, 0.11 mg/mL BSA, 2.25% NP-40 for 30 min at 20°C. In the experiments we used 100 nM of the unlabeled dsoligonucleotides and a total incubation volume of 20 μL . The products were resolved by PAGE using a 10% non-denaturing polyacrylamide gel and 0.5X TBE buffer for 40 min at 20°C, and analyzed by staining with SyBrGold (Molecular Probes: 5 μL in 50 mL of 1X TBE) for 10 min and visualized by fluorescence.

5X TBE buffer: 0.445M Tris, 0.445 M Boric acid, 10 mM EDTA pH 8.0).

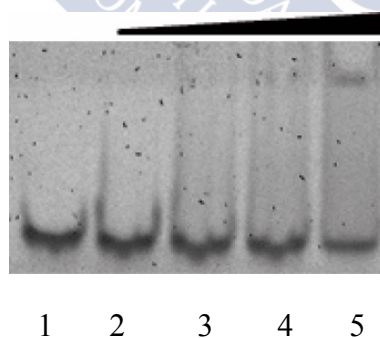


Figure S18. DNA binding of $\text{Ru}(\text{brC})_2$ with a non specific DNA, studied by EMSA: lanes 1-5: Random ds-DNA (5'-AGCGCACGTCCTGCACGTC-3', only one strand is shown; lanes 2-5: 200, 400, 600, 800 nM of $\text{Ru}(\text{brC})_2$.

Quantum yield calculation

The quantum yield of photolysis of the linker ($\Phi_{\text{unc}} \approx 0.56$) was obtained by comparing its photodegradation rate with that of $[\text{Ru}(\text{bpy})_2\text{PPh}_3\text{-GABA}]^+$. Isoabsorbant solutions of **1a** and $[\text{Ru}(\text{bpy})_2\text{PPh}_3\text{-GABA}]^+$ ($\Phi_{\text{unc}} \approx 0.2$)^[S10] were simultaneously irradiated and aliquots of both solutions were HPLC-analyzed after specific irradiation times. Their normalized peak areas were plotted against the total irradiation time, and adjusted to a monoexponential decay to compare the rates of photolysis.^[S11]

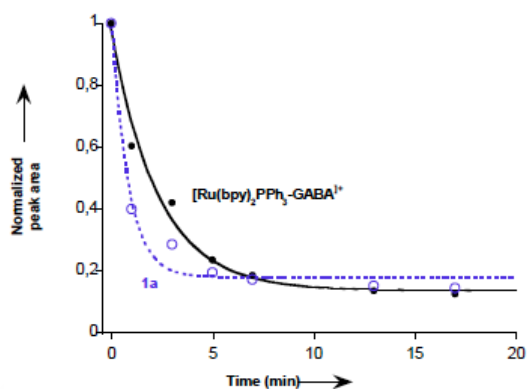


Figure S19. Relative uncaging kinetics of **1a** and $[\text{Ru}(\text{bpy})_2\text{P}(\text{Ph})_3\text{-GABA}]^+$. Best fit to a single exponential is shown.

UV spectra of 2

UV measurements were made in a Jasco V-630 spectrophotometer coupled to a Jasco ETC-717 temperature controller, using a standard Hellma semi-micro cuvette (140.002-QS). Measurements were made at 20 °C.

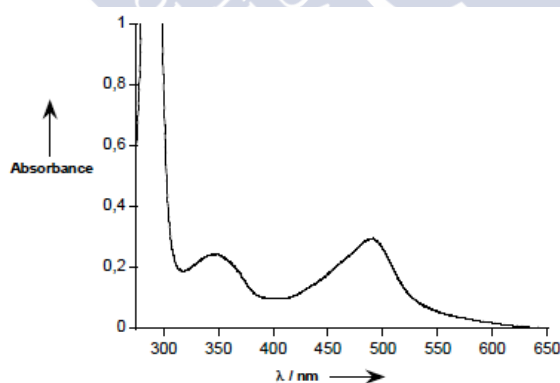
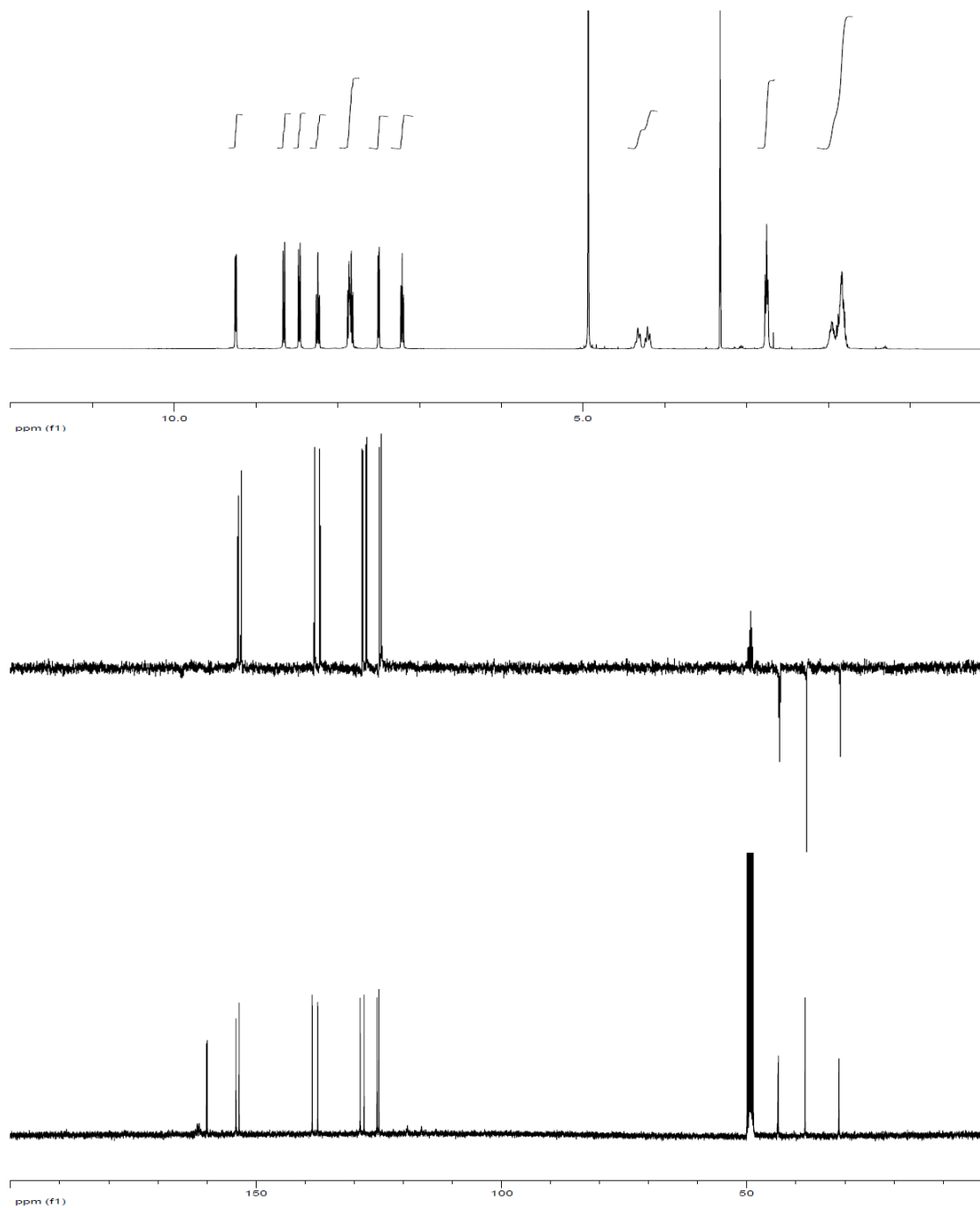
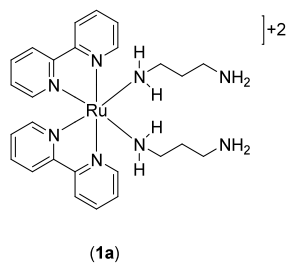


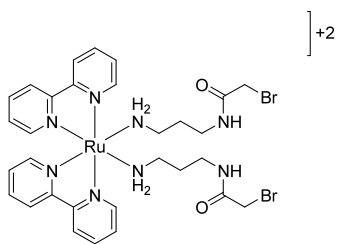
Figure S20. UV-Vis spectra of **2** (40 μM) recorded in aqueous solution.

NMR spectra

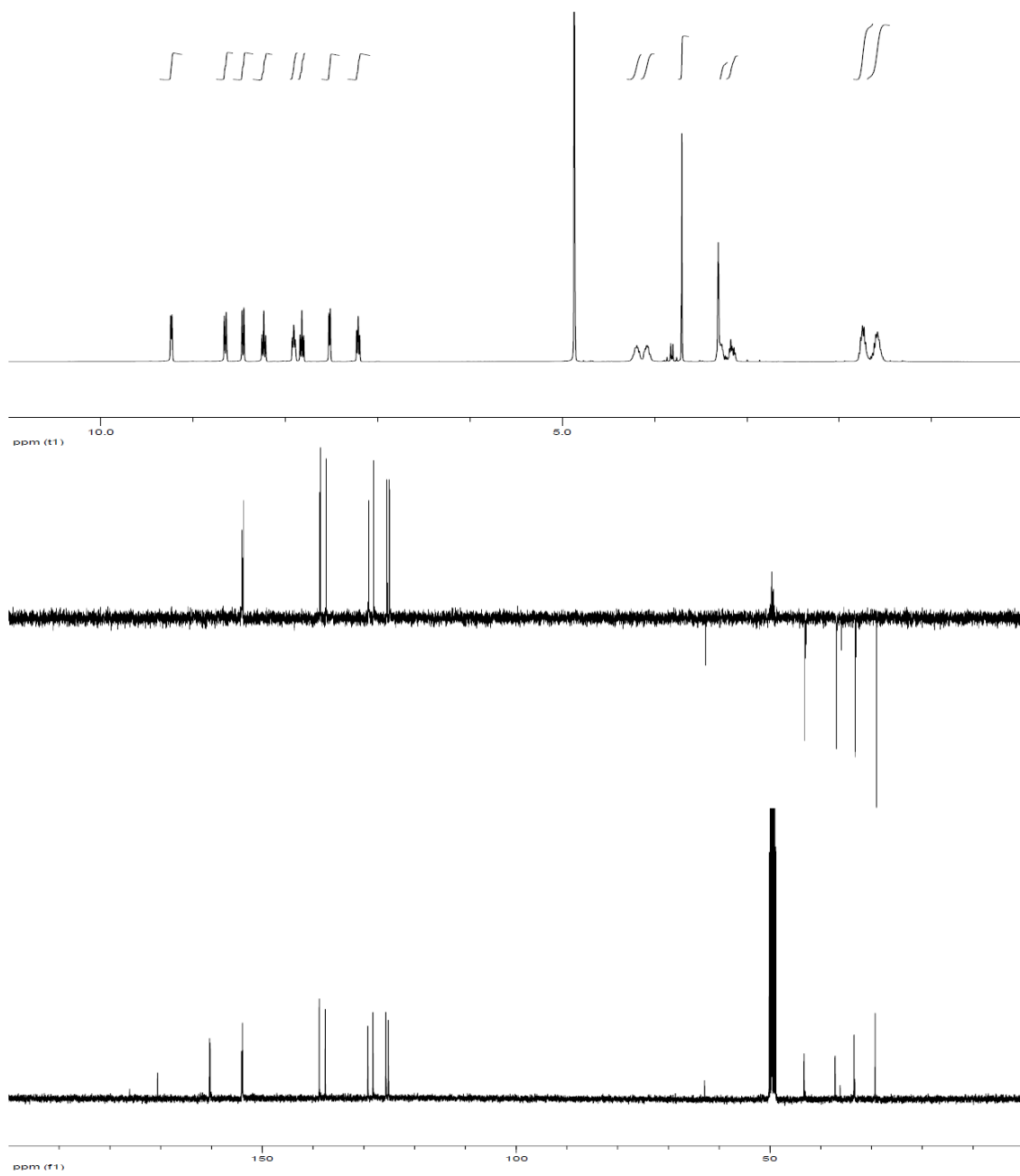
bis(2,2'-bipyridine) bis(N-(propane-1,3-diamine) ruthenium(II) (1a)



bis(2,2'-bipyridine) bis(N-(3-aminopropyl)-2-bromoacetamide) ruthenium(II) (2)



(2)



References

- [S1] a) K. C. Engman, P. Sandin, S. Osborne, T. Brown, M. Billeter, P. Lincoln, B. Nordén, B. Albinsson, L. M. Wilhelmsson. *Nucl. Acids Res.* **2004**, *17*, 5087-5095. b) G. Kallansrud, B. Ward. *Anal Biochem.* **1996**, *236*, 134-138.
- [S2] C. E. McCusker, J. K. McCusker, *Inorg. Chem.* **2011**, *50*, 1656-1669.
- [S3] All the ruthenium complexes were obtained as trifluoroacetic salts, as TFA is used for purifications.
- [S4] W.S. Hancock, J.E. Battersby *Anal. Biochem.* **1976**, *71*, 260–264.
- [S5] A. Jiménez-Balsa, E. Pazos, B. Martínez-Albardonedo, J. L. Mascareñas, M. E. Vázquez *Angew. Chem. In. Ed.* **2012**, *51*, 8825-8829.
- [S6] Zayat, L.; Salierno, M.; Etchenique, R. *Inorganic Chemistry* **2006**, *45*, 1728-1731.
- [S7] Schnaith, L. M. T.; Hanson, R. S.; Que, L. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 569-573.
- [S8] Abalea, V.; Cillard, J.; Dubos, M. P.; Anger, J. P.; Cillard P.; Morel, I. *Carcinogenesis* **1998**, *19*, 1053–1059.
- [S9] TAMRA is the acronym of the tetramethylrhodamine dye.
- [S10] L. Zayat, M. G. Noval, J. Campi, C. I. Calero, D. J. Calvo, R. Etchenique *ChemBioChem* **2007**, *8*, 2035–2038.
- [S11] a) Y. H. Huang, S. R. Sinha, O. D. Fedoryak, G. C. R. Ellis-Davies, D. E. Bergles *Biochemistry* **2005**, *44*, 3316–3326; b) Y. Zhao, Q. Zheng, K. Dakin, K. Xu, M. L. Martinez, W.-H. Li *J. Am. Chem. Soc.* **2004**, *126*, 4653–4663.

Parte experimental del Capítulo 3

Reversible Supramolecular Assembly at Specific DNA Sites: Nickel- Promoted Bivalent DNA Binding with Designed Peptide and Bipyridyl–Bis(benzamidine) Components

Mateo I. Sánchez, Jesús Mosquera, M. Eugenio Vázquez, José L. Mascareñas *Angew. Chem. Int. Ed.* **2014**, *53*, 9917.

Supporting material





Peptide synthesis

All peptide synthesis reagents and amino acid derivatives were from *GL Biochem* (Shanghai) and *Novabiochem*; amino acids were used as protected Fmoc amino acids with the standard side chain protecting scheme. C-terminal amide peptides were synthesized on a 0.1 mmol scale using a Fmoc-PAL-PEG-PS resin from *Applied Biosystems*. All solvents were dry and synthesis grade. Peptides were synthesized using an automatic *PS3* peptide synthesizer from *Protein Technologies*. Amino acids were coupled in 4-fold excess using HBTU/HOBt as activating agent. Each amino acid was incubated with the HBTU/HOBt mixture for 30 seconds in DMF before being added onto the resin. Peptide bond-forming couplings were conducted for 30 min to 45 min.

Deprotection of the temporal Fmoc protecting group: Removal of the temporal Fmoc groups was performed by treating the resin with 20% piperidine in DMF solution for 10 min.

N-terminal Aba (4-acetamidobenzoic acid) coupling: The resin containing the desired peptide was loaded with 4 equiv of HATU, another 4 equiv of 4-acetamidobenzoic acid and 6 equiv DIEA 0.2 M in DMF for 30 min. The resin was washed with DMF (2×5 mL, 2 min) and filtered.

Cleavage/deprotection step: The resin containing the desired peptide was washed with CH_2Cl_2 (2×5 mL, 2 min) and filtered. A mixture of CH_2Cl_2 (50 μL), H_2O (25 μL), triisopropylsilane (25 μL , TIS), and TFA (900 μL) was added to 40 mg of resin, and the mixture was shaken for 2.5 h. The resin was filtered and the liquid was added over 10 volumes of cold Et_2O . After 5 min the suspension was centrifuged and the resulting white residue was washed twice with Et_2O and dried under an Ar current.

Synthesis of peptides brHis₂, tbrHis₂, mbrHis₂

The peptides synthesis and deprotection step were carried out following standard conditions described before, the resulting crude mixtures were purified by RP-HPLC obtaining the desired peptides on yields between 30-50%.

Peptide brHis₂

Sequence: Aba–DPAAHKRAHNTEAARRSRARKLQR–NH₂.

ESI-MS $[\text{MH}]^+$ calcd. for $\text{C}_{123}\text{H}_{207}\text{N}_{52}\text{O}_{34} = 2958.3$, found = 2958.2. UV (H_2O) λ_{max} , nm (ϵ): 265 ($17,500 \text{ M}^{-1} \text{ cm}^{-1}$).

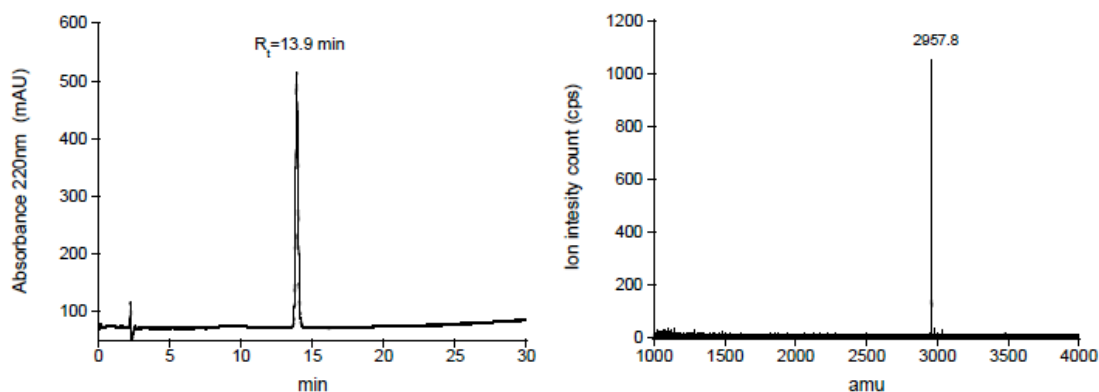


Figure S1. Left: HPLC of **brHis₂** after purification. Right: MALDI spectrum.

Peptide **tbrHis₂**

Sequence: *Aba*–AHKRAHNTEAARRSRARKLQR–*NH₂*.

ESI-MS $[MH]^+$ calcd. for $C_{111}H_{189}N_{48}O_{30}$ = 2674.5, found = 2674.6. UV (H_2O) λ_{max} , nm (ϵ): 265 ($17,500 M^{-1} cm^{-1}$).

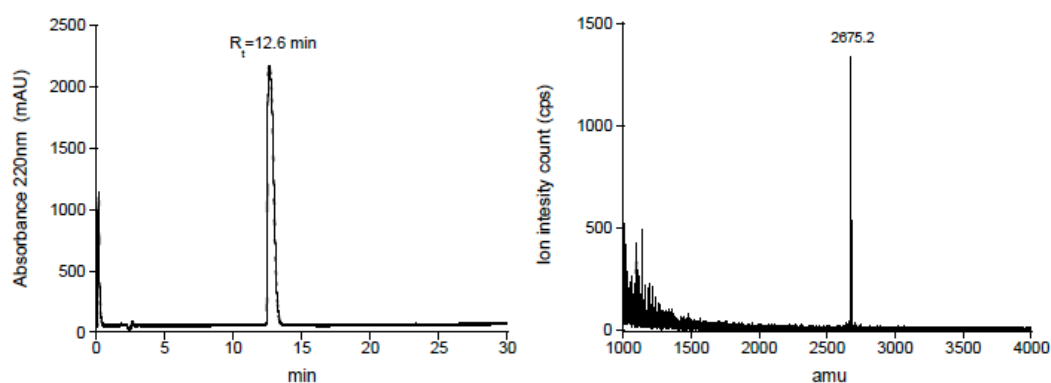


Figure S2. Left: HPLC of peptide **tbrHis₂** after purification (220 nm). Right: MALDI spectrum.

Peptide **mbrHis₂**

Sequence: *Aba*–DPAACKRAHNTEAARRSRARKLQR–*NH₂*.

ESI-MS $[MH]^+$ calcd. for $C_{120}H_{205}N_{50}O_{34}S$ = 2924.3, found = 2923.9. UV (H_2O) λ_{max} , nm (ϵ): 265 ($17,500 M^{-1} cm^{-1}$).

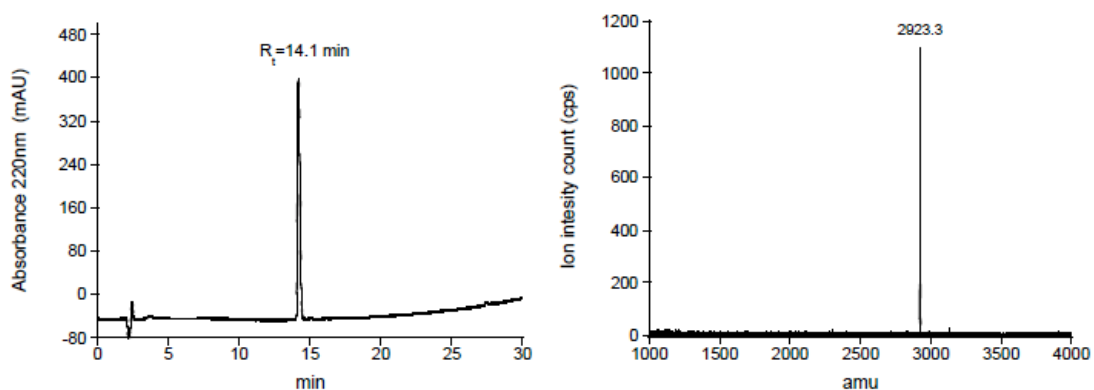


Figure S3. Left: HPLC of peptide **mbrHis₂** after purification (220 nm). Right: MALDI profile.

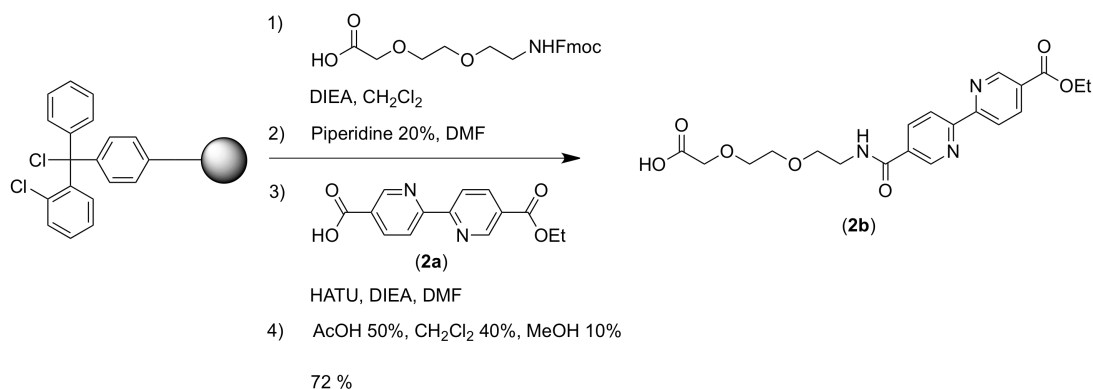
Synthesis of linker **2b** and bisbenzamidine derivatives **3a** and **3b**

All reagents were acquired from commercial sources: DMF and TFA were purchased from *Scharlau*, CH_2Cl_2 from *Panreac*, CH_3CN from *Merck*.

The formation of **3a** and **3b** was monitored by analytical RP-HPLC with an *Agilent 1100* series LC/MS using an *Eclipse XDBC18* ($4.6 \times 150 \text{ mm}$, $5 \mu\text{m}$) analytical column. Compounds were detected by UV absorption; standard conditions for analytical RP-HPLC consisted on a linear gradient from 5% to 95% of solvent B for 30 min at a flow rate of 1 mL/min (A: water with 0.1% TFA, B: acetonitrile with 0.1% TFA).

Final products were purified by RP-HPLC with an *Agilent 1100* series equipped with a binary pump system and a UV-visible detector using a *Phenomenex Luna C18 100A* ($250 \times 10 \text{ mm}$, $5 \mu\text{m}$) semipreparative column. The fractions containing the products were freeze-dried, and their identity confirmed by ESI^+ -MS and NMR. Compounds were isolated as TFA salts after HPLC purification.

Bipyridine **2a** was kindly provided by courtesy of Dr. Gustavo Rama^[S1] and was synthesized according to methods previously described.^[S2]

Synthesis of 2b

8-(9-Fluorenylmethyloxycarbonyl-amino)-3,6-dioxaoctanoic acid (Fmoc-O₂Oc-OH) (385 mg, 1.0 mmol) and DIEA (516 mg, 4.0 mmol, 4 eq) was added over 2-Chlorotrityl resin (0.25 mmol) in 2.5 mL of CH₂Cl₂ for 50 min after that, the resin was washed with CH₂Cl₂/MeOH/DIEA (17:2:1) (3 x 5 mL, 2 min), CH₂Cl₂ (3 x 5 mL, 2 min), finally, the Fmoc was removed by treating the resin with 20% piperidine in DMF (1 x 5 mL, 10 min).

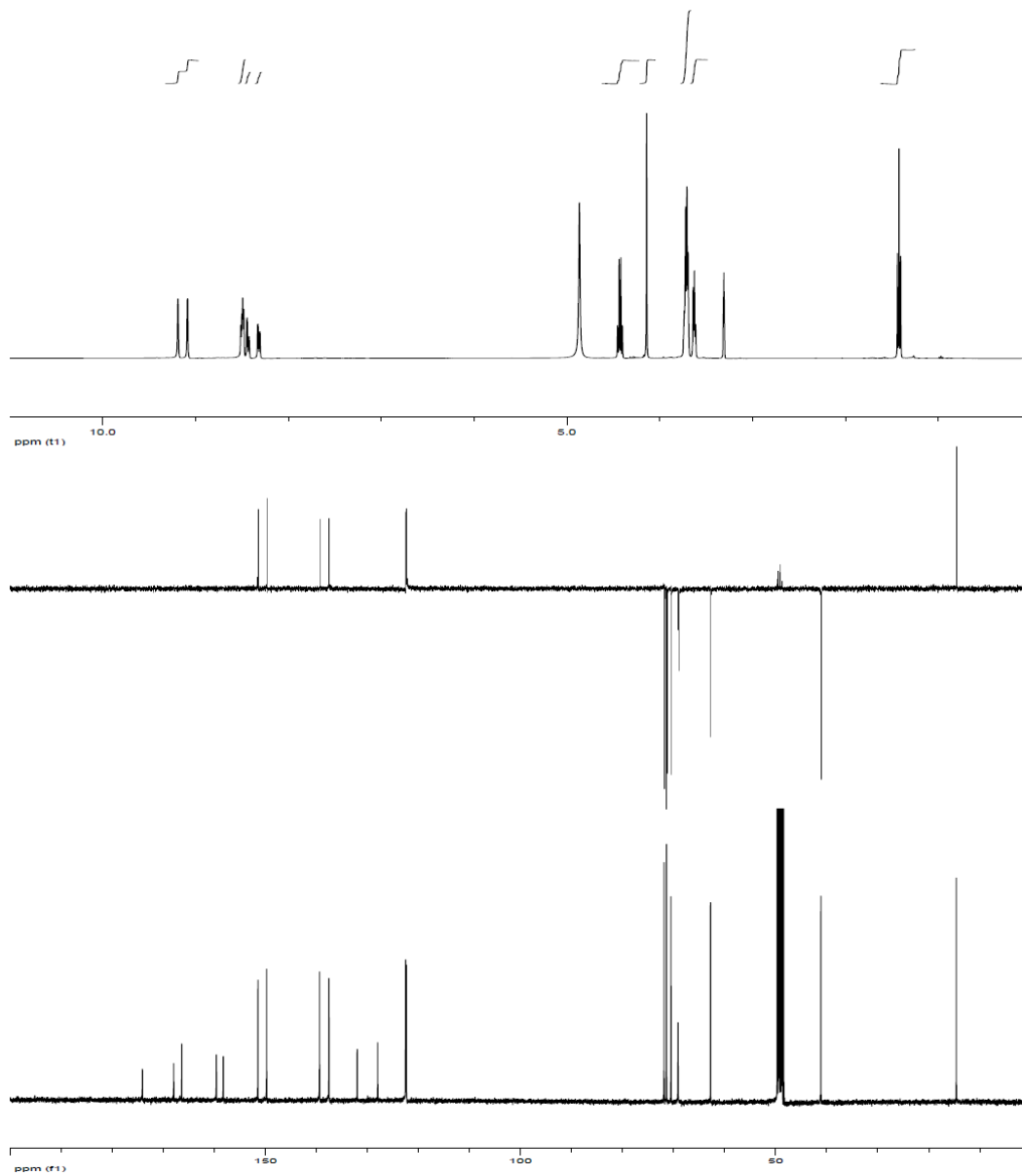
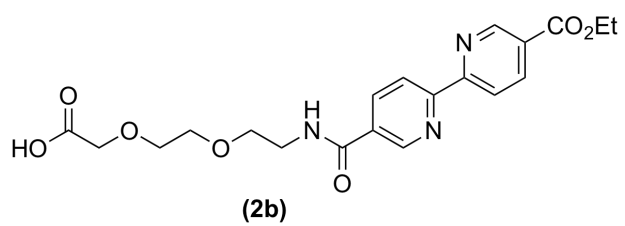
A solution of bipyridine 2a (204 mg, 0.75 mmol), DIEA/DMF 0.2 M (5.6 mL, 1.5 mmol) and HATU (285 mg, 0.75 mmol) in 3.5 mL of DMF was added over the unprotected resin for 50 min. The resin was washed with DMF (3 x 5 mL, 2 min) and CH₂Cl₂ (2 x 5 mL, 2 min) and then was treated with the deprotection cocktail AcOH/ CH₂Cl₂/MeOH (5/4/1) (400 mg of resin / 10 mL of cocktail); after filtration the crude was concentrated under reduced pressure.

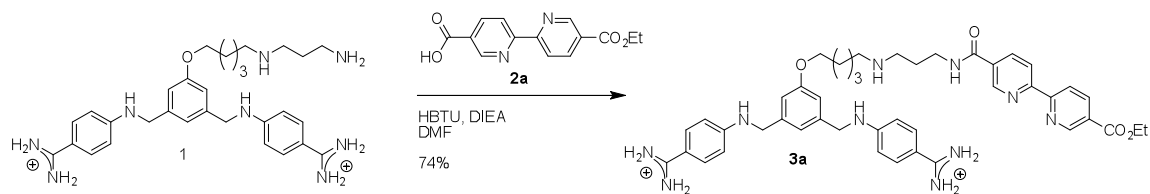
The residue after concentration was purified by HPLC (gradient: 15% B, 5 min; 15%→95 % B, 30 min). The appropriated fractions were collected, concentrated and freeze-dried to obtain the desired product as pale white powder (74 mg, 0.18 mol, 72 %).

¹H NMR (400 MHz, MeOD *d*-₄ δ): 1.42 (t, *J* = 7.12 Hz, 3H), 3.63 (t, *J* = 5.20 Hz, 2H), 3.69-3.74 (m, 6H), 4.14 (s, 2H), 4.43 (q, *J* = 7.12 Hz, 2H), 8.32 (dd, *J* = 8.27, 1.78 Hz, 1H), 8.43 (d, *J* = 8.33 Hz, 1H), 8.48-8.51 (m, 2H), 9.08 (s, 1H), 9.18 (s, 1H).

¹³C NMR (MeOD *d*-₄ δ): 14.6 (CH₃), 41.1 (CH₂), 62.7 (CH₂), 69.1 (CH₂), 70.5 (CH₂), 71.4 (CH₂), 71.8 (CH₂), 122.3 (CH), 122.5 (CH), 127.9 (C), 131.9 (C), 137.5 (CH), 139.3 (CH), 149.7 (CH), 151.4 (CH), 158.2 (C), 159.5 (C), 166.3 (C), 167.9 (C), 174.0 (C).

ESI-MS: [M+H]⁺ calc. for C₂₀H₂₄N₃O₇ 418.1609 found 418.1605 C₂₀H₂₃N₃O₇ (M.W. 417.1536).



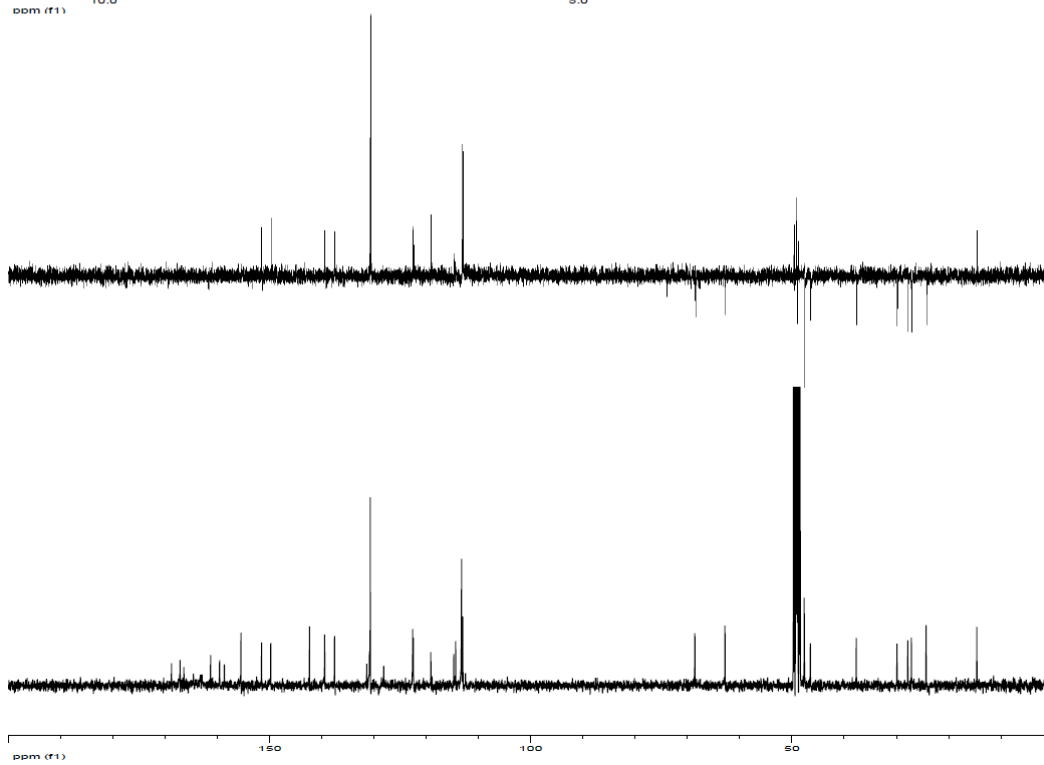
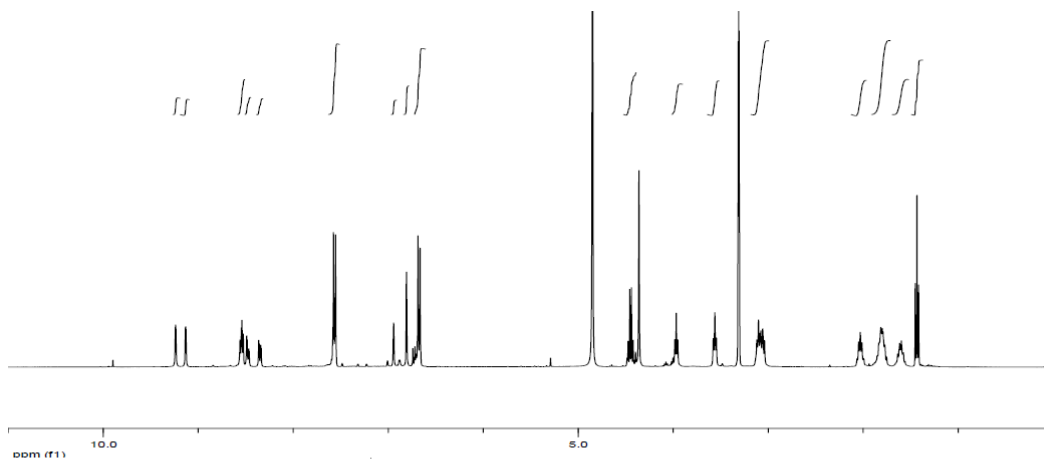
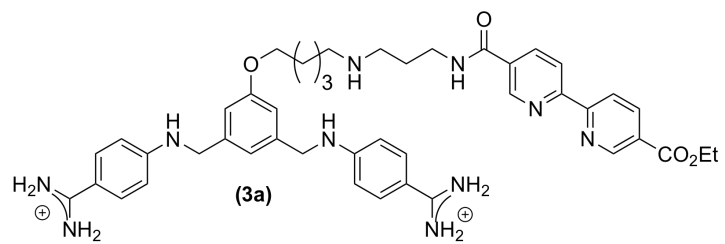
Synthesis of 3a

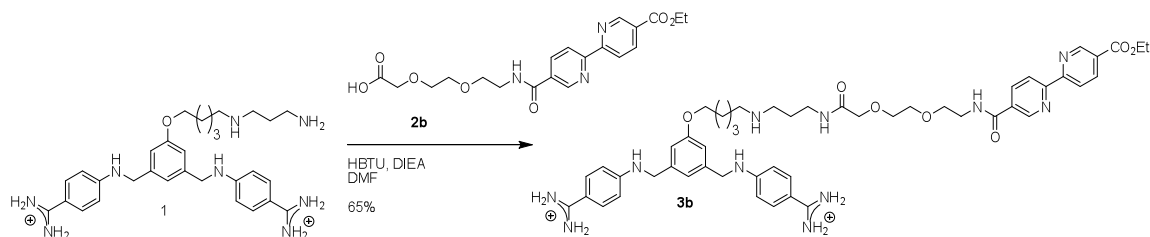
A fresh mixture of **2a** (5.8 mg, 14 μmol) and HBTU (6.0 mg, 15.2 μmol) in DIEA/DMF 0.195 M (270 μL) was stirred for 5 min at room temperature. To the mixture was added a solution of the benzamidine **1**^[S3] (15 mg, 17.5 μmol) in DIEA/DMF (132 μL). The resulting solution was stirred at room temperature until HPLC analysis showed the disappearance of the starting material, (1 h). The residue after concentration was purified by HPLC (gradient: 15% B, 5 min; 15% \rightarrow 95 % B, 30 min). The appropriated fractions were collected, concentrated and freeze-dried to obtain the desired conjugate **3a** as pale white powder TFA salt (11.7 mg, 10.4 μmol , 74 %).

¹H NMR (400 MHz, MeOD *d*-₄ δ): 1.43 (t, $J = 7.1$ Hz, 3H), 1.60 (td, $J = 14.9, 7.4$ Hz, 2H), 1.75-1.85 (m, 4H), 2.03 (q, $J = 6.6$ Hz, 2H), 3.08 (td, $J = 15.5, 7.4$ Hz, 4H), 3.56 (t, $J = 6.4$ Hz, 2H), 3.97 (t, $J = 6.1$ Hz, 2H), 4.36 (s, 4H), 4.45 (q, $J = 7.1$ Hz, 2H), 6.67 (d, $J = 8.9$ Hz, 4H), 6.80 (s, 2H), 6.94 (s, 1H), 7.56 (d, $J = 8.9$ Hz, 4H), 8.35 (dd, $J = 8.3, 2.0$ Hz, 1H), 8.48 (dd, $J = 8.3, 1.9$ Hz, 1H), 8.54 (dd, $J = 8.0, 5.7$ Hz, 2H), 9.13 (s, 1H), 9.23 (s, 1H).

¹³C NMR (MeOD *d*-₄ δ): 14.6 (CH₃), 24.2 (CH₂), 27.1 (CH₂), 27.8 (CH₂), 29.8 (CH₂), 37.7 (CH₂), 46.4 (CH₂), 47.6 (CH₂), 49.0 (CH₂), 62.8 (CH₂), 68.5 (CH₂), 112.9 (C), 113.2 (C), 114.3 (C), 114.7 (C), 119.1 (C), 122.5 (C), 128.1 (C), 130.7 (C), 131.3 (C), 137.5 (C), 139.4 (C), 142.3 (C), 149.7 (C), 151.5 (C), 155.4 (C), 158.6 (C, TFA), 159.5 (C), 161.2 (C), 166.3 (C), 167.1 (C), 168.7 (C).

ESI-MS: $[\text{M}+\text{H}]^+$ calc. for C₄₄H₅₃N₁₀O₄ 785.4246 found 785.4238. C₅₀H₅₅F₉N₁₀O₁₀ (M.W. 1127.0183).

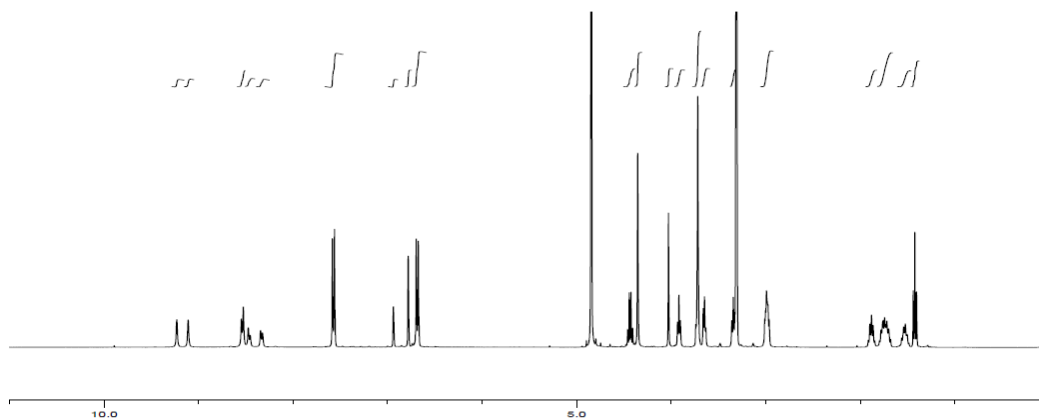
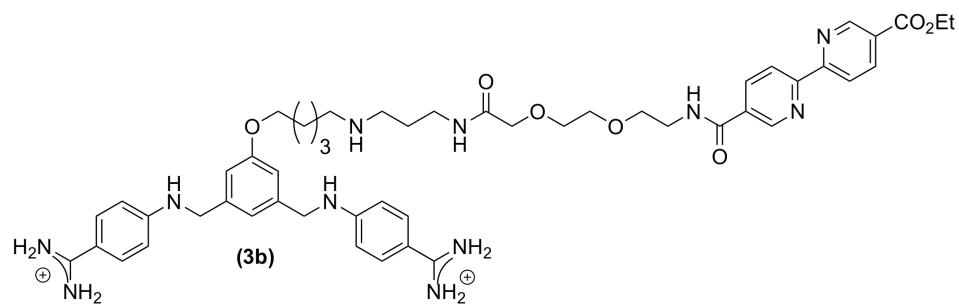


Synthesis of 3b

A mixture of **2b** (3.8 mg, 9.2 μmol) and HBTU (3.9 mg, 10.0 μmol) in DIEA/DMF 0.195M (176 μL) was stirred for 5 min, and to the mixture was added a solution of the benzamidinium **1** (10 mg, 11.5 μmol) in DIEA/DMF (87 μL). The resulting mixture was incubated at room temperature until HPLC analysis showed the disappearance of the starting material (1 h). The residue after removal of the solvents under vacuum was purified by HPLC (gradient: 15% B, 5 min; 15% \rightarrow 95 % B, 30 min). The appropriated fractions were collected, concentrated and freeze-dried to obtain the desired conjugate **3b** as pale white powder TFA salt (7.6 mg, 6.0 μmol , 65 %).

¹H NMR (400 MHz, MeOD *d*-₄ δ): 1.42 (t, $J = 7.12$ Hz, 3H), 1.48-1.56 (m, 2H), 1.68-1.80 (m, 4H), 1.88 (q, $J = 6.85$ Hz, 2H), 2.93-3.01 (m, 2H), 3.34 (t, $J = 6.60$ Hz, 2H), 3.65 (t, $J = 5.35$ Hz, 2H), 3.71 (s, 6H), 3.92 (t, $J = 6.11$ Hz, 2H), 4.02 (s, 2H), 4.35 (s, 4H), 4.43 (q, $J = 7.12$ Hz, 2H), 6.69 (d, $J = 8.86$ Hz, 4H), 6.78 (s, 2H), 6.93 (s, 1H), 7.57 (d, $J = 8.87$ Hz, 4H), 8.33 (d, $J = 8.00$ Hz, 1H), 8.46 (d, $J = 8.06$ Hz, 1H), 8.54 (d, $J = 8.21$ Hz, 2H), 9.11 (s, 1H), 9.23 (s, 1H).

ESI-MS: $[\text{M}+\text{H}]^+$ calc. for C₅₀H₆₄N₁₁O₇ 930.4985 found 930.4990. C₅₆H₆₆F₉N₁₁O₁₃ (M.W. 1272.1748).



Oligonucleotide sequences

Double strand oligonucleotides (only one strand is shown) were supplied by *Thermo Fischer* and their sequences were:

API^{hs}•AT 5'-ACGAACG TCATAATTT CCTC-3'

API^{hs}•sAT 5'-ACGAACG TCATAATT CCTC-3'

mAPI^{hs}•AT 5'-ACGAACG TCGTAATTT CCTC-3'

API^{hs}•mAT 5'-ACGAACG TCATAGTTT CCTC-3'.

Fluorescence spectroscopy

Measurements were made with a Jobin-Yvon *Fluoromax-3*, (DataMax 2.20) coupled to a Wavelength Electronics LFI-3751 temperature controller, using the following settings: increment: 1.0 nm; integration time: 0.2 s; excitation slit width: 3.0 nm; emission slit width: 6.0 nm at 20 °C. Excitation wavelength 329 nm and the emission spectra were acquired from 345 to 500 nm.

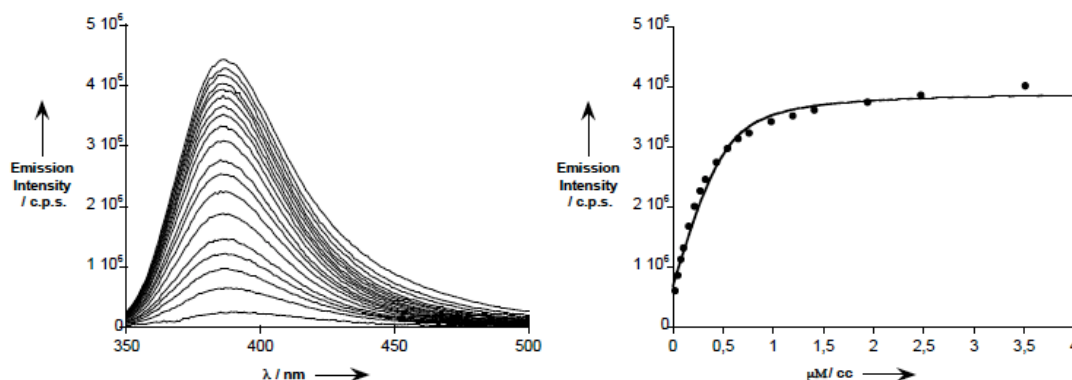


Figure S4: Left: fluorescence emission spectra of **3a** in the presence of increasing concentrations of **API^{hs}•AT**. Right: Plot of the fluorescence emission at 389 nm against the total ds-DNA concentration and best-fit curve to a 1:1 binding mode. Estimated $k_d = 33.1 \pm 2.1$ nM.

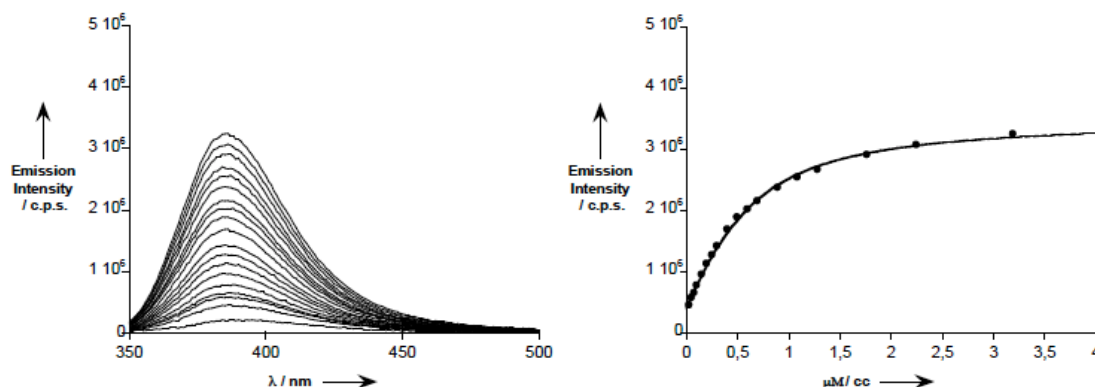


Figure S5: Left: fluorescence emission spectra of **3a** in the presence of increasing concentrations of **API^{hs}•sAT**. Right: Plot of the fluorescence emission at 389 nm against the total ds-DNA concentration and best-fit curve to a 1:1 binding mode. Estimated $k_d = 167.3 \pm 27.8$ nM.

EMSA controls

EMSA was performed with a *BioRad Mini Protean* gel system, powered by an electrophoresis power supplies PowerPac Basic model, maximum power 150 V, frequency 50.60 Hz at 140 V (constant V).

The protocol for fluorescence EMSAs involved addition of peptide solutions in 20 mM Tris-HCl pH 7.5 to 18 mM Tris-HCl pH 7.5, 90 mM KCl, 1.8 mM MgCl₂, 9% glycerol 0.11 mgmL⁻¹ BSA, 2.25% NP-40 and 100 nM of DNA (4 °C, 30 min) and loading into the gel. Products were resolved by PAGE on a 10% nondenaturing polyacrylamide gel and 0.5X TBE buffer over 40 min at 4 °C. and analyzed if correspond by staining with SyBrGold (Molecular Probes: 5 μL in 50 mL of 1x TBE) for 10 min, followed by fluorescence visualization.

5X TBE buffer without EDTA (0.445M Tris, 0.445 M Boric acid).

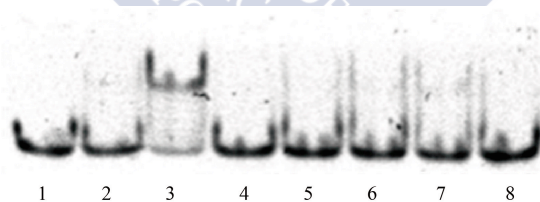


Figure S6. EMSA analysis of the DNA binding recognition properties of **brHis₂** with different metals. Lanes 1-8: **API^{hs}/T•A** (100 nM); lanes 2-8: 1000 nM of **brHis₂**, and **3a**. Lanes 2: 5 equiv of Co(ClO₄)₂, lane 3: 5 equiv of Ni(ClO₄)₂; lane 4: 5 equiv of CdCl₂; lane 5: 5 equiv of ZnSO₄; lane 6: 5 equiv of CuSO₄; lane 7: 5 eq of HgCl₂; lane 8: 5 equiv of Fe(NH₄)₂(SO₄)₂.

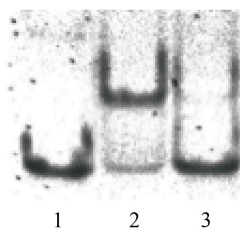


Figure S7. EMSA analysis of the DNA binding recognition properties of **tbrHis₂**. Lanes 1-3: **AP1^{hs}/T·A** (100 nM); lanes 2: 1000 nM of **brHis₂, 3a** and 5 equiv of Ni(ClO₄)₂; lane 3: 1000 nM of **tbrHis₂, 3a** and 5 equiv of Ni(ClO₄)₂.

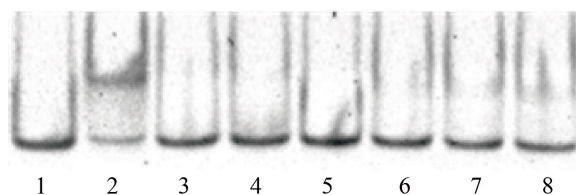


Figure S8. EMSA analysis of the DNA binding recognition properties of **mbrHis₂** with different metals. Lanes 1-8: **AP1^{hs}/T·A** (100 nM); lane 2: 1000 nM of **brHis₂, 3a** and 5 equiv of Ni(ClO₄)₂; Lanes 3-8 : 1000 nM of **mbrHis₂**, and **3a**; lane 3: 5 equiv of Ni(ClO₄)₂, lane 4: 5 equiv of Fe(NH₄)₂(SO₄)₂; lane 5: 5 equiv of ZnSO₄; lane 6: 5 equiv of CuSO₄; lane 7: 5 equiv of Co(ClO₄)₂; lane 8: 5 equiv of HgCl₂.

References

- [S1] Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CIQUS) and Departamento de Química Inorgánica. Universidade de Santiago de Compostela. 15782 Santiago de Compostela, Spain.
- [S2] H. Ishida, Y. Maruyama, M. Kyakuno, Y. Kodera, T. Maeda, S. Oishi, *ChemBioChem* **2006**, *7*, 1567-1570.
- [S3] M. I. Sánchez, O. Vázquez, M. E. Vázquez, J. L. Mascareñas, *Chemistry – A European Journal* **2013**, *19*, 9923-9929.

Parte experimental del Capítulo 4

Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides

Jesús Mosquera, Adrián Jiménez-Balsa, Verónica I. Dodero, M. Eugenio Vázquez, José L. Mascareñas *Nature Communications* **2013**, *4*, 1874.

Supporting material





General

All reagents were acquired from commercial sources: DMF and TFA were purchased from *Scharlau*, CH₂Cl₂ from *Panreac*, CH₃CN from *Merck*, [2,2':6',2''-terpyridin]-4'-ylmethanol was purchased from *HetCat*. The rest of reagents were acquired from *Sigma-Aldrich*. All peptide synthesis reagents and amino acid derivatives were purchased from *GL Biochem (Shanghai) Ltd.* and *Novabiochem*. Amino acids were purchased as protected Fmoc amino acids with the standard side chain protecting scheme, except for the orthogonally protected Fmoc-Lys(Alloc)-OH and Fmoc-Lys(Mtt)-OH.

Reactions were followed by analytical RP-HPLC with an *Agilent 1100* series LC/MS using an *Eclipse XDB-C8* analytical column (4.6 x 150 mm, 5 μm). Electrospray Ionization Mass Spectrometry (ESI/MS) was performed with an *Agilent 1100* Series LC/MSD model in positive scan mode using direct injection of the purified peptide solution into the MS. Standard conditions for analytical RP-HPLC consisted on an isocratic regime during the first 5 min, followed by a linear gradient from 5% to 75% of solvent B for 30 min at a flow rate of 1 mL/min (A: water with 0.1% TFA, B: acetonitrile with 0.1% TFA). Compounds were detected by UV absorption at 220, 270, 304 and 330 nm. Purification were performed by semipreparative RP-HPLC with an *Agilent 1100* series LC using a *Luna 5u C18(2) 100A* (5 μm, 10 × 250 mm) reverse-phase column from *Phenomenex*. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI/MS) was performed with a *Bruker Autoflex MALDI/TOF* model in positive scan mode by direct irradiation of the matrix-absorbed peptide. Concentrations were measured using the listed extinction coefficients.

Reverse phase purifications of the terpyridines was performed with a Büchi Sepacore preparative system consisting on a C-615 pump manager with two C-605 pump modules for binary solvent gradients, a C-660 fraction collector and a C-635 UV Photometer. Reverse phase conditions consisted on 5 min isocratic 5% B, followed by linear gradient from 5 to 75% B for 40 min, with a flow rate of 30 mL/min using a pre-packed preparative cartridge (150 × 40 mm) with reverse phase RP18ec silica gel (Büchi order number 54863).

Calf thymus DNA was acquired from *Sigma-Aldrich*. Oligonucleotides were purchased from *Thermo Fisher Scientific GmbH* on a 0.2 mmol scale as freeze-dried solids. After solving in H₂O *milliQ* their concentrations were measured by UV absorption at 260 nm with a *BioRad SmartSpec Plus Spectrophotometer*. Absorbance was measured twice and concentrations were calculated applying Lambert-Beer's equation. The molar extinction coefficients of single strand oligonucleotides were calculated by using the following formula,^[S1]

$$\epsilon_{(260\text{ nm})} = \{(8.8 \times \#T) + (7.3 \times \#C) + (11.7 \times \#G) + (15.4 \times \#A)\} \times 0.9 \cdot 10^3 \text{ M}^{-1} \text{ cm}^{-1}$$

Where #A, #T, #C, #G stand for the number of each type of bases in the DNA strand. The concentration of Rhodamine-labeled oligonucleotides was calculated using the molar extinction of rhodamine. Oligonucleotides were hybridized by mixing complementary sequences at equal molar concentration (50 μM), heating at 90 °C for 10 min and then slowly cooling the mixture to rt over 1h.

EMSA experiments were carried as out described in the main manuscript (methods section). The labeling of oligonucleotides with ³²P was accomplished in the following manner: 1 μL of the single strand oligonucleotide which will be labeled (10 μM), 1 μL kinase buffer 10X (40 mM, Tris HCl pH 7.5, 10 mM MgCl₂, 5 mM DTT), 2 μL ATP* (32 γ-ATP [5000

Ci/mmol]) and 0.5 μ L Ty kinase (10 units / μ L); after shaking the mixture is heated for 1 h at 37 $^{\circ}$ C, and then quenched by heating for 10 min at 95-100 $^{\circ}$ C.

DNA annealing is carried out by incubating the selected oligonucleotide with another solution containing a complementary oligonucleotide; the mixture is heated for 10 min at 90 $^{\circ}$ C and left to reach rt again.

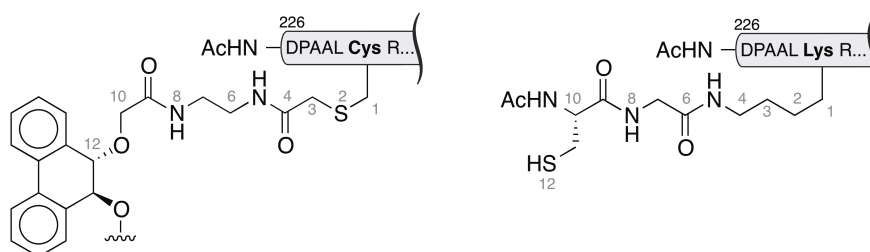


Figure S1. Comparison between the Morii's linker and our linker for N-terminal covalent dimers of GCN4 basic regions. Left: Linker used by Morii et. al. in their studies of N-terminal covalent dimers of GCN4 basic regions. Right: Linker used in this work. Both linkers are geometrically equivalents.

Peptide Synthesis

C-terminal amide peptides were synthesized following standard peptide protocols (Fmoc/tBu strategy) on a 0.1 mmol scale using a 0.19 mmol/g loading Fmoc-PAL-PEG-PS resin from Applied Biosystems, using a PS3 automatic peptide synthesizer from Protein Technologies. The amino acids were coupled in 4-fold excess using HBTU as activating agent. Each amino acid was activated for 30 seconds in DMF before being added onto the resin, and couplings were conducted for 30 min. Deprotection of the temporal Fmoc protecting group was performed by treating the resin with 20% piperidine in DMF for 10 min. Manual couplings were monitored using the TNBS (trinitrobenzene sulfonate) test.^[S2]

The resin cleavage-deprotection was accomplished by shaking the resin-bound peptides for 2 h (approx. 0.025 mmol), in 3 mL of the cleavage cocktail (25 μ L of H₂O, 25 μ L of TIS (triisopropylsilane), 50 μ L of DCM and 900 μ L of TFA). The resin is filtered, and the TFA filtrate is added to ice-cold diethyl ether (30 mL). After 10 min, the precipitate is centrifuged and washed again with 20 mL of ice-cold ether. The solid residue is dried under argon, dissolved in acetonitrile/water 1:1 (2 mL) and purified by semipreparative RP-HPLC. The collected fractions are lyophilized and stored at -20 $^{\circ}$ C.

Deprotection of Lys(Alloc) side chain. The resin-bound peptide (approx 0.05 mmol) was treated overnight at rt with a deoxygenated mixture of methylmorpholine (55 μ L, 10 eq.), triphenylphosphine (20 mg, 1.5 equiv), phenylsilane (62 μ L, 10 equiv) and Pd(OAc)₂ (3.4 mg, 0.3 equiv) in CH₂Cl₂ (2.7 mL). The resin was filtered and washed with DMF (2 \times 1.5 mL \times 2 min), diethyldithiocarbamate (DEDTC, 25 mg in 5 mL of DMF, 2 \times 1.5 mL \times 5 min), DMF (2 \times 1.5 mL \times 2 min) and CH₂Cl₂ (2 \times 5 mL \times 2 min).

Deprotection of Lys(Mtt) sidechain. The resin-bound peptide (approx. 0.05 mmol) was treated twice with a mixture of TFA (250 μ L), triisopropylsilane (50 μ L) and CH₂Cl₂ (4.7 ml) for 5 min. The resin was then filtered and washed with CH₂Cl₂ (2 \times 5 mL \times 2 min).

Synthesis of the peptide 1

The peptide was elongated by the previously described SPPS procedures, and the lysine at position 231 was inserted as Lys(alloc). Once introduced the acetyl group at the N terminus, the protecting alloc group was removed (see Deprotection of Lys(Alloc) side chain) and the glycine and Cys(tButhio) were coupled using SPPS standard protocols. Finally, the N-terminus was acetylated, treated with cleavage cocktail and purified.

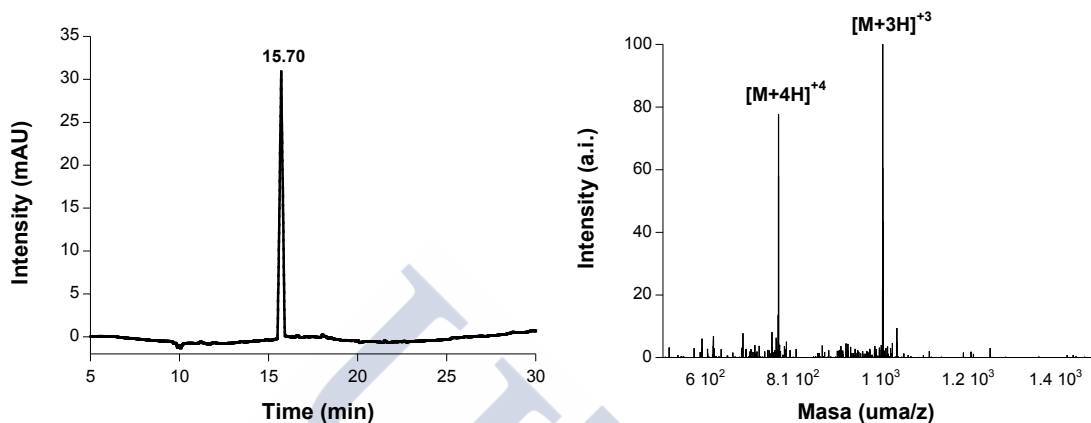
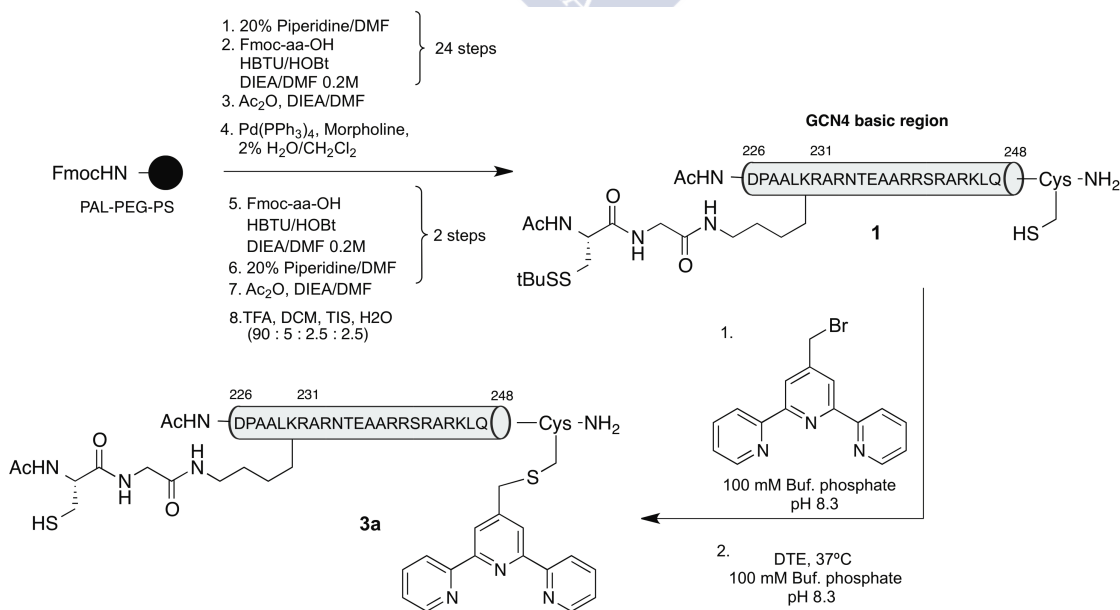


Figure S2. Left: HPLC chromatogram of the purified peptide **1** (222 nm). Right: ESI⁺-MS of peptide **1**. Peptide **1**: ESI-MS [MH]⁺ calcd. for C₁₂₄H₂₂₂N₄₉O₃₆S₃ = 3071.6, found = 3071.3. UV (H₂O) λ_{max}, nm: 222.

Synthesis of the peptide 3a

The purified peptide **1** (8 mg, approx. 0.002 mmol) was dissolved in a deoxygenated 100 mM phosphate buffer, pH 8.3 (800 μL). 4'-(bromomethyl)-2,2':6',2''-terpyridine (2.6 mg, 0.008 mmol) was dissolved in deoxygenated acetonitrile (800 μL) and added to the peptide solution. The mixture was stirred for 30 min and then dithioerythritol (DTE, 6.2 mg, 0.04 mmol) dissolved in buffer phosphate pH 8.3 (200 μL) were added. The reaction was stirred for 2 h at 37 °C, and the resulting crude purified by RP-HPLC. After lyophilization the product was obtained as white powders on yields between 60-50%.



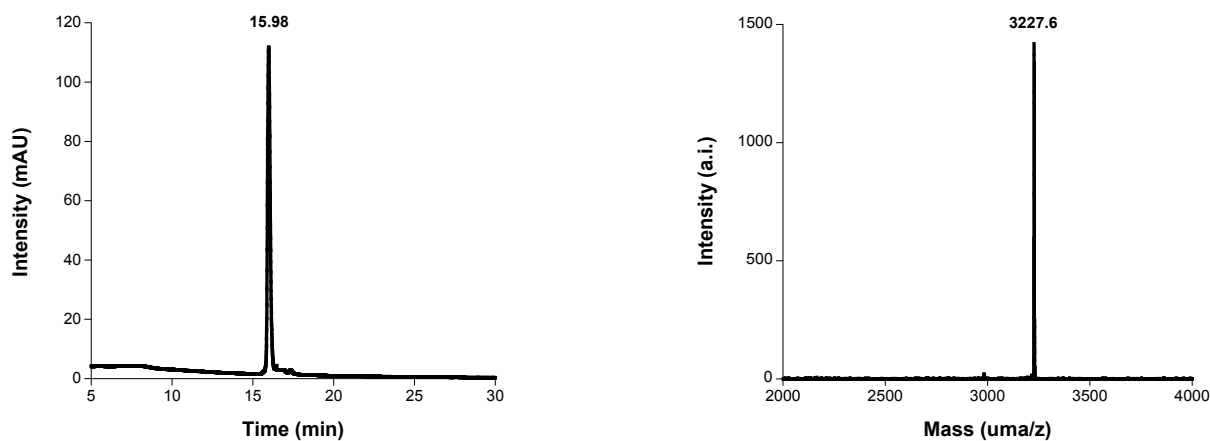


Figure S3. Left: HPLC chromatogram of the purified peptide **3a**. Right: MALDI of peptide **3a**. Peptide **3a**: ESI-MS $[MH]^+$ calcd. for $C_{136}H_{225}N_{52}O_{36}S_2 = 3228.7$, found = 3228.8. UV (H_2O) λ_{max} , nm (ϵ): 280 ($17,800 M^{-1} cm^{-1}$).

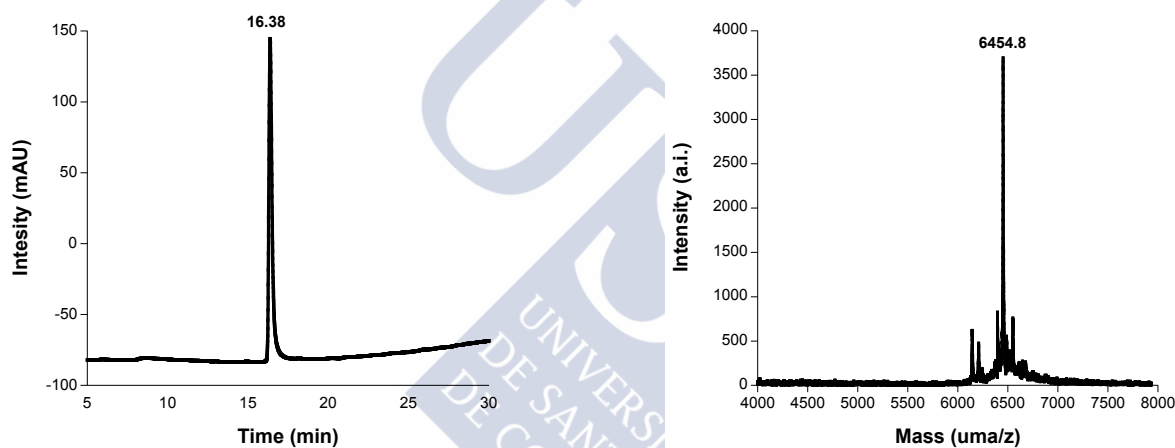


Figure S4. Left: HPLC chromatogram of the purified dimer **(3a)₂SS**. Right: MALDI of **(3a)₂SS**. Dimer **(3a)₂SS**: ESI-MS $[MH]^+$ calcd. for $C_{272}H_{447}N_{104}O_{72}S_4 = 6454.4$, found = 6455.0. UV (H_2O) λ_{max} , nm (ϵ): 280 ($35,600 M^{-1} cm^{-1}$).

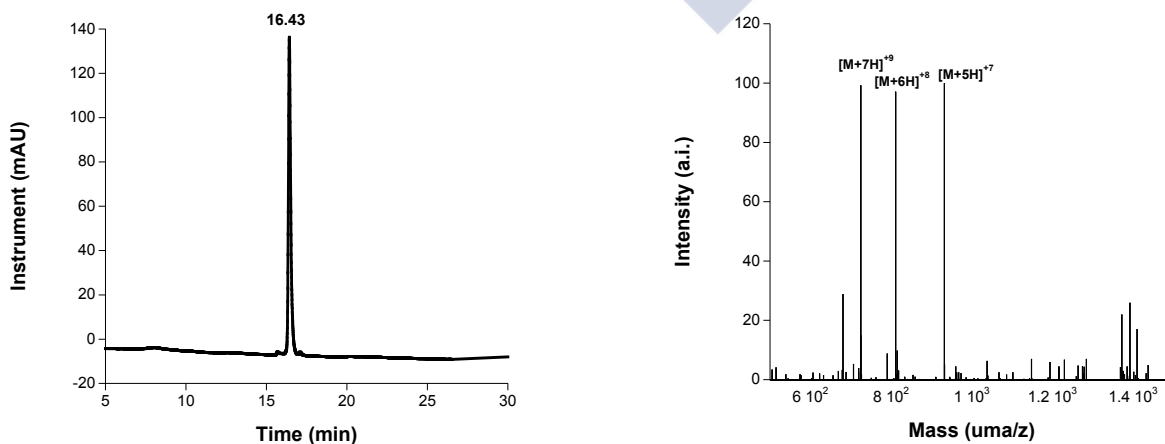


Figure S5. Left: HPLC chromatogram of the purified dimer **(3a)₂Ni**. Right: ESI⁺-MS of **(3a)₂Ni**. Dimer **(3a)₂Ni**: ESI-MS $[M-H]^+$ calcd. for $C_{272}H_{447}N_{104}NiO_{72}S_4 = 6514.4$, found = 6515.0.

Synthesis of the peptide 3b

The purified peptide **1** (4 mg, approx. 0.001 mmol) was dissolved in a deoxygenated 100 mM phosphate buffer, pH 8.3 (400 μ L). 2-Bromo-4'-nitroacetophenone (1 mg, 0.004 mmol) was dissolved in deoxygenated acetonitrile (400 μ L) and was added to peptide solution. The mixture was stirred for 30 min and then dithioerythritol (DTE, 3.1 mg, 0.02 mmol) dissolved in buffer phosphate pH 8.3 (100 μ L) was added. The reaction was stirred for 2 h at 37 $^{\circ}$ C, and the resulting crude purified by RP-HPLC. After lyophilization the product was obtained as white powders on yields between 60-50%.

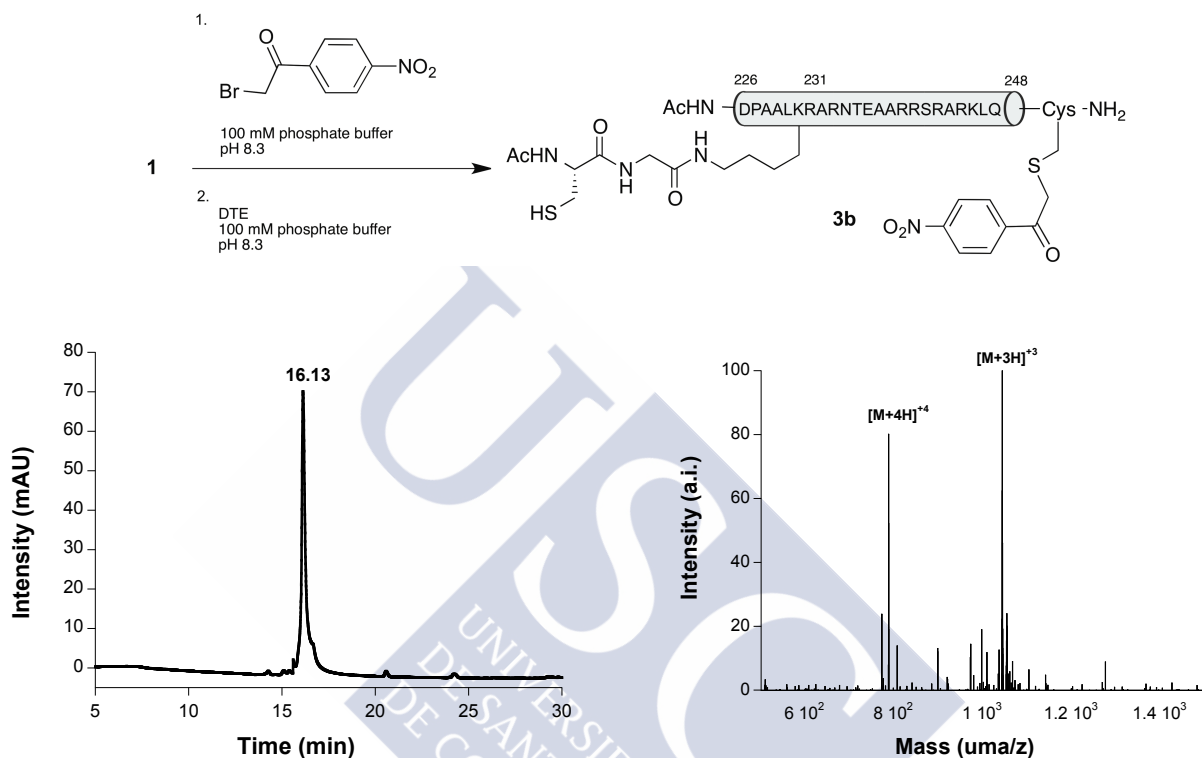


Figure S6. Left: HPLC chromatogram of the purified peptide **3b**. Right: ESI⁺-MS of **3b**. Peptide **3b**: ESI-MS [MH]⁺ calcd. for C₁₂₈H₂₁₉N₅₀O₃₉S₂ = 3146.5, found = 3146.4. UV (H₂O) λ_{max} , nm (ϵ): 265 (17,500 M⁻¹ cm⁻¹).

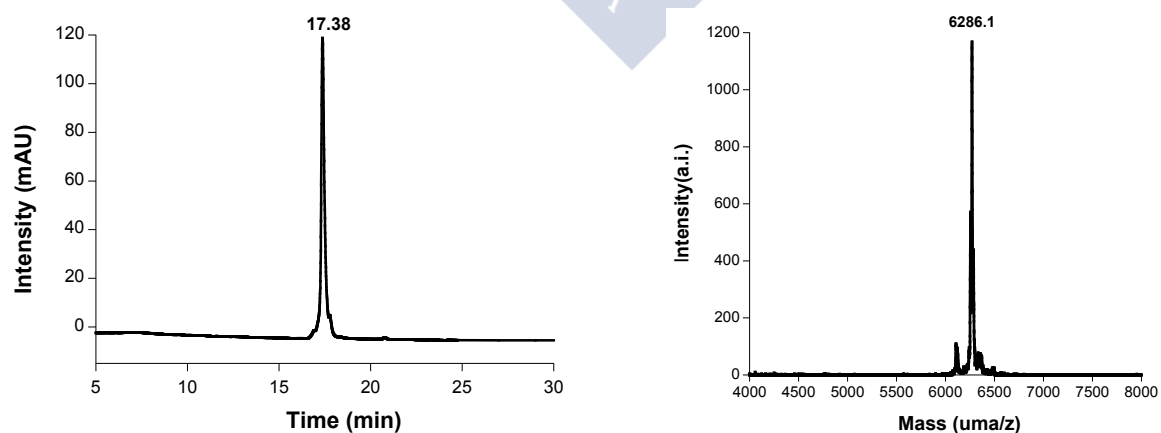


Figure S7. Left: HPLC chromatogram of the purified dimer (**3b**)₂SS at 222 nm. Right: MALDI of (**3b**)₂SS. Dimer (**3b**)₂SS: ESI-MS [MH]⁺ calcd. for C₂₅₆H₄₃₅N₁₀₁O₇₈S₄ = 6287.2, found = 6287.0. UV (H₂O) λ_{max} , nm (ϵ): 265 (35,000 M⁻¹ cm⁻¹).

Synthesis of the peptide 4

After coupling the Lys(Mtt) residue to the resin, the Mtt group was removed by using the protocol previously described. 5,5''-dimethyl-[2,2':6',2''-terpyridine]-4'-carboxylic acid (61 mg, 0.2 mmol) and HATU (76 mg, 0.2 mmol) was dissolved in DIEA (0.12 M in DMF, 2.5 mL), this solution was stirred for 1 min and then added onto the resin (approx. 0.05 mmol). After 45 min, the resin was filtered off and washed with DMF (3 × 1.5 mL × 2 min). Next, the Fmoc was removed under standard conditions, and the peptide was elongated by the previously described SPPS procedure with the lysine at position 231 inserted as Fmoc-Lys(alloc)-OH. Once introduced the acetyl group at the N terminus, the protecting group alloc was removed and the glycine and cysteine were coupled using SPPS standard protocols. Finally, the N-terminus was acetylated, the residue treated with cleavage cocktail and the crude purified by RP-HPLC.

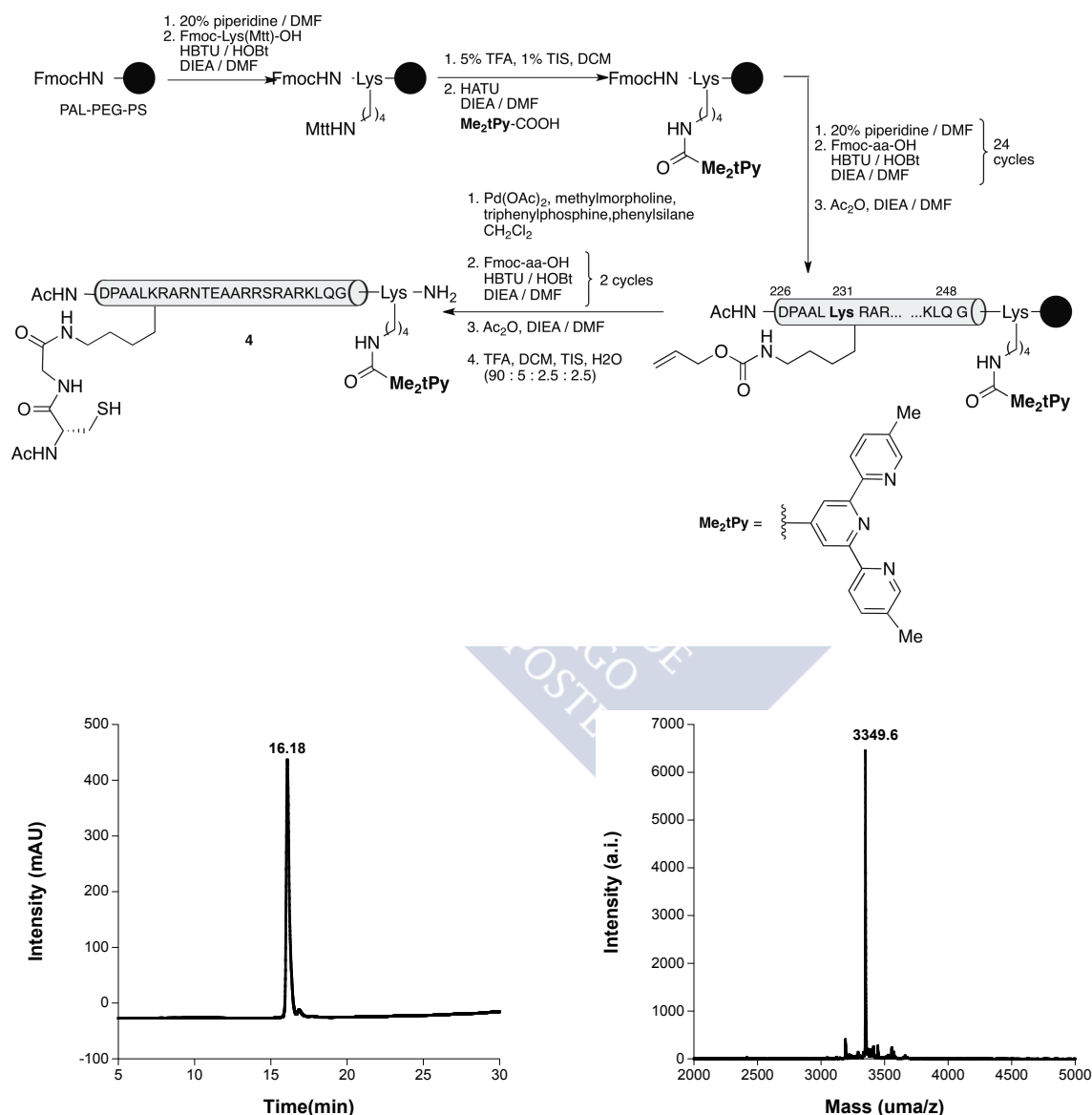


Figure S8. Left: HPLC chromatogram of the purified peptide **4**. Right: ESI⁺-MALDI of **4**. Peptide **4**: ESI-MS [MH]⁺ calcd. for C₁₄₃H₂₃₇N₅₄O₃₈S = 3352.8, found = 3352.2. UV (H₂O) λ_{max}, nm (ε): 280 (18,809 M⁻¹cm⁻¹).

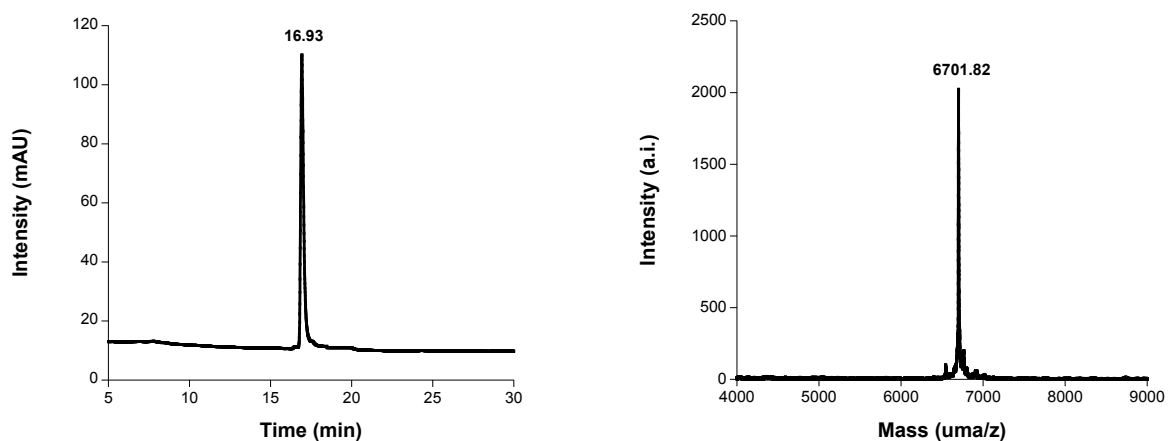


Figure S9. Left: HPLC chromatogram of the purified dimer $(4a)_2SS$. Right: MALDI of $(4a)_2SS$. Dimer $(4)_2SS$: ESI-MS $[MH]^+$ calcd. for $C_{286}H_{471}N_{108}O_{76}S_2 = 6702.6$, found = 6703.0. UV (H_2O) λ_{max} , nm (ϵ): 280 ($37,618 M^{-1}cm^{-1}$).

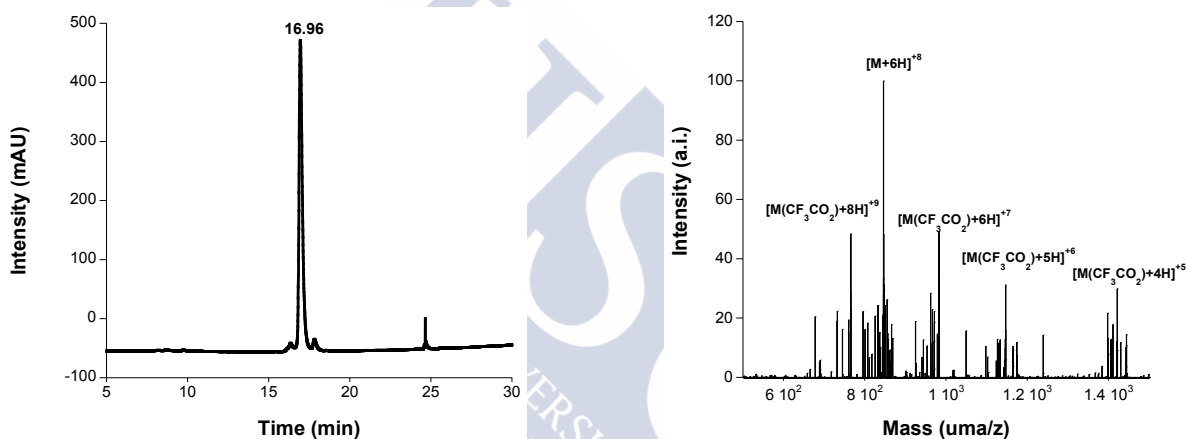


Figure S10. Left: HPLC chromatogram of the purified dimer $(4)_2Ni$. Right: ESI⁺-MS of $(4)_2Ni$. Dimer $(4)_2Ni$: ESI-MS $[M-H]^+$ calcd. for $C_{286}H_{471}N_{108}NiO_{76}S_2 = 6761.3$, found = 6762.0.

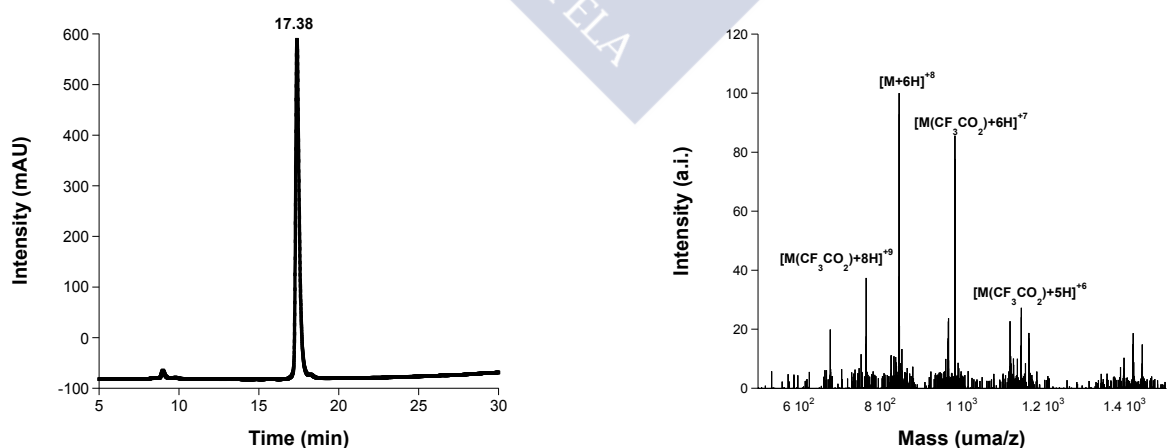


Figure S11. Left: HPLC chromatogram of the purified dimer $(4)_2Fe$. Right: ESI⁺-MS of $(4)_2Fe$. Dimer $(4)_2Fe$: ESI-MS $[M-H]^+$ calcd. for $C_{286}H_{471}N_{108}FeO_{76}S_2 = 6758.4$, found = 6756.4.

The ESI-TOF of the crude resulting from adding 5 equiv of $\text{Ni}(\text{ClO}_4)_2$ to $(4)_2\text{SS}$ shows products in which the peptide is complexed to one Ni^{+2} atom. ESI-MS $[\text{MH}]^+$ calcd. for $\text{C}_{286}\text{H}_{470}\text{N}_{108}\text{NiO}_{76}\text{S}_2 = 6759.3$, found = 6759.8.

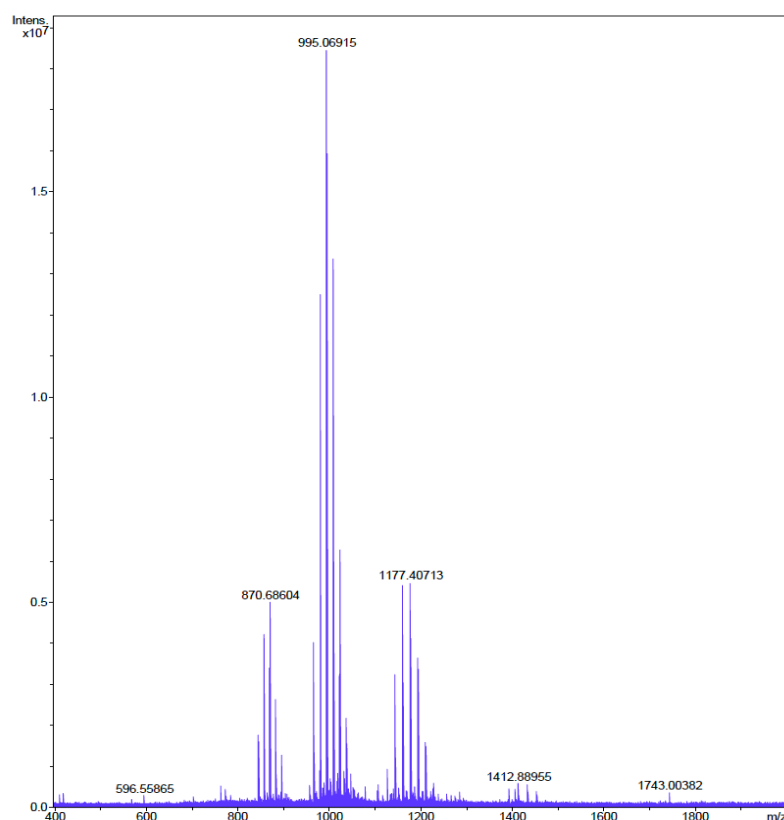
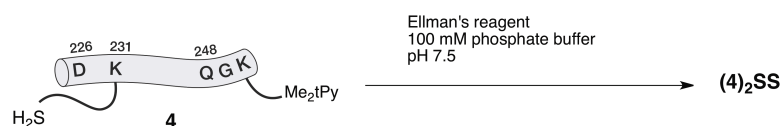


Figure S12. ESI-TOF of $(4)_2\text{SS}$ with 5 equiv of $\text{Ni}(\text{ClO}_4)_2$. The labels correspond to the highest peaks of each set, the rest of the peaks are the peptide with different numbers of ClO_4^- .
 $[\text{M}(\text{ClO}_4)_3+6\text{H}+\text{Ni}]^{+5}=1412.8$ $[\text{M}(\text{ClO}_4)_3+7\text{H}+\text{Ni}]^{+6}=1177.4$ $[\text{M}(\text{ClO}_4)_2+7\text{H}+\text{Ni}]^{+7}=995.8$
 $[\text{M}(\text{ClO}_4)_3+9\text{H}+\text{Ni}]^{+8}=870.7$.

General method to obtain the disulfide dimers

The monomeric thiol-containing peptide was dissolved in 100 mM buffer phosphate (pH 7.5) to give an approximate peptide concentration of 2 mM. Ellman's reagent^[S5] (0.8 mM in acetonitrile, 0.4 equiv) was added, and the resulting mixture was stirred for 20 min at rt. The reaction was quenched using 0.1% aqueous TFA, and the crude purified by RP-HPLC. The collected fractions were lyophilized and stored at $-20\text{ }^\circ\text{C}$, after characterization. The dimeric peptides were obtained in an approximated 70% yield.



Binding studies

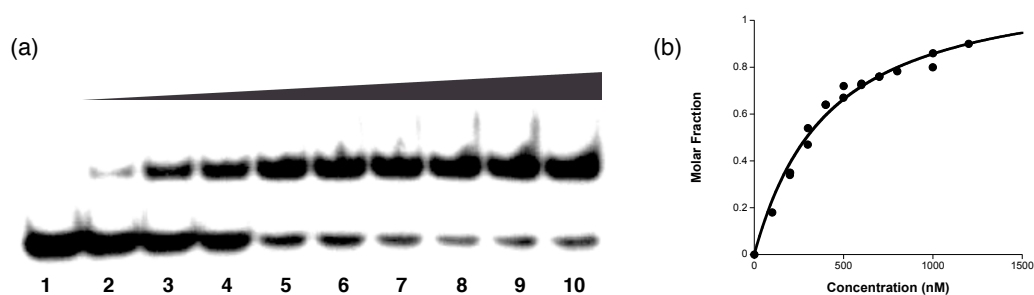


Figure S13: Radioactive EMSA titration of peptide (4)₂SS with dsDNA *inv*. (a) Lanes 1-10: dsDNA *inv* (100 nM, ~ 100 pM labeled with P³²); Lanes 1-10: 0, 100, 200, 300, 400, 500, 600, 800, 1000 and 1200 nM of peptide (4)₂SS in presence of 5 equiv of Ni(ClO₄)₂. (b) binding isotherm resulting from densitometry analysis of the three different gels. The curve represents the best fit to the data using nonlinear analysis with the DynaFit program (Biokib Software) to the equation derived using a 1:1 model ($K_d=375 \pm 40$ nM at 4 °C).^[S3]

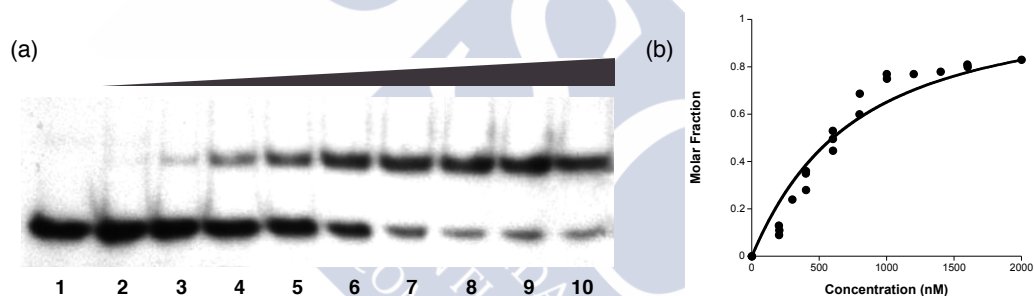


Figure S14. Radioactive EMSA titration of peptide (4)₂Ni with dsDNA *dir*. (a) Lanes 1-10: dsDNA *dir* (100 nM, ~ 100 pM labeled with P³²); Lanes 1-10: 0, 100, 200, 300, 400, 600, 800, 1200 and 1600 nM of peptide 4 with 5 eq. of Ni(ClO₄)₂. (b) Binding isotherm resulting from densitometry analysis of the three different gels. The curve represents the best fit to the data using nonlinear analysis with the DynaFit program (Biokib Software) to the equation derived using a 1:1 model ($K_d=299 \pm 26$ nM at 4 °C for (4)₂Ni).

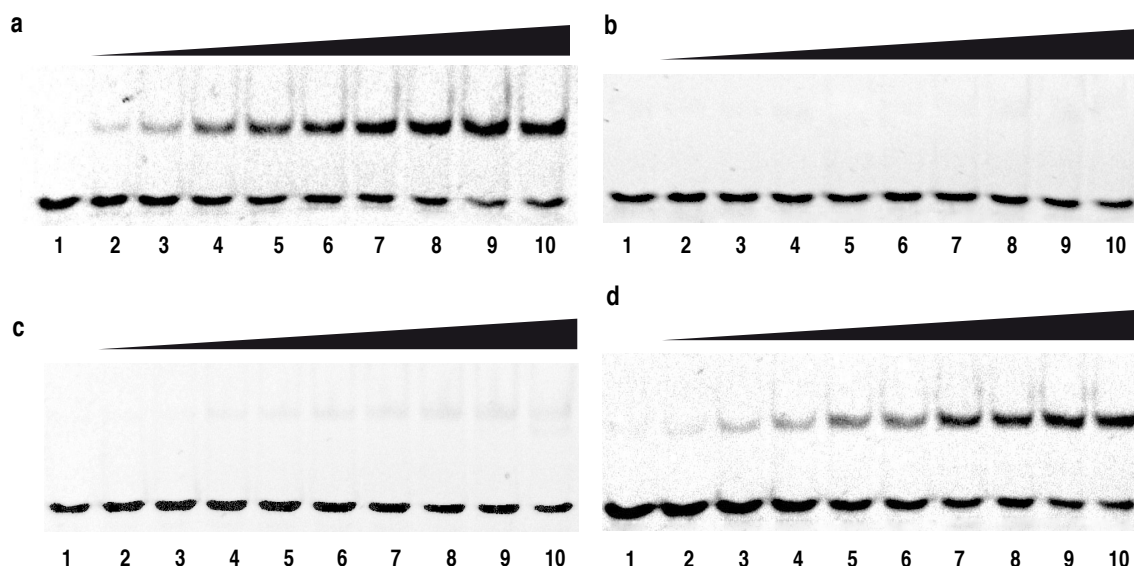


Figure S15. Fluorescence (SYBR gold staining) EMSA analysis of the DNA binding properties of peptide 4 and its dimeric derivatives. (a) Lanes 1-10: dsDNA *dir* (100 nM); lanes 1-10: 0, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800 nM of peptide 4 with 5 equiv of $\text{Ni}(\text{ClO}_4)_2$, using excess of TCEP. (b) Lanes 1-10: dsDNA *inv* (100 nM); lanes 1-10: 0, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800 nM of peptide 4 with 5 equiv of $\text{Ni}(\text{ClO}_4)_2$. (c) Lanes 1-10: dsDNA *dir* (100 nM); lanes 1-10: 0, 100, 200, 300, 400, 500, 600, 700, 800, 900 of the peptide $(4)_2\text{SS}$ in presence of 5 equiv of $\text{Ni}(\text{ClO}_4)_2$. (d) Lanes 1-10: dsDNA *inv* (100 nM); lanes 1-10: 0, 100, 200, 300, 400, 500, 600, 700, 800, 900 of the peptide $(4)_2\text{SS}$ in presence of 5 equiv of $\text{Ni}(\text{ClO}_4)_2$. The protocol involved addition of peptide solutions in 20 mM Tris-HCl pH 7.5 to 18 mM Tris-HCl pH 7.5, 90 mM KCl, 1.8 mM MgCl_2 , 1.8 mM EDTA, 9% glycerol 0.11 mgmL⁻¹ BSA, 2.25% NP-40 and 100 nM of DNA (4 °C, 30 min) and loading into the gel. Oligonucleotides: *dir*: 5'-CGG ATGA CG TCAT TTTTTC-3'; *inv*: 5'-CGG TCAT CG ATGA TTTTTC-3';

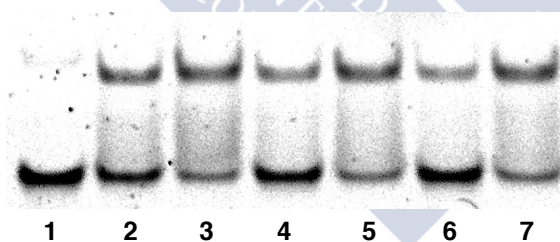


Figure S16. Fluorescence (SYBR gold staining) EMSA analysis of the DNA binding properties of peptide $(3a)_2\text{SS}$ with different metals. Lane 1-7: dsDNA *inv* (100 nM); Lane 2,3: 300 and 600 nM of $(3a)_2\text{SS}$; Lane 4,5: 300 and 600 nM of $(3a)_2\text{SS}$ and 5 eq. of $\text{Ni}(\text{ClO}_4)_2$; Lane 6,7: 300 and 600 nM of $(3a)_2\text{SS}$ and 5 eq. of $\text{Co}(\text{ClO}_4)_2$. Oligonucleotide sequences (binding sites underlined, only one strand shown): *inv*: 5'-CGG TCAT CG ATGA TTTTTC-3'.

Circular Dichroism spectroscopy

CD measurements were made in a 2 mm cell at 4 °C. Samples contained 5 μM of corresponding dsDNA (when present) and 5 μM of peptides in 10 mM phosphate buffer (pH 7.5) and 100 mM of NaCl. The CD spectra of the peptides (when measured in the presence of DNA) were calculated as the difference between the spectrum of the peptide/DNA mixture and the measured spectrum of a sample of the DNA oligonucleotide.

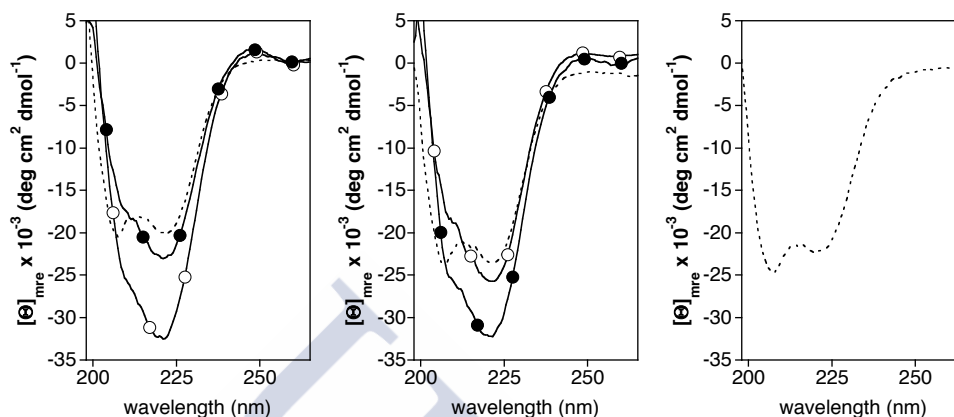
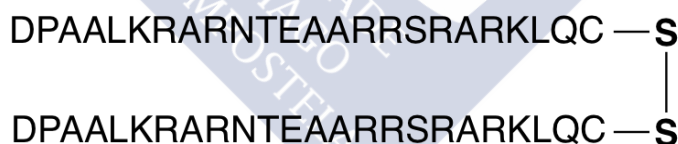
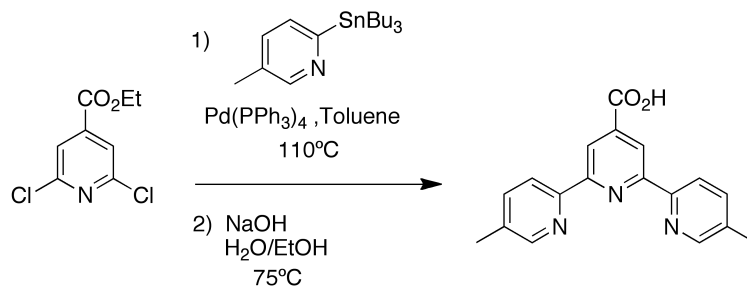


Figure S17. Left: circular dichroism spectra of a 5 μM solution of **(4)₂SS** in the presence of 5 eq. of $\text{Ni}(\text{ClO}_4)_2$ (dashed line); same solution in the presence of 5 μM of its target DNA **inv** (white circles) or 5 μM of the non-consensus **dir** (black circles); center: circular dichroism spectra of a 5 μM solution of **4** in the presence of 5 equiv of $\text{Ni}(\text{ClO}_4)_2$ (dashed line); same solution in the presence of 5 μM of **inv** (white circles) or 5 μM of the target DNA **dir** (black circles). Right: circular dichroism of the basic region control peptide disulfide dimer indicated below. CD measurements were made in 10 mM phosphate buffer (pH 7.5) and 100 mM of NaCl at 4 °C.



Synthesis of terpyridines

5,5''-dimethyl-[2,2':6',2''-terpyridine]-4'-carboxylic acid



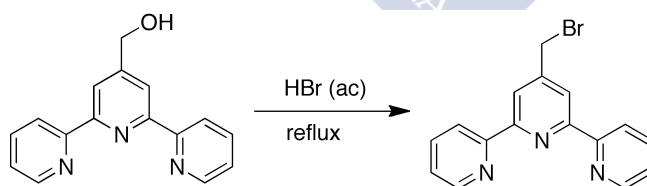
A mixture of 5-methyl-2-(tributylstannyl) pyridine (3.45 g, 9.03 mmol) and Pd(PPh₃)₄ (255 mg, 0.22 mmol) and ethyl 2,6-dichloroisonicotinate (800 mg, 3.64 mmol),^[S4] in degassed toluene (15 mL) was refluxed under argon for 2 days. After removal of the solvent, the black residue was treated with aqueous HCl (11 mL, 6 M). The suspension was extracted with CH₂Cl₂, and the organic phase was washed again with HCl (5 mL, 6 M). The aqueous solution was added dropwise into cold aq. ammonia (50 mL, 10%). The precipitate was filtered and washed with cold water. The ethyl 5,5''-dimethyl-[2,2':6',2''-terpyridine]-4'-carboxylate was suspended in a mixture of EtOH (11 mL) and 2.5 M aqueous NaOH (11 mL) and heated at 75 °C for 14 h. The reaction crude was concentrated under reduced pressure, and the residue purified by preparative reverse-phase chromatography (Buchi Sepacore,). The combined fractions were concentrated and freeze-dried to obtain the desired product (white solid, 642 mg, 2.105 mmol, 58% overall yield).

¹H RMN δ (400 MHz, DMSO): 2.37 (s, 6H), 7.87 (d, J = 8.1 Hz, 2H), 8.53 (d, J = 8.1 Hz, 2H), 8.57 (broad signal, s, 2H), 8.75 (broad signal, s, 2H).

¹³C RMN δ (400 MHz, DMSO): 18.3 (CH₃), 120.0 (CH), 121.4 (CH), 135.4 (C), 139.4 (CH), 141.1 (C), 149.2 (CH), 151.2 (C), 155.5 (C), 166.4 (C).

ESI-MS: [M+H]⁺ calc. for C₁₈H₁₆N₃O₂ = 306.33; found 306.12 [MH]⁺.

4'-(bromomethyl)-2,2':6',2''-terpyridine



A solution of [2,2':6',2''-terpyridin]-4'-ylmethanol (100 mg, 0.38 mmol) was refluxed in concentrated hydrobromic acid (47% in water, 5 mL) for 5 h. The solution was cooled to room temperature and brought to pH 8 with aqueous saturated Na₂CO₃. Then, the solution was extracted CH₂Cl₂ (3 x 10 mL). The organic layer was dried with Na₂SO₄ and concentrated under low pressure yielding the desired crude product (99.2 mg, 80%).

¹H RMN δ (250 MHz, CDCl₃): 8.54 (d, 2H); 8.45 (d, 2H); 8.32 (s, 2H); 7.70 (t, 2H); 7.18 (t, 2H); 4.41 (s, 2H).

¹³C RMN δ (250 MHz, CDCl₃): 156.0 (C); 155.5 (C); 149.0 (C); 148.1 (CH); 136.8 (CH); 124.0 (CH); 121.2 (CH); 120.7 (CH); 30.9 (CH₂).

ESI-MS: [M+H]⁺ calc. for C₁₆H₁₂BrN₃ = 327.02; found 327.1 [MH]⁺.

References

- [S1] a) K. C. Engman, P. Sandin, S. Osborne, T. Brown, M. Billeter, P. Lincoln, B. Nordén, B. Albinsson, L. M. Wilhelmsson. *Nucl. Acids Res.* **2004**, *17*, 5087-5095. b) G. Kallansrud, B. Ward. *Anal Biochem.* **1996**, *236*, 134-138.
- [S2] W. Hancock, W. & Battersby, J. *Anal. Biochem.* **1976**, *71*, 260–264.
- [S3] Kuzmic, P. Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase *Anal. Biochem.* **1996**, *237*, 260-273.
- [S4] This compound was synthesized following known procedures from acid derivate: Rohan J. Kumar, & Leif Hammarstrom. *Chem. Eur. J.* **2010**, *16*, 2830– 2842.
- [S5] a) Hermanson, G. T. *Bioconjugate Techniques*. Academic Press; 2nd Ed, 2008; b) S. Arpicco, F. Dosio, P. Brusa, P. Crosasso, L. Cattel, *Bioconjugate Chem.* **1997**, *8*, 327 – 337.

