TESIS DE DOCTORADO

METALLOPEPTIDES AND METALLOPROTEINS IN CHEMICAL BIOLOGY: FROM DNA BINDING TO INTRACELLULAR CATALYSIS

Soraya Learte Aymamí

ESCUELA DE DOCTORADO INTERNACIONAL DE LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOCTORADO EN CIENCIA Y TECNOLOGÍA QUÍMICA

SANTIAGO DE COMPOSTELA / LUGO

AÑO 2020

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DEL AUTOR DE LA TESIS

METALLOPEPTIDES AND METALLOPROTEINS IN CHEMICAL BIOLOGY: FROM DNA BINDING TO INTRACELLULAR CATALYSIS

Dña. Soraya Learte Aymamí

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento, y declaro que:

1) *La tesis abarca los resultados de la elaboración de mi trabajo.*
2) *En su caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.*
3) *La tesis es la versión definitiva presentada para su defensa y coincide con la versión enviada en formato electrónico.*
4) *Confirmo que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.*

En **Santiago de Compostela**, **22 de diciembre de 2020**

Firma electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DEL DIRECTOR/TUTOR DE LA TESIS

D./Dña. **José Luis Mascareñas Cid**

En condición de: **Tutor/a y director/a**

Título de la tesis: **Metallopeptides and metalloproteins in chemical biology: from DNA binding to intracellular catalysis.**

INFORMA:

Que la presente tesis, se corresponde con el trabajo realizado por D/Dña **Soraya Learte Aymamí**, bajo mi dirección/tutorización, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director/tutor de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.

En **Santiago de Compostela**, **22 de diciembre de 2020**

Firma electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A mis padres.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

TABLE OF CONTENTS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Abbreviations

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Θ	molar elipcity
°C	celsius degree
λ	wavelength
A	adenine
aa	amino acid
Aba	4- acetamidobenzoic acid
Ac	acetyl
alloc	allyloxycarbonyl
approx	approximately
ATP	adenosine triphophate
a.u.	arbitraty units
Boc	*tert*-butyloxycarbonyl
Bp	Base pair
Bpy	2,2´-bipyridine
BSA	bovine serum albumin
C	cytosine
calcd.	calculated
CD	circular dichroism
cm	centimeter
ds	double-stranded
DCM	dicloromethane
DEDTC	diethyldithiocarbamate
DIEA	N,N´-diisopropylethylamine
DMF	N,N´-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
Equiv/eq	equivalent

ESI	Electro Spray Ionization
Fmoc	9-fluorenylmethoxycarbonyl
G	guanine
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HBTU	(2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High Performance Liquid Chromatography
K_D	dissociation constant
µg	microgram
µm	micrometer
µM	micromolar
mAU	absorbance milliunits
mdeg	millidegrees
mg	milligram
min	minute
mL	milliliter
mM	millimolar
mmol	millimole
mre	molar ellipticy per residue
mRNA	messenger ribonucleic acid
m/z	mass charge relation
M	molar
MALDI	Matrix Assisted Laser Desorption Ionization
MeOH	methanol
MQ	milli-Q water
MS	Mass Spectrometry
MW	molecular weight
nm	nanometer
nM	nanomolar

NMM	N-methylmorpholine
NMR	Nuclear Magnetic Resonance
NP-40	octylphenoxypolyethoxyethanol
PAGE	PoliAcrilamide gel electrophoresis
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl chloride
PBS	Phosphate Buffer Saline
PDB	Protein Data Bank
PNA	peptidonucleic acid
PPh_3	triphenylphosphine
$PhSiH_3$	phenylsilane
rt	room temperature
RNA	ribonucleic acid
RNApol	RNA polymerase
RNAr	ribosomal ribonucleic acid
RNAt	transfer ribonucleic acid
RP-HPLC	Reverse Phase High-Performance Liquid Chromathography
SPPS	Solid-Phase Peptide Synthesis
t-Bu	*tert*-butyl
T	thymine
TBE	tris-boric-EDTA solution
TCEP	tris(2-carboxyethyl)phosphine
TF	transcription factor
TFA	trifluoroacetic acid
TIS	triisopropylsilane
TMR/TAMRA	5-carboxytetramethylrhodamine
Tris	tris(hydroxymethyl)aminometane
Trt	triphenylmethyl(trityl)
UV-Vis	Ultraviolet-visible

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Preface

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chemical Biology: Origins and Current Perspective

Defining 'Chemical Biology' has been a challenge since the term came into wide use in the 1990s. *Nature Chemical Biology* defines "Chemical Biology" as both the use of chemistry to advance a molecular understanding of biology, and the harnessing of biology to advance chemistry.[1] Therefore, Chemical Biology is a field at the interface between Chemistry and Biology that dissolves the boundaries between both disciplines.[2] The intrinsic interdisciplinarity of Chemical Biology complicates the precise definition of its scope, which covers a wide variety of topics, such as biomolecular recognition, enzyme catalysis, cell signaling, cell internalization and transport, biosynthesis, molecular mechanisms of gene expression and regulation, bioconjugation methods or bio-probe design, among others.

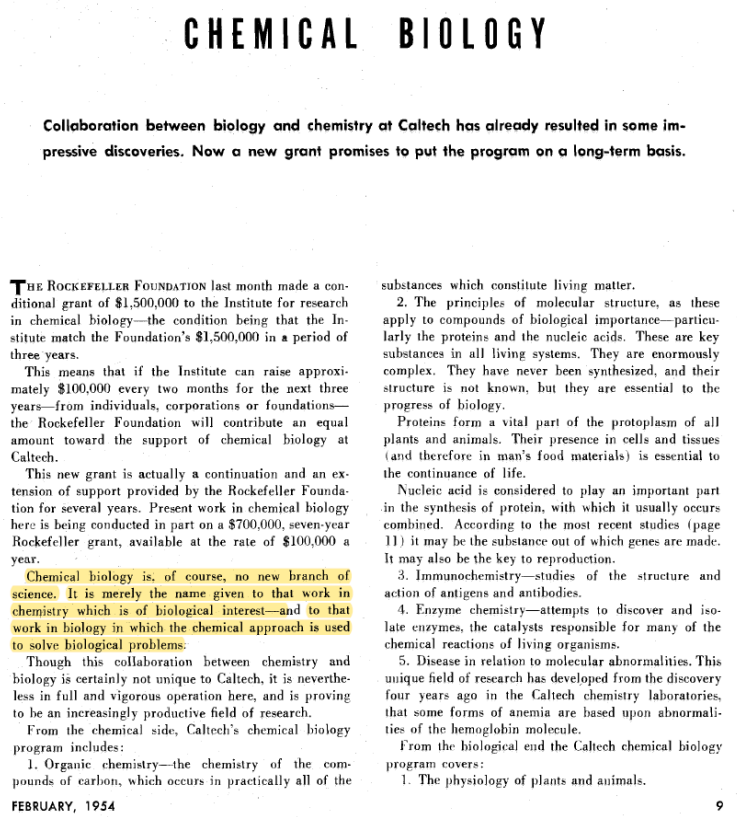
Research in this interdisciplinary field contributed to many scientific breakthroughs that begins in the 19th Century. In 1828, the German chemist Friedrich Wöhler isolated urea,[3] a compound found in urine, from mixing chemicals such as ammonium chloride and silver cyanate. Previously, urea had only been obtained from living organisms such as humans and dogs. At this time, there was a widespread belief in a "vital force" necessary for all biological compounds, but Wöhler's research showed that biological compounds could be made from inorganic materials.

Cellular imaging, which is very important in Chemical Biology, was also developed during this time, and useful compounds like aniline dyes for staining cells were invented.[4] In addition, chemicals began to be used to treat certain conditions by targeting specific pathogens. For example, the chemical compound Salvarsan, prepared by Paul Elrich in the 19th Century, was used to treat syphilis by targeting the bacteria that caused it.[5] Salvarsan was a huge improvement over the previous treatment at the time, which involved administering mercury, and caused serious problems such as blackened and loose teeth, kidney failure, and even death from poisoning. In the latter half of the 19th Century, the Swiss biochemist Friedrich Miescher used chemical compounds to isolate and break down the nuclei of cells.[6] He obtained substances that would be later termed "nucleic acids", and we know now that nucleic acids make up DNA, the genetic information of the cell.

Chemical Biology began to be thought of as a separate field in the 20th Century and although is in 1954 when it is for first time defined as an independent discipline,[7] the

[1] *Nat. Chem. Biol.* **2005**, *1*, 121.
[2] S. Aldridge, B. Imperiali, *Chem. Commun.* **2003**, *9*, 445-447.
[3] F. Wöhler, *Ann. Phys.* **1828**, *88*, 253-256.
[4] C. Weigert, *Anatom. Beiträge zur Lehre von den Pocken*. **1874**, Breslau.
[5] P. Ehrlich, *Berliner klinische Wochenschrif.* **1910**, 2346-2347.
[6] F. Miescher, *Medicinisch-chemische Untersuchungen*, **1871**, *4º edition.*
[7] *Eng. & Sci*. **1954**, *17*, 9-13.

term only came into widespread use in the 1990s. Indeed, in 1987 Arthur Kornberg, one of the biggest advocates of the integration of Chemistry and Biology, was still trying to convince his fellow biochemists of the value of Chemistry in biological research: *"Much of life can be understood in rational terms if expressed in the language of chemistry"*.[8]



CHEMICAL BIOLOGY

Collaboration between biology and chemistry at Caltech has already resulted in some impressive discoveries. Now a new grant promises to put the program on a long-term basis.

THE ROCKEFELLER FOUNDATION last month made a conditional grant of $1,500,000 to the Institute for research in chemical biology—the condition being that the Institute match the Foundation's $1,500,000 in a period of three years.

This means that if the Institute can raise approximately $100,000 every two months for the next three years—from individuals, corporations or foundations—the Rockefeller Foundation will contribute an equal amount toward the support of chemical biology at Caltech.

This new grant is actually a continuation and an extension of support provided by the Rockefeller Foundation for several years. Present work in chemical biology here is being conducted in part on a $700,000, seven-year Rockefeller grant, available at the rate of $100,000 a year.

Chemical biology is, of course, no new branch of science. It is merely the name given to that work in chemistry which is of biological interest—and to that work in biology in which the chemical approach is used to solve biological problems.

Though this collaboration between chemistry and biology is certainly not unique to Caltech, it is nevertheless in full and vigorous operation here, and is proving to be an increasingly productive field of research.

From the chemical side, Caltech's chemical biology program includes:

1. Organic chemistry—the chemistry of the compounds of carbon, which occurs in practically all of the substances which constitute living matter.

2. The principles of molecular structure, as these apply to compounds of biological importance—particularly the proteins and the nucleic acids. These are key substances in all living systems. They are enormously complex. They have never been synthesized, and their structure is not known, but they are essential to the progress of biology.

Proteins form a vital part of the protoplasm of all plants and animals. Their presence in cells and tissues (and therefore in man's food materials) is essential to the continuance of life.

Nucleic acid is considered to play an important part in the synthesis of protein, with which it usually occurs combined. According to the most recent studies (page 11) it may be the substance out of which genes are made. It may also be the key to reproduction.

3. Immunochemistry—studies of the structure and action of antigens and antibodies.

4. Enzyme chemistry—attempts to discover and isolate enzymes, the catalysts responsible for many of the chemical reactions of living organisms.

5. Disease in relation to molecular abnormalities. This unique field of research has developed from the discovery four years ago in the Caltech chemistry laboratories, that some forms of anemia are based upon abnormalities of the hemoglobin molecule.

From the biological end the Caltech chemical biology program covers:

1. The physiology of plants and animals.

FEBRUARY, 1954 9

Figure 1. Extract of Ref. 7 (Pag. 9) where the term Chemical Biology was defined for first time.

Fortunately, at the beginning of the 21st Century, Chemical Biology became an established field of research at the crossroads between Chemistry and Biology, and there have been enormous advances, to the point that most problems in biology can now be approached from a molecular perspective. Certainly, many of the spectacular achievements in biomedicine have been made possible thanks to developments in chemical sciences, which provide the means for synthesizing, modifying and analyzing increasingly complex systems.

Therefore, Chemical Biology is a rich and diverse field that defies a closed definition and that is constantly growing. It involves many topics: from the development of new bioconjugation methods for the selective modification of biomolecules to the design of new sensing strategies that allow the visualization of hidden phenomena, including the application of the rich properties of biological molecules, protein and nucleic acids in materials science or nanotechnology. Chemical Biology will be without any doubt at the core of many relevant advances in Science.

[8] A. Kornberg, *Biochemistry* **1987**, *26*, 6888-6891.

Chapter I: Metal-Promoted DNA Binding by Peptides and Miniproteins

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

General Introduction

1. Central Dogma of Biology

Living organisms are governed by their genes, the concrete sequence of nucleic acids that codes for proteins.[9] The flow of genetic information was postulated by Francis Crick in 1957 and was published one year later.[10] The idea of this postulate was called *Central Dogma of the Biology*, which gave DNA a key role in biology, and opened the modern era of molecular biology.

However, according to the Crick´s scheme, genetic information flows unidirectionally from the DNA to proteins via RNA. This directionality of the process was, and still is, one of the most controversial issues of this first proposal. This dogma has been actualized in the past years introducing new pathways for the flux of the information, taking account of epigenetic modification of gene expressions by proteins or the regulatory role of some RNAs.[11]

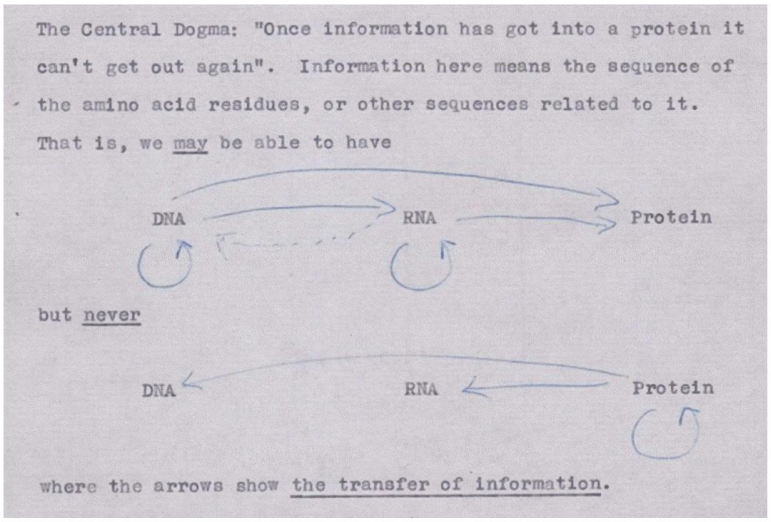


Figure 2. Crick´s first outline of the Central Dogma, from an unpublished note made in 1956 (Source: Wellcome Library, London).

[9] H. Lodish, A. Berk, L. S. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, *Molecular Cell Biology*, W. H. Freeman, **2000**, chapter 9.

[10] F. H. C. Crick, *Symp. Soc. Exp. Biol.* **1958**, *12*, 138-163.

[11] M. Cobb, *PLoS Biol.* **2017**, *15*, e2003243.

2. DNA: Function and Structure

2.1 Biological Function

Discovered in 1869 by Friedrich Miescher, DNA was identified as the genetic material in experiments carried out in the 40s by Oswald Avery, Colin MacLeod, and Maclyn McCarty. X-ray diffraction work of Rosalind Franklin and the observations of Erwin Chargaff were studied by James Watson and Francis Crick to build a model of DNA that we are familiar with today.

Deoxyribonucleic acid, or DNA, is the hereditary material in humans and almost all other living organisms. DNA is made up of four unique chemical bases, or nucleotides, including: adenine (A), cytosine (C), guanine (G) and thymine (T). This alphabet of nucleotides provides instructions for sequences of amino acids (three nucleotide combinations code for each of 20 amino acids), which the body uses to build proteins. A segment of DNA that codes for a protein is called a gene. The central dogma of biology states that a gene is transcribed into a messenger ribonucleic acid (mRNA) transcript, which is then translated into a protein.[12] Before the final mRNA is translated into protein, the pre-mRNA transcript contains both introns and exons. *Introns* are removed from the transcript and the *exons* are the portion of the gene that gets directly translated into proteins.[13]

The human genome consists of three billion base pairs of DNA and approximately 20.000 genes. The portion of the human genome that is protein-coding is very small, accounting for less than 2% of the genome. Of the remaining 98% of the genome, not used to encode proteins, its function is not fully understood. Some of it encodes RNA molecules with biological functions other than coding for proteins (an exception to the central dogma), and other regions have important roles in regulating the expression of genes.[14]

2.2 Structure of DNA

Based on its structure, DNA can be considered as a biopolymer formed by condensation of deoxyribonucleotides, which contain three primary structural components: a nitrogenous base, a pentose sugar and at least one phosphate (which bonds with the 3´-OH group of a sugar with the 5´-OH group of another sugar unit). The presence of the phosphate groups, negatively charged at physiological pH and oriented towards the exterior, makes of DNA a polyanionic polymer.

The heterocyclic bases are the molecular moieties of DNA responsible for the code. Adenine and guanine are purine bases, and cytosine and thymine are pyrimidines. Adenine pairs to thymine, while cytosine and guanine interact to one another. These complementary bases are bonded together via hydrogen bonds, which can be easily broken apart when the DNA needs to unzip and duplicate itself.

---

[12] F. Crick, *Nature* **1970**, *227*, 561-563.

[13] H. Pearson, *Nature* **2006**, *441*, 398-44.

[14] E. Pennisi, *Science* **2007**, *316*, 1556-1557.

DNA has a regular structure of a right-handed double helix forming antiparallel dimers. The three-dimensional structure of double-stranded DNA (dsDNA) is caused by the geometry and conformational preferences of pentoses and the hydrophobic nature of the bases, which tend to minimize their contact surface within the aqueous media.[15] This helix can adopt different structures depending on the conditions of the environment such as pH, ionic strength, solvent, etc. But the most common and relevant structure in physiological conditions is the B-DNA.[16]

B-DNA has a diameter of 20 Å approximately with 10 pairs of bases per each turn. The distance between the consecutive pairs is around 3.4 Å and the rotation per residue is 36°. Despite being a very regular and conserved structure, there are some deviations in the torsion angle, that are characteristic of the sequence and also play an important role in codifying for the interaction with proteins.[17] The geometry of the sugars leads to the formation of two grooves of different sizes, named major and minor grooves. The major groove is relatively wide and shallow, while the minor groove is narrow and deep. The width is also dependent on the sequence, being the regions rich in A-T narrower while G-C regions are wider (Figure 3).

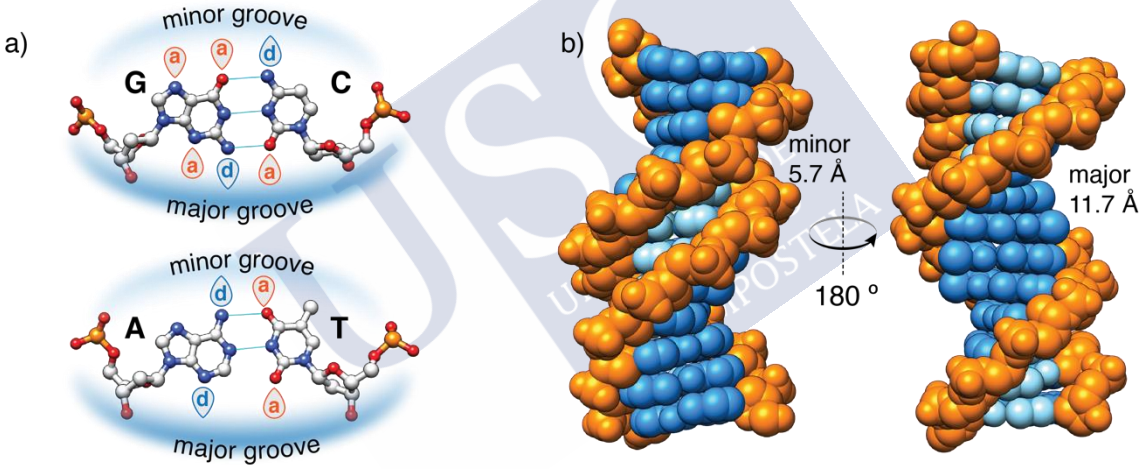


Figure 3. a) Structures of the nitrogenous bases forming the corresponding base pairs, showing the patterns of hydrogen-bond donors (d) and acceptors (a). b) B-form DNA model where the minor groove (light blue) and major groove (dark blue) can be observed (Source: adapted from *Eur. J. Org. Chem.* **2018**, *2018*, 249-261).

[15] S. Neidle, *Nucleic Acid Structure and Recognition*, Oxford University Press, USA, **2002**.

[16] G. M. Blackburn, M. J. Gait, D. Loakes, D. M. Williams, *Nucleic Acids in Chemistry and Biology*, RSC Publishing, Cambridge, **2006**.

[17] a) Z. Otwinowski, R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, P. B. Sigler, *Nature* **1988**, *335*, 321-329; b) R. E. Dickerson, H. R. Drew, *J. Mol. Biol.* **1981**, *149*, 761-786.

3. Proteins: Basic Concepts

3.1 Brief History

The word protein means primary substance, according to Mulder and Berzelius, who proposed the name in 1838. However, the first discovered protein was the famous description of gluten published by Jacopo Beccari, professor of chemistry at the University of Bologna, in 1745.[18] From the 19th Century the chemists started systematic and scientific studies of proteins as a unique family of biomolecules, though proteins were too complex to be studied by the analytic techniques available at that time. Scientists found that heating the proteins in concentrated acid afforded simple crystalline substances, called *amino acids*, which could be isolated and studied with their rudimentary methods. Modern protein chemistry can be dated back to 1820, when Henri Braconnot prepared glycine from gelatin in order to verify if proteins behaved like starch and are decomposed by acids, producing sugars.

By the beginning of the 20th Century, 18 amino acids were identified as components of proteins.[19] Soon, the three-letter abbreviations became tiresome to write, and a one-letter abbreviation system was finally established by Šorm in 1961.[20] Although the chemical structure of proteins has been known for a long time, the conformation of proteins was not established until the first protein structures were solved by X-ray crystallography in 1958 (myoglobin).[21] However, some years before, in 1951, Linus Pauling, Robert Corey and Herman Branson proposed the α-helix and the β-sheet structures, now known to form the backbones of tens of thousands of proteins.

3.2 Structure

The structure of large molecules such as proteins can be described at several levels of complexity. Four levels of protein structure are commonly defined. A description of all covalent bonds (mainly peptide bonds and disulfide bonds) linking amino acid residues in a polypeptide chain is its *primary structure*. The most important element of primary structure is the sequence of amino acids. *Secondary structure* refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns. *Tertiary structure* describes all aspects of the three-dimensional folding of a polypeptide. When a protein has two or more polypeptide subunits, their arrangement in space is referred to as *quaternary structure*.

The protein structure, reactivity and function are defined by the sequence of its amino acids, thus by its *primary structure*. The amino acids contain a carboxyl group (-COOH), an amino group ($-NH_2$), a hydrogen atom, and a variable side chain (-R). Most proteins are formed by the combination of 20 monomers, known as standard amino acids, that

[18] C. H. Bailey, *Cereal Chemistry* **1941**, *18*, 555-561.

[19] H. B. Vickery, C. L. A. Schmidt, *Chem. Rev*. **1931**, *9*, 169-318.

[20] F. Šorm, B. Keil, J. Vaněček, V. Tomášek, O. Mikeš, B. Meloun, V. Kostka, V. Holeyšovský, *Collect. Czech. Chem. Commun*. **1961**, *26*, 531-578.

[21] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, *Nature* **1958**, *181*, 662-666.

differ between each other in the chemical nature of the functional groups of their side chains.[22]

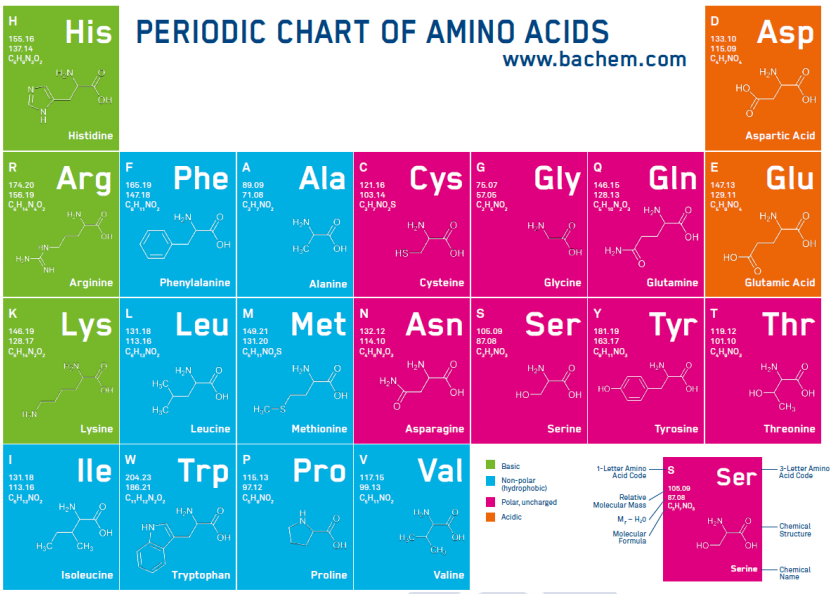


Figure 4. The "*periodic*" table of the 20 natural amino acids (Source: *www.bachem.com)*.

Proteins are polymers formed by condensation of the carboxyl group of an amino acid with the amino group of another amino acid, resulting in amide or peptide bond (Figure 5). The formation of amide bonds requires an energy input; however, they are kinetically stable once formed.[23] The peptide C–N bond is slightly shorter than a typical C–N bond in amines, which indicates that the bond between the carbonyl carbon and the amide nitrogen has partial double bond character; as a consequence, the atoms associated with the peptide bond are coplanar and the peptide bond cannot rotate freely.[24]

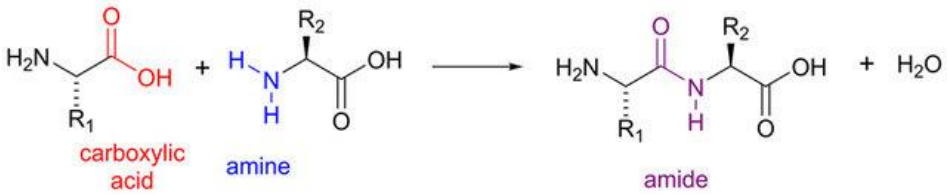


Figure 5. Formation of the peptide or amide bond.

The term *secondary structure* refers to any chosen segment of a polypeptide chain and describes the local spatial arrangement of its main-chain atoms, without considering its relationship to other segments. There are a few types of secondary structure that are particularly stable and occur widely in proteins, being the most frequent conformations

[22] a) D. Voet, J. G. Voet, C. W. Pratt, *Fundamentals of Biochemistry* **2007**, John & Sons, New York; b) B. Alberts, A. Johnson, J. Lewis, M. Raff, K Roberts, P. Walter, *Molecular Biology of the Cell* **2008**, Garland Publishing, New York, Chapter 3.

[23] R. B. Martin, *Biopolymers* **1998**, *45*, 351-353.

[24] D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry* **2017**, W.H. Freeman Company, New York, Chapter 4.

the α-helix and β-sheet. Where a regular pattern is not found, the secondary structure is sometimes referred to as undefined or as a random coil.

α-Helix

The α-helix is the most common secondary structure element, so that about 30% of the residues in proteins are in α-helical conformation.[25] The α-helix is right-handed, with 3.6 amino acid residues per helical turn, where the backbone of the polypeptide is wound around an imaginary axis and side chains of the amino acids are oriented to the exterior of the helix. This structure is stabilized by hydrophobic effects resulting from the compaction of the peptide backbone, where the carbonyl group of each peptide residue form a hydrogen bond with the amide group located four residues along the sequence (*i*,*i*+*4*).[24,26] The folding of a random coil into an *α*-helix is entropically unfavourable. Peptides that form helices in solution are typically found as a complex mixture of several helices, or most frequently, central helices with frayed-coil ends.[27]

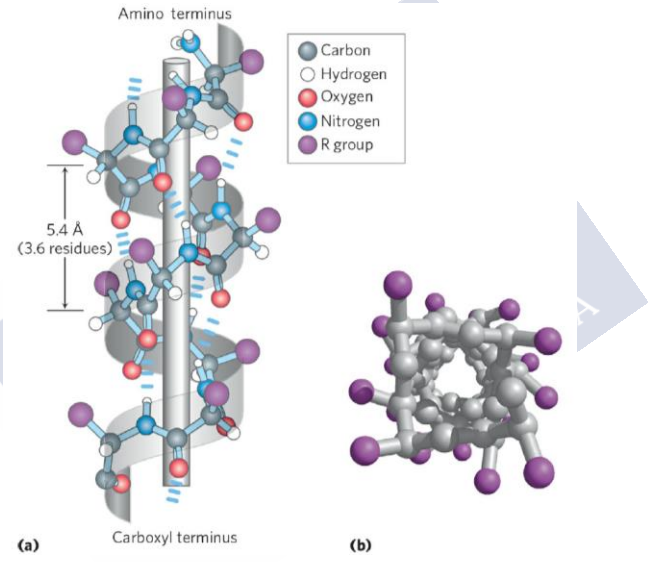


Figure 6. a) Ball-and-stick model showing the intrachain hydrogen bonds. b) The α helix viewed from one end, looking down the longitudinal axis (Source: adapted from[24]).

β-Sheet

This secondary structure is formed by two or more short polypeptide chains (5-8 residues) called β-strands. These strands are fully extended in a zig-zig form, creating a distance between residues of 3.5 Å. The residues of one strand are linked to the residues of a second strand via hydrogen bonding between N–H and C=O of the amide bonds. The individual segments that form a β-sheet are usually nearby on the polypeptide chain but can also be quite distant from each other in the linear sequence of the polypeptide; they may even be in different polypeptide chains. The R groups of adjacent amino acids

[25] D. J. Barlow, J. M. Thornton, *J. Mol. Biol.* **1988**, *201*, 601-619.

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

[27] D. T. Clarke, A. J. Doig, B. J. Stapley, G. R. Jones, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 7232-7237.

protrude from the zigzag structure in opposite directions, creating the alternating pattern.

We can find different substructures depending on the parallel or antiparallel orientation between the adjacent strands, being the antiparallel the most stable due to the minimal distortion of the inter-chain hydrogen bonding. In this β-sheet structure the side chains are oriented perpendicularly to the plane of the sheet in an alternating pattern.[24,26]

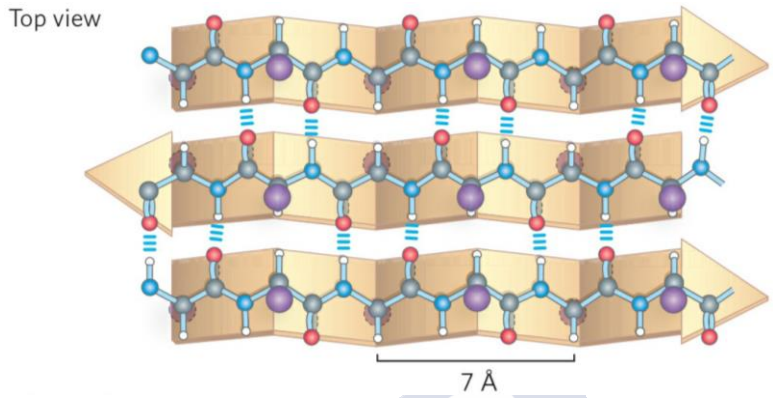


Figure 7. Antiparallel β-sheet conformation of polypeptide chains. Top view reveals the R groups extending out from the β-sheet and emphasizes the pleated shape formed by the planes of the peptide bonds. Hydrogen-bond crosslinks between adjacent chains are also shown (Source: adapted from[24]).

4. Gene Expression: General Concepts and Regulation

Gene expression is the process by which the information encoded in a gene is used to synthetize functional gene products. As we have commented before, these products are often proteins but in non-protein coding gene the resulting products are functional RNA, as transfer RNA (tRNA) or ribosomal RNA (rRNA). The process of gene expression is used by all known life (eukaryotes, prokaryotes and viruses) to generate the macromolecular machinery necessary for life. This process involves two main stages*: transcription* and *translation*.[9] Through the process of transcription, the information encoded in the DNA template strand is copied into a new molecule of messenger RNA (mRNA), which is transported to the cytoplasm. There, the translation process takes place using this mRNA to direct the protein synthesis.

The identity of an organism is determined by its genes and more specifically by its expression pattern, which defines its phenotype.[28] All cells of a multicellular organism have the same genetic material, but not all of them are expressed in the same way and at the same time. *Gene regulation* includes a wide range of sophisticated mechanisms that are used by cells to increase or decrease the production of specific gene products. Thus, for example, the genes encoding the essential metabolic or structural components of a cell may be continuously expressed in all cell types, while other genes (in fact, most of them) are transiently expressed under particular circumstances. This implies that the morphology and proper functioning of cell depend largely on the expression of genetic material occurring in a controlled manner. Any step of gene expression may be modulated, from the transcription step to post-translational modification of a protein. However, it has been observed that the regulation occurs mainly at the stage of transcription[29] and specifically in the initiation of the process,[30] hence the discussion will focus on this stage.

4.1 Transcription

This step involves the synthesis of RNA molecules and takes place in the nucleus. During the transcription a strand of RNA is made, complementary to DNA strand, but with a ribose instead of deoxyribose and thymine replaced by uracil. For a protein-coding gene, the RNA copy, or transcript, carries the information needed to build a polypeptide or protein.

The main enzyme involved in this step is *RNA polymerase*, which builds an RNA strand in the 5´to 3´direction, adding each new nucleotide to the 3´end of the strand. In eukaryotes, there are multiple types of RNA polymerases which make the various types of RNA. Protein-coding genes are transcribed by RNA polymerase II to yield mRNAs; ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) are transcribed by RNA polymerases I and III.

---

[28] P. Taylor, R. Lewontin, *The Genotype/Phenotype Distinction* **2017**, The Stanford Encyclopedia of Philosophy, Edward N. Zalta.

[29] R. D. Kornberg, *Trends cell biology* **1999**, *9*, M46-M49.

[30] P. Cramer, *Nat. Struct. Mol. Biol.* **2007**, *14*, 686-687.

Generally speaking, polymerases are large enzymes that work together with several other specialized cell proteins.[31] These cell proteins, called *transcription factors*, help to determine which DNA sequences should be transcribed and when the transcription process should occur. The process consists of three steps:[32]

- Initiation: RNA polymerase (and its associated transcription factors) bind to the promoter and trigger the local unwinding of the DNA double helix.
- Elongation: Using one of the two DNA strands as a template, RNA polymerase synthesizes an RNA chain. There is a brief time during this process when the newly formed RNA is bound to the unwound DNA. At that moment, an adenine (A) in the DNA binds to an uracil (U) in the RNA.
- Termination: is the ending of transcription and occurs when RNA polymerase crosses a stop (termination) sequence in the gene.

Once termination is complete, the RNA molecule falls off the DNA template. At this point, in the case of mRNA, it undergoes a process in which noncoding nucleotide sequences, called introns, are clipped out of the pre-mRNA strand. This process, known as splicing, "tidies up" the molecule and removes nucleotides that are not involved in protein production. Then, a sequence of adenine nucleotides called a poly-A tail is added to the 3' end of the mRNA molecule. This sequence signals to the cell that the mRNA molecule is ready to leave the nucleus and enter the cytoplasm.

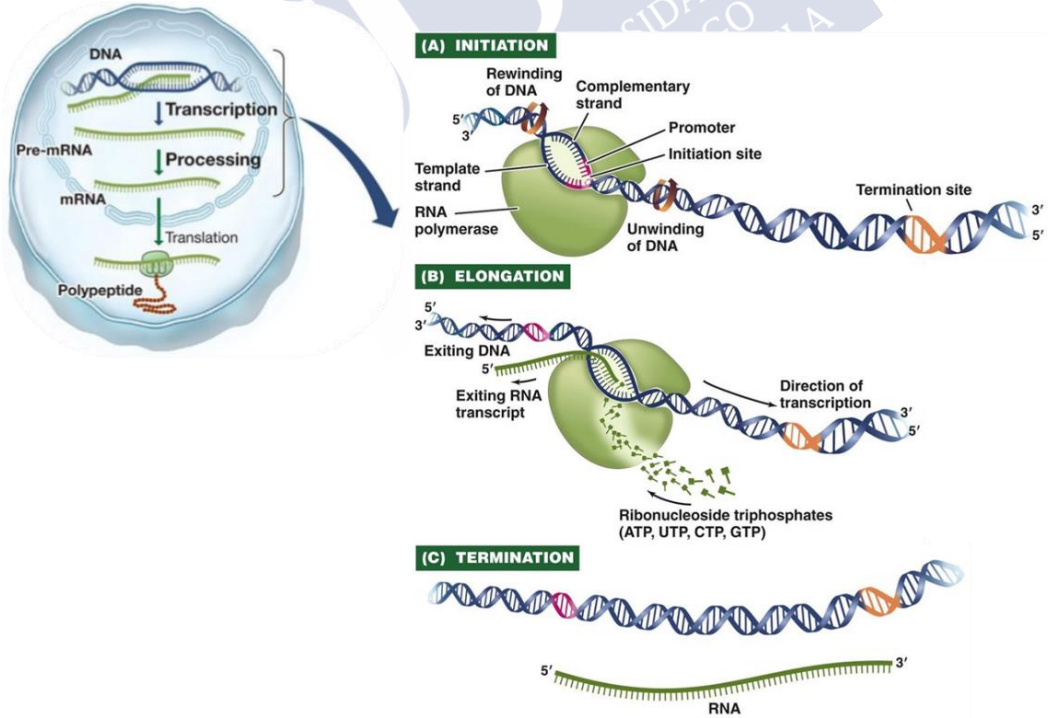


Figure 8. Illustration of the three stages of transcription process (Source: adapted from *Principles of Life* **2012**, Sinauer Associates and W.H. Freeman, Figure 10.5).

[31] G. M. Cooper, *The Cell: A Molecular Approach* **2000**, Sunderland (MA): Sinauer Associates, Chapter 6.

[32] J. D. Watson, T. A. Baker, S. P. Bell, A. A. Gann, M. Levine, R. M. Losick, *Molecular Biology of the Gene* **2013**, Pearson.

4.2 Transcription Factors

The complex transcription machinery of cells is primarily regulated by a set of proteins, transcription factors (TFs), that bind DNA at specific sites. Every TF can have from one to several dozens of such specific sites on the DNA. Upon binding to the site, TFs form a stable protein-DNA complex that can either activate or repress transcription of nearby genes, depending on the actual control mechanism.[33] The fact that approximately 5-10% of genes encode transcription factors highlights the importance of this family of proteins.

Understanding how proteins recognize their DNA-binding sites has a long history. On the basis of early low-resolution X-ray structures of nucleic acid duplexes, it was soon realized that the major groove of the DNA helix offered a set of base-specific hydrogen bond donors, acceptors, and nonpolar groups that could be recognized by a complementary set of donors and acceptors presented in the protein amino acid side chains. Accordingly, the idea soon evolved that short DNA sequences could serve as binding sites that were specifically read by a complementary sequence of amino acids.[34] This mechanism of protein-DNA recognition, now commonly referred to as *direct readout*, is evident in nearly all of the >1500 structures of protein-DNA complexes that have been solved and deposited in the Protein Data Bank (PDB). Nevertheless, as was realized many years ago, there is not a simple recognition code or one-to-one correspondence between DNA and protein sequences.[35]

Although elements of *direct readout* contribute to nearly all protein-DNA complexes, these structures also reveal that bound DNA frequently deviates from a standard B-form double helix. In some cases, these deviations can clearly contribute to the DNA-binding selectivity.[36] A bend or some other deformation of the DNA helix is required to establish a set of hydrogen bonds or nonpolar interactions between the protein and DNA that are much less likely to occur in the absence of the deformation. From such observations, it was coined the term *indirect readout*, to indicate those protein-DNA interactions that depend on base pairs that are not directly contacted by the protein. This broad definition includes situations where the DNA sequence creates or facilitates a DNA structure that is subsequently recognized by a protein, but also when the protein-DNA contact is mediated by a water molecule.[37]

Therefore, protein-DNA recognition utilizes a continuum of readout mechanisms that depend on the structural features and flexibility of both macromolecules. Moreover, in

[33] M. Slutsky, L. A. Mirny, *Biophys. J.* **2004**, *87*, 4021-4035.

[34] a) N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci.* **1976**, *73*, 804-808; b) M. A. Viswamitra, O. Kennard, P. G. Jones, G. M. Sheldrick, S. Salisbury, L. Falvello, Z. Shakked, *Nature* **1978**, *273*, 687-688.

[35] Z. Otwinowski, R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, P. B. Sigler, *Nature* **1988**, *335*, 321-329.

[36] Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **1993**, *365*, 512-520.

[37] C. L. Lawson, H. M. Berman, *Indirect readout of DNA sequence by proteins* **2008**, Protein-Nucleic Acid Interactions, Structural Biology, Cambridge, UK: R. Soc. Chem.

many cases it is necessary some type of cooperative oligomerization of the proteins in order to achieve high affinity recognition.[38]

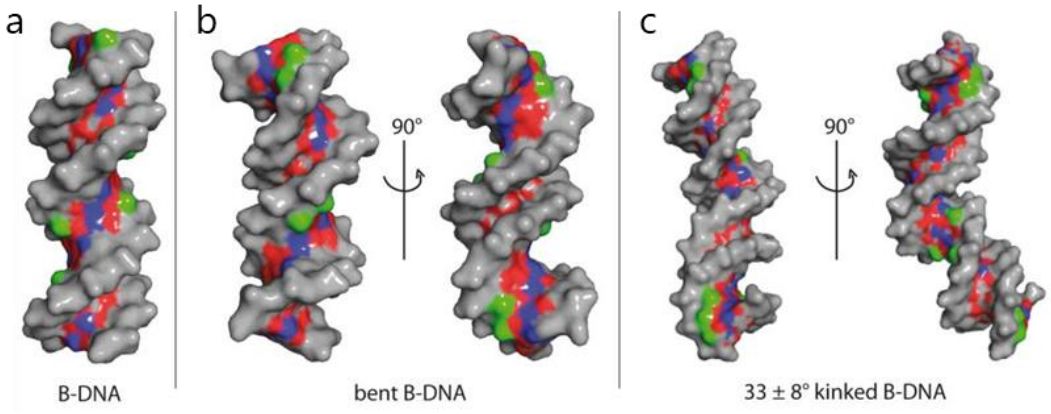


Figure 9. Base readout in the major and minor groove. a) Accessibility of functional groups in the major groove in a B-DNA. b) Locally increased readout accessibility in minor groove of bent B-DNA. c) Enlarged binding capability due to kink-induced shape alteration in B-DNA. Hydrogen bond donors in blue, acceptors in red and thymine methyl group in green (Source: adapted from *Int. J. Mol. Sci.* **2014**, *15*, 12335-12363).

Although transcription factors recognize DNA using a variety of folds, in many cases they share similar structural recognition motifs, which facilitates a classification into families. In many of these families the α-helix is the most frequently used secondary structure element for specific DNA recognition in the major groove. However, some helices can also insert into the minor groove, as occurs in the lac repressor family.[39] In these cases, kinking of the DNA is required to open the minor groove and thereby accommodate the helix. Furthermore, although less common than α-helices, β-strands and loops embedded in the mainly β-domain structures are used by some proteins to recognize specific DNA sequences.[40,36] Here, we present the most relevant families of TFs.

Helix-Turn-Helix Proteins

The proteins in this "family" span a broad range of protein folds that contain a conserved bihelical motif termed the helix-turn-helix (HTH), but are generally dissimilar in structure outside the HTH region. This motif, composed of approximately 22 amino acids, consists of two α-helices connected by a tight bend, although the length of each helix varies among different subclasses of this broadly defined family. The "recognition helix" of the HTH motif binds DNA through a series of hydrogen bonds and hydrophobic interactions with exposed bases, and the other helix stabilizes the interaction between the protein and DNA, but does not play a particularly strong role in the direct recognition.[41]

[38] a) R. Moretti, A. Z. Ansari, *Biochimie.* **2008**, *90*, 1015-1025; b) D. J. Segal, C. F. Barbas, *Curr. Opin. Chem. Biol.* **2000**, *4*, 34-39.

[39] a) M. A. Schumacher, K. Y. Choi, H. Zalkin, R. G. Brennan, *Science* **1994**, *266*, 763-770; b) P. Van Roey, C. A. Waddling, K. M. Fox, M. Belfort, V. Derbyshire, *EMBO J.* **2001**, *20*, 3631-3637.

[40] W. S. Somers, S. E. V Phillips, *Nature* **1992**, *359*, 387-393.

[41] a) C. W. Garvie, C. Wolberger, *Mol. Cell.* **2001**, *8*, 937-946; b) R. Wintjens and M. Rooman, *J. Mol. Bio.* **1996**, *262*, 294-313.

The HTH motif is very common in prokaryotic transcription factors, and usually operates in the form of oligomers with other TFs. When this motif is found in eukaryotes it is called *homeodomain*, and typically recognizes sequences involved in the regulation of the development. An important difference with the bacterial motif, is that this homeodomain can bind the target DNA sequences as monomers,[42] because it presents a C- or N-terminal tails that interacts with the adjacent minor groove.

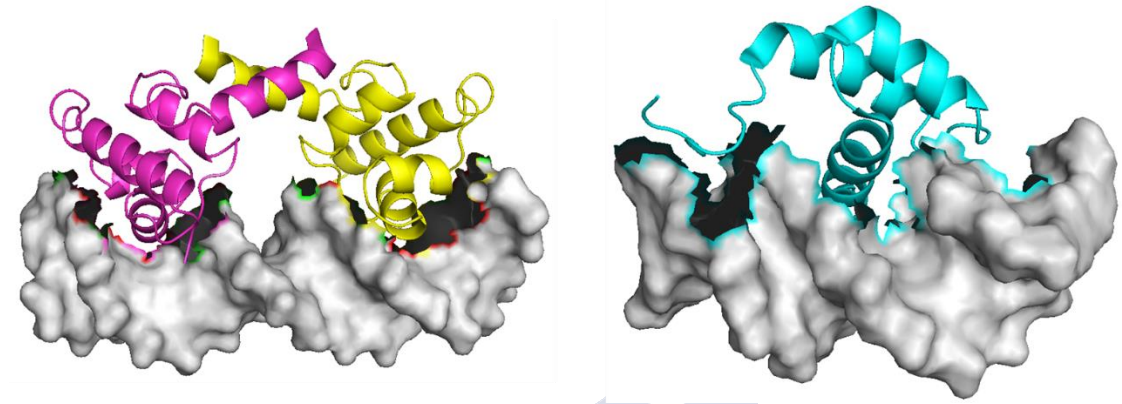


Figure 10. *Left*: HTH domain of the λ repressor of bacteriophage lambda (PDB 1LMB). *Right*: Antennapedia drosophila melanogaster homeodomain (PDB 1AHD).

Basic Region-Leucine Zipper (bZIP) and Helix-Loop-Helix Zipper (HLHbZIP) variants

These two classes of dimeric eukaryotic DNA binding proteins have a common mechanism of binding DNA and distinct but related modes of dimerization. The basic region-leucine zipper (bZIP) motif is employed for DNA recognition by a wide number of eukaryotic transcription factors involved in the control of cellular growth.[43] The bZIP proteins consist of long, uninterrupted α helices of about 60 residues. These proteins are formed by two distinct regions: a positively charged segment (*basic region*, BR) linked to a sequence of heptad repeats of leucine residues (*leucine zipper*). The basic region, a highly conserved segment rich in basic amino acids, is located at the N-terminus and is constituted by approximately 20 amino acids. This region is inserted in the major groove of DNA and therefore is responsible for most of the direct contacts with the DNA bases and phosphates. The other region known as leucine zipper is composed of about 30 amino acids located at the C-terminus. The leucine residues of its sequence form a parallel coiled-coil that provides the hydrophobic environment necessary for dimerization.[44] The leucine zipper and basic regions are connected through a spacer that is 6 amino acids in length.

A very important aspect in DNA binding by bZIP proteins is that in the absence of their target DNA sequence, the basic region is poorly structured and only adopts the characteristic α-helical conformation upon specific DNA binding. From a thermodynamic

[42] a) S. Khorasanizadeh, F. Rastinejad, *Curr. Biol.* **1999**, *9*, R456-458; b) P. W. H. Holland, *Wiley Interdiscip. Rev. Dev. Biol.* **2013**, *2*, 31-45.

[43] H. C. Hurst, *Protein Profile*, **1995**, *2*, 101-168.

[44] a) K. A. Lee, *Cell Sci.* **1992**, *103*, 9-14; b) E. K. O´Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science* **1991**, *254*, 539-544; c) T. K. Kerppola, T. Curran, *Curr. Opin. Struct. Biol.* **1991**, *1*, 71-79.

point of view, the interaction of bZIP proteins with DNA has a strongly unfavorable entropic term associated with the folding of basic region into α-helix. Due to this, the monomers cannot bind by themselves to their cognate DNA sequences with sufficient affinity. Only with the simultaneous interaction of two chains as homo- or heterodimers, the unfavorable entropic term is compensated by the enthalpy gain, allowing complexation with the target DNA.[45]

One of the best characterized bZIP proteins is GCN4, a transcriptional activator of yeast which preferentially interacts with palindromic sequences ATF/CREB and CRE (5´-ATGAcgTCAT-3´) and the pseudopalindromic sequence AP1 (5´-ATGAcTCAT-3´) as a homodimer (Figure 11, left).[46]

The basic region-helix-loop-helix (bHLHZIP) proteins share with the bZIP proteins a very similar mode of DNA binding. The salient difference lies in the dimerization region, which is composed of two helices separated by a loop (Figure 11, right). Both bZIP and bHLH proteins have many members that form both homo- and heterodimers, a feature that expands the repertory of DNA sequences that the proteins can recognize.

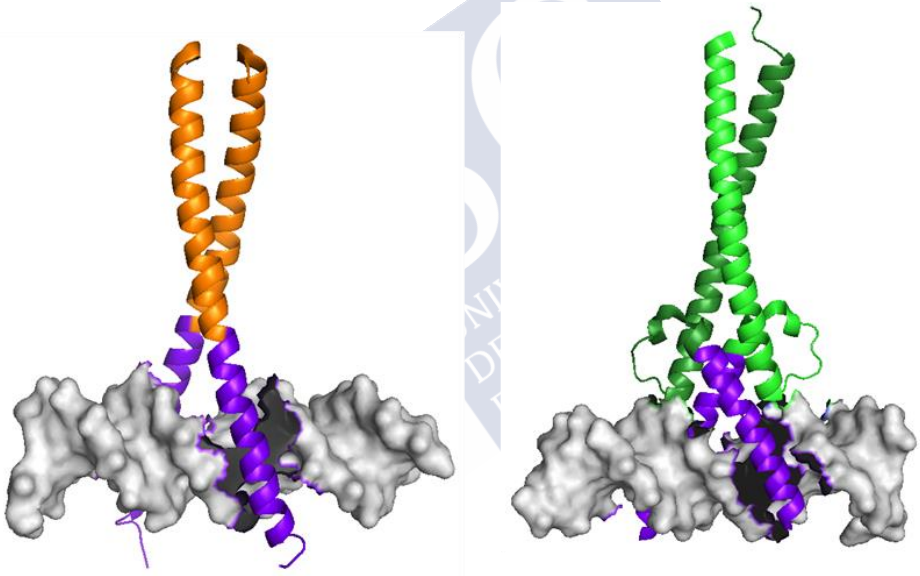


Figure 11. *Left*: bZIP binding domain of the transcription factor GCN4 (PDB 1YSA). *Right*: X-ray structure of the Myc/Max complex and its HLH binding domain (PDB 1NKP).

Zinc Finger Proteins

The zinc finger (ZF) family is the most abundant class of DNA binding proteins in the human genome, and it is formed by multiple copies of a compact, ~30-amino acid DNA-binding domain. ZFs are the most minimal of DNA binding domains, with a relatively short α-helix, a two-stranded antiparallel β-sheet, and a Zn^{2+} ion coordinated by cysteine

[45] a) M. Weiss, T. Ellenberger, C. Wobbe, I. Lee, S. Harrison, K. Struhl, *Nature* **1990**, *347*, 575-578; b) K. T. O'Neil, W. F. DeGrado, J. D. Shuman, C. Ampe, *Biochemistry*. **1991**, *30*, 9030-9034.

[46] a) I. A. Hope, K. Struhl, *Cell*. **1986**, *46*, 885-894; b) W. Keller, P. König, T. J. Richmond, *J. Mol. Biol*. **1995**, *254*, 657–667.

and histidine residues.[47] Zinc fingers are classified by the type and order of the zinc coordinating residues (Cys_2His_2, Cys_4, Cys_6 ,etc.), but the most common is the Cys_2His_2. Zinc fingers often occur as tandem repeats with two, three, or more fingers that can bind in the major groove, typically spaced at 3-bp (base pair) intervals. The α-helix of each domain (the recognition helix) makes sequence-specific contacts to DNA bases in the major groove; residues from a single recognition helix can contact four or more bases to yield an overlapping pattern of contacts with adjacent zinc fingers.

In the case of the Cys_2His_2 family, its members typically contain several fingers in tandem, and the system is highly modular. Of its 30 amino acids, 25 are folded into a compact unit, upon coordination of Zn^{2+}, consisting of an α-helix packed against a β-hairpin (ββα-domain) while the other five amino acids work as linkers between consecutive zinc fingers (Figure 12, left). In these structures, the Zn^{2+} is not directly involved in DNA binding but it is essential for the stability of the ββα-structure. It exhibits a tetrahedral coordination, where the two cysteines are located on the β-sheet and the two histidines are in the α-helix.[48]

In addition to the above zinc-finger group, there are other families of Zn^{2+}-containing transcription factors, the most prominent members being those belonging to the nuclear hormone receptor family, which bind DNA as noncovalent dimers (Figure 12, right).

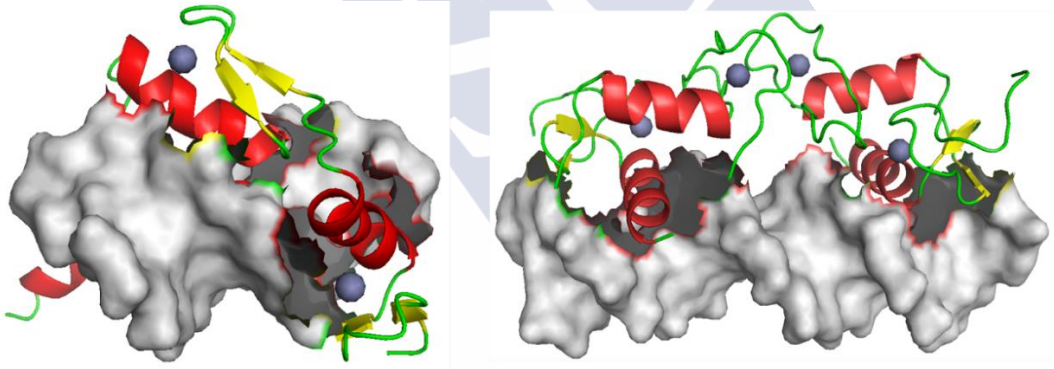


Figure 12. *Left*: Crystal structure of a ZIF268 8 transcription factor-DNA complex (PDB 1ZAA). *Right*: Crystal structure of the estrogen receptor DNA-binding domain (PDB 1HCQ).

Other DNA-Binding Motifs

While many transcription factors can be classified into the general categories presented above, some others combine features of different families. This is the case for GAGA factor, a 519 residue-long, that comprises a single classical Cys_2-His_2 zinc finger core, and an N-terminal extension containing two basic regions, BR1 and BR2. The zinc finger core binds in the major groove and recognizes the first three GAG bases of the consensus site in a manner similar to that seen in other classical zinc finger-DNA complexes. Unlike the latter, which require tandem zinc finger repeats with a minimum of two units for high

[47] a) N. P. Pavletich, C. O. Pabo, *Science* **1991**, *252*, 809-817; b) A. Klug, J. W. R. Schwabe, *FASEB J.* **1995**, *9*, 597-604.

[48] a) M. S. Lee, G. P. Gippert, K. V. Soman, D. A. Case, P. E. Wright, *Science* **1989**, *245*, 635-637; b) J. Miller, A. D. McLachlan, A. Klug, *J. Trace Elem. Exp. Med.* **2001**, *14*, 157-169.

affinity binding, the GAGA factor makes use of only a single finger complemented by BR1 and BR2. BR2 forms a helix that interacts in the major groove recognizing the last G of the consensus, while BR1 wraps around the DNA in the minor groove and recognizes the A in the fourth position of the consensus (Figure 13, left).[49]

Another interesting example is the *AT-hook* motif, a cationic peptide (RKPRGRPKK) found in eukaryotic HMG-I(Y) nuclear proteins.[50] This peptide interacts with DNA mainly through the minor groove. Although AT-hook binds to their target DNA sites with poor affinity (in the mM range), HMG-I(Y) proteins exploit the cooperative effect of three AT-hook repeats to achieve high DNA binding affinity.[51] NMR and crystallography studies have provided a detailed structural picture of the interaction between the AT-hook and the DNA and have shown that its central RGR core is deeply inserted into the minor groove in an extended concave conformation. The side chains of the Arg residues, extend at the bottom of the groove with their guanidinium groups hydrogen-bonded to the base edges in the groove, whereas various lysine residues in the sequence introduce additional electrostatic contacts with the phosphate components of the DNA backbone (Figure 13, right).[52]

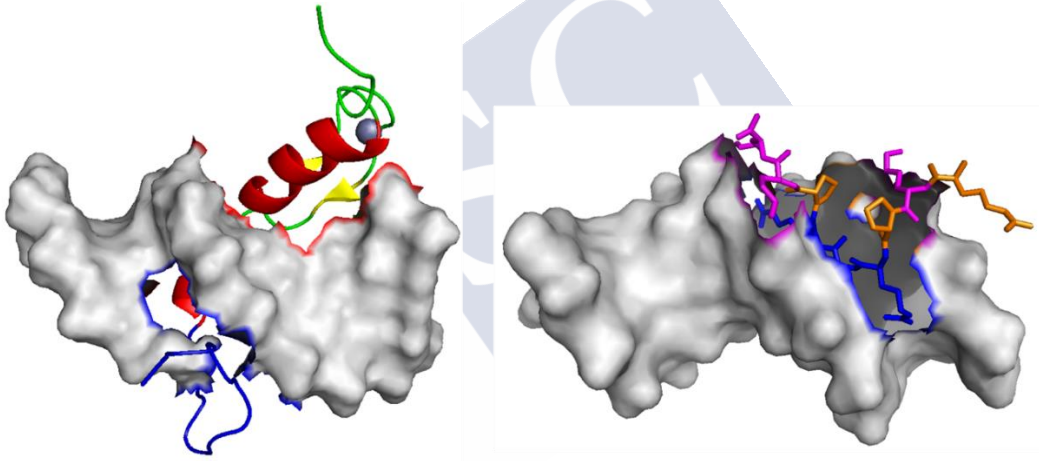


Figure 13. *Left*: Structure of the GAGA factor–DNA complex (PDB 1YUI). *Right*: The complex of the AT-hook bound to DNA. RGR central motif (blue), deeply inserted in the minor groove and lysine residues (pink) establishing additional contacts with DNA (PDB 3UXW).

[49] J. G. Omichinski, P. V. Pedone, G. Felsenfeld, A. M. Gronenborn, G. M. Clore, *Nat. Struct. Biol.* **1997**, *4*, 122-132.

[50] a) L. Aravind, D. Landsman, *Nucleic Acids Res*. **1998**, *26*, 4413-4421; b) R. Reeves, M. S. Nissen, *J. Biol. Chem.* **1990**, 265, 8573-8582.

[51] a) C. C. Cheng, Y. H. Jian, C. J. Lo, J. W. Cheng, *J. Chin. Chem. Soc*. **1998**, *45*, 619-624; b) B. H. Geierstanger, B. F. Volkman, W. Kremer, D. E. Wemmer, *Biochemistry* **1994**, *33*, 5347-5355.

[52] a) J. R. Huth, C. A. Bewley, M. S. Nissen, J. N. Evans, R. Reeves, A. M. Gronenborn, G. M. Clore, *Nat. Struct. Biol.* **1997**, *4*, 657-665; b) E. Fonfría-Subirós, F. Acosta-Reyes, N. Saperas, J. Pous, J. A. Subirana, J. L. Campos, *PLoS One*. **2012**, *7*, e37120.

5. DNA Binding by Synthetic Peptides

It is known that mis-regulation of the TFs activity is at the origin of a number of diseases, and thus, there is growing interest in understanding the molecular basis of specific DNA recognition, as well as in developing designed DNA-binding agents that can modulate or mimic the action of the natural TFs and related proteins. Most artificial DNA-binding agents described to date are small-molecules.

DNA recognition by small molecules is an intensively studied and attractive research field, not only for fundamental reasons, but also because these compounds can present relevant anticancer and/or antimicrobial properties. In fact, some important clinical chemotherapeutic agents are DNA-targeted drugs. Between the classical DNA-binding agents we can highlight cationic species, that establish electrostatic interactions with the DNA phosphate backbone, intercalating agents and small organic molecules that insert into the DNA minor groove.[53]

However, most of these classical DNA-binding agents possess several limitations, such as a relatively low affinity and specificity. Therefore, their use as therapeutic agents in medicine is problematic due to inevitable secondary toxicities. Moreover, many of these molecules present poor stability in physiological media and a limited cellular uptake capacity.

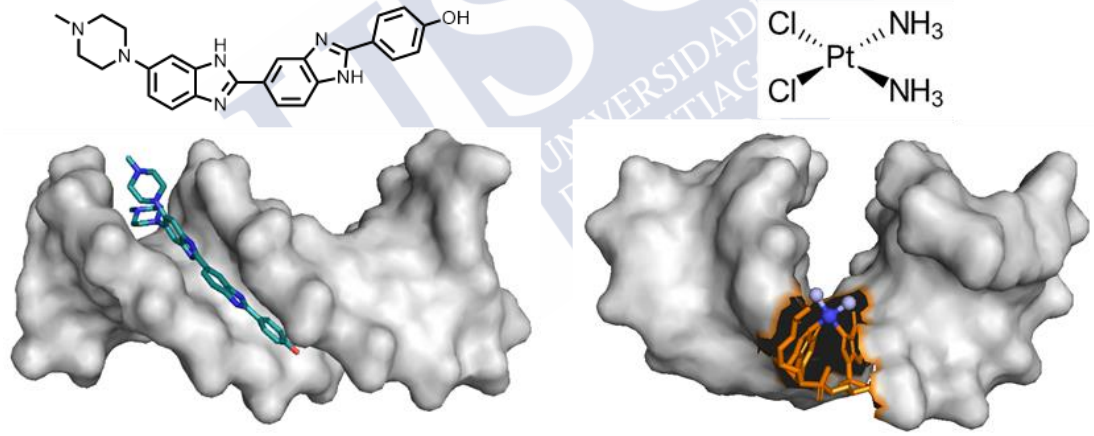


Figure 14. *Left:* Hoechst 33258 structure and its binding to the minor groove of B-DNA (PDB 8BNA). *Right*: Cisplatin´s structure and X-ray crystal structure of DNA containing a cisplatin 1,2-d(GpG) intra-strand cross-link (PDB 3LPV).

In addition to those classic DNA-binding agents, researchers, inspired by natural TFs, have also explored the potential of designed peptides for the specific recognition of DNA, and have developed increasingly sophisticated systems that not only display

[53] a) L. M. Wilhelmsson, P. Lincoln, B. Nordén, *Sequence-specific DNA Binding Agents*, **2006**, RSC Publishing; b) A. Lorente, M. J. Fernández, *An. Quím.* **2008**, *104*, 280-289; c) M. J. Hannon, *Chem. Soc. Rev*. **2007**, *36*, 280-295; d) J. B. Chaires, *Curr. Opin. Struct. Biol.* **1998**, *8*, 314-320; e) S. Neidle, *Nat. Prod. Rep.* **2001**, *18*, 291-309.

excellent binding properties, but are also endowed with new properties not found in their natural counterparts.[54]

These miniature versions of the natural transcription factors may have important future applications in gene-based medicine. The growing amount of information on the human genome, in conjunction with our increased understanding of molecular mechanisms of many major diseases, will unveil a lot of new genetic targets that can be exploited for the control of illnesses. Thus, it will be very important to obtain molecules that can be delivered to selective sites in the genome, and effectively discriminate between closely related DNA sequences.

However, the design of high-affinity DNA binding peptides is not an easy task and indeed the progress in this area has been relatively slow, mainly because of the lack of simple rules for protein-DNA recognition and our poor understanding of protein folding.[55]

It is well established that the DNA-binding domains of transcription factors cannot interact with their consensus DNA sites when isolated from the rest of the protein. Although one might think that the elimination of residues of natural TF located away from the DNA binding domain should not influence the DNA binding, it has been found that these regions usually have a decisive role for the formation of stable DNA complexes.[56] Only in the case of Cys_2-His_2 zinc fingers there is a certain DNA-recognition code that has allowed the rational design of artificial versions that recognize specific DNA sites, albeit obtaining a good binding requires to use multimeric systems.[57]

Therefore, in the absence of tools for efficient de novo design of DNA-binding peptides, the preferred approach has been trial and error studies that take as reference the binding mode of transcription factors. Some of the most successful approaches are described below.

5.1 Artificial Dimerization of bZIP Basic Regions

The structural simplicity of bZIP proteins makes them particularly amenable to the design of simplified non-natural versions. In 1990 the group of Peter Kim, demonstrated that a disulfide bond could effectively substitute the leucine zipper region of the bZIP protein GCN4, without significantly altering the DNA recognition properties of the system. In fact, the disulfide dimer of a truncated basic region of GCN4 TF, was able to recognize the AP1 target sequence (5'-ATGA C TCAT-3') of the natural binding domain with high affinity (nM). However, in contrast to the natural transcription factor that binds to its target DNA at room temperature with high affinity, the dimers only bind at low temperatures (4 °C). This seems to point out that the leucine zipper is performing an additional role to a mere

[54] a) E. Pazos, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *ChemBioChem*. **2011**, *12*, 1958-1973; b) M. E. Vázquez, A. M. Caamaño, J. L. Mascareñas, *Chem. Soc. Rev*. **2003**, *32*, 338-349.

[55] A. Z. Ansari, A. k. Mapp, *Cur. Opin. Chem. Biol.* **2002**, *6*, 765-772.

[56] C. O. Pabo, E. Peisac, R. A. Grant, *Annu. Rev. Biochem*. **2001**, *70*, 313-340.

[57] a) K. J. Brayer, D. Segal, *Cell Biochem. Biophys*. **2008**, *50*, 111-131; b) S. Negi, M. Imanishi, M. Matsumoto, Y. Sugiura, *Chem. Eur. J.* **2008**, *14*, 3236-3249.

dimerization element. As expected, the reduced monomeric peptide did not display significant binding.[58]

This group also identified, two years later, the sequence of the minimal GCN4 basic region fragment that displays specific DNA binding (in the form of a disulfide dimer). Thus, a 23-residue peptide featuring a C-terminal Gly-Gly-Cys linker binds to both ATF/CREB and AP1/GCRE sites in a sequence-specific manner with high affinity in its oxidized form as a dimer (Figure 15a).[59]

Goddard's group reported another example that this sulfur-sulfur dimerization strategy can be extended to make heterodimers or even trimers which recognize the predicted composite DNA sequences.[60] The design consisted of three fragments (each corresponding to the basic region of the v-Jun transcription factor) linked by disulfide bonds through Cys residues at the C and/or N terminus of the recognition helices. Gel retardation experiments showed that the trimeric peptide with these basic regions was able to recognize a 16 base-pair site with a Kd of 5 nM at 4 °C.

After the discovery of Peter Kim, several research groups developed related strategies, showing that the basic regions of GCN4 could be dimerized using different types of connectors (Figure 15). One of the first examples was reported in 1993 by the Schepartz group, where they used coordination complexes to connect the two basic regions of the GCN4 transcription factor. By exploiting the kinetic stability and well-defined geometry of bis(terpyridyl)iron(II) coordination complexes, they were able to assemble isomeric complexes featuring the basic regions in different relative orientations, and to study the influence of geometric changes in the coordination sphere of the DNA-binding affinity and specificity (Figure 15b). Unsurprisingly, they found that dimerization was not enough for DNA binding, and that the geometry of the metal complexes influenced not only the binding affinity of the dimers, but also the sequence selectivity.[61]

[58] R. V. Talanian, C. J. McKnight, P. S. Kim, *Science* **1990**, *249*, 769-771.

[59] R. V. Talanian, C. J. McKnight, R. Rutkowski, P. S. Kim, *Biochemistry* **1992**, *31*, 6871-6875.

[60] C. Park, J. L. Campbell, W. A. Goddard III, *Proc. Nat. Acad. Sci. USA*. **1993**, *90*, 4892-4896; b) C. Park, J. L. Campbell, W. A. Goddard III, *J. Am. Chem. Soc*. **1995**, *117*, 6287-6291.

[61] B. Cuenoud, A. Schepartz, *Science* **1993**, *259*, 510-513.

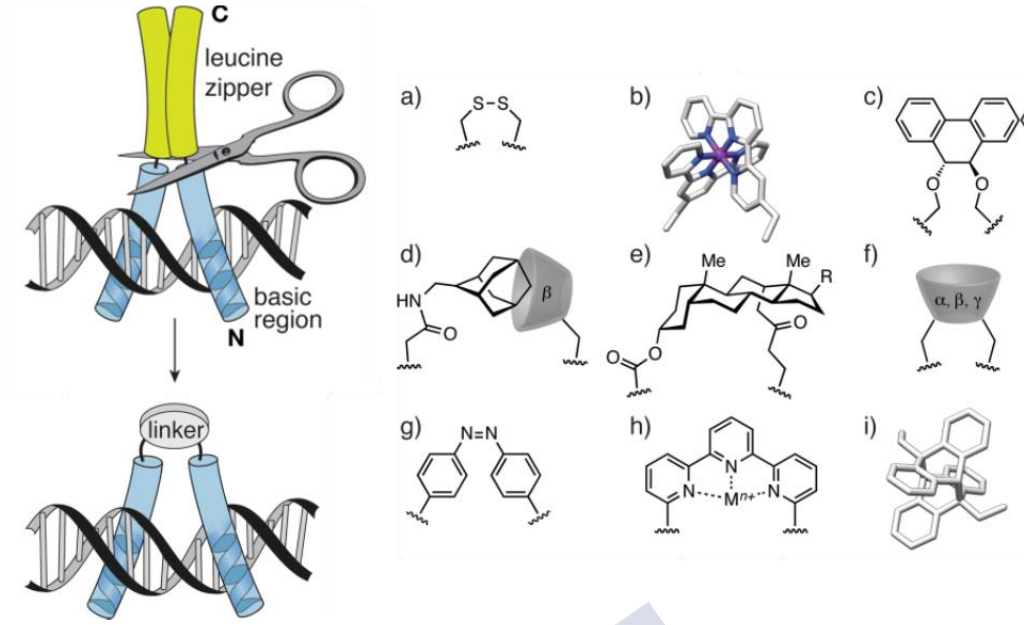


Figure 15. Artificial GCN4 dimers. a) Kim's disulfide dimer. b) Schepartz's transition-metal coordination compound linker. c) Morii's chiral linker derived from trans-9,10-dihydro-phenanthrene-9,10-diol. d) Morii's supramolecular linker based on a β-cyclodextrin–adamantane complex. e) Madder's steroid linker. f) Madder's cyclodextrin dimer. g) Mascareñas' cis/trans azobenzene photoswitch linker. h) Peacock's metal switch linker. i) Peacock's photocontrolled anthracene dimer (Source: adapted from *Eur. J. Org. Chem.* **2018**, *2018*, 249-261).

Following those earlier reports, other synthetic dimers that take advantage of the simplicity of the bZIP architecture have been described. Morii et al. synthesized dimers through a chiral bisphenyl unit, which allowed them to study the influence of the chirality of the connector in the DNA binding (Figure 15c).[62] Moreover, this group was the first to report the use of a noncovalent linker between GCN4 basic regions; peptides derived from the GCN4 basic region and equipped with an adamantyl group at the N-terminus and a β-cyclodextrin at a C-terminal cysteine were able to dimerize and recognize specific DNA sequences (Figure 15d).[63] As in the case of the disulfide dimer, each of the basic regions by itself is not capable of high-affinity DNA recognition, whereas the complex between the b-cyclodextrin and its compound generates a dimer that specifically binds DNA with almost native affinity, although low temperatures (4 °C) are required.

The effect of the geometry in metal complexes was also exploited by Peacock's group, who in 2014 reported a dimer of the GCN4 basic region based on a terpyridine linker. This linker should undergo a conformational change upon metal ion coordination, which in turn affects the relative orientation of the basic regions and influences the DNA binding of the dimer. CD and UV/Vis spectroscopy demonstrated that coordination of this metal-chelating GCN4 dimer to Cu^{II} or Zn^{II} ions favoured DNA binding (Figure 15h).[64]

[62] T. Morii, M. Simomura, S. Morimoto, I. Saito, *J. Am. Chem. Soc.* **1993**, *115*, 1150-1151.

[63] a) M. Ueno, A. Murakami, K. Makino, T. Morii, *J. Am. Chem. Soc.* **1993**, *115*, 12575-12576; b) T. Morii, J. Yamane, Y. Aizawa, K. Makino, Y. Sugiura, *J. Am. Chem. Soc.* **1996**, *118*, 10011-10017.

[64] E. Oheix, A. F. A. Peacock, *Chem. Eur. J.* **2014**, *20*, 2829-2839.

A more recent paper by Madder also involved the use of (α, β, γ)-cyclodextrins for dimerization (Figure 15f). In their case, the cyclodextrins were not used as receptors, but as scaffolds for covalent attachment of the two GCN4 basic regions.[65] Madder's group has also reported the application of steroid scaffolds for the homodimerization of GCN4 basic regions, as well as for the heterodimerization of the basic regions of related c-Myc/Max basic Helix–Loop–Helix transcription factors.[66] The use of a steroidal cholic acid moiety as dimerizer element was justified, because it provides a rigid and synthetically accessible scaffold with which to attach the peptides, with reported benefits for improving the cell uptake of the conjugates (Figure 15e).

In addition to static dimerization elements, our group reported the application of photoswitchable azobenzene units as linkers between the GCN4 basic regions. Our design took advantage of the large conformational change between the cis and trans isomers to modulate the DNA-binding affinity of the synthetic construct (Figure 15g).[67] Peacock´s group also described in 2016, an interesting interplay between the DNA binding of GCN4 peptides and anthracene photodimerization (Figure 15i).[68]

5.2 Monomeric bZIP Basic Regions

Isolated, monomeric alpha-helices of bZIP TFs are not be able to bind their target DNA, in part because of the entropic cost associated with the folding of the peptide chain into the appropriate α-helical conformation. Therefore, several of the strategies described to develop this type of peptide DNA binders are based on stabilizing the α-helical folding, for instance by introducing conformational constraints via stapling techniques.[69] Although the use of this tactic is quite common, and it has been used with different peptides and proteins, such as DNA-binding proteins,[70] only a couple of groups have developed stapled versions of the basic region of GCN4. The first example was described by Taylor and consisted of a constrained GCN4 basic region analogue incorporating two Lys^{i}-Asp^{i+4} side chain lactam bridges. The resulting compound presents a greater helical content than the natural domain, although the authors did not study DNA binding.[71] Later, the same authors reported other analogues containing a single lactam bridge, and confirmed an increase in the affinity compared to that of the natural basic region, although with low DNA affinities (K_D 0.65 μM).[72]

---

[65] Y. Ruiz García, J. Zelenka, Y. V. Pabon, A. Iyer, M. Buděšínský, T. Kraus, C. I. E. Smith, A. Madder, *Org. Biomol. Chem.* **2015**, *13*, 5273-5278.

[66] a) L. L. G. Carrette, T. Morii, A. Madder, *Eur. J. Org. Chem*. **2014**, *14*, 2883-2891; b) Y. R. García, Y. V. Pabon-Martínez, C. I. E. Smith, A. Madder, *Chem. Commun.* **2017**, *53*, 6653-6656.

[67] A. M. Caamaño, M. E. Vázquez, J. Martínez-Costas, L. Castedo, J. L. Mascareñas*, Angew. Chem. Int. Ed.* **2000**, *39*, 3104-3107; *Angew. Chem*. **2000**, *112*, 3234-3237.

[68] G. A. Bullen, J. H. R. Tucker, A. F. A. Peacock, *Chem. Commun*. **2015**, *51*, 8130-8133.

[69] a) A. D. De Araujo, H. N. Hoang, W. M. Kok, F. Diness, P. Gupta, T. A. Hill, R. W. Driver, D. A. Price, S. Liras, D. P. Fairlie, *Angew. Chem. Int. Ed.* **2014**, *53*, 6965-6969; b) Y. H. Lau, P. De Andrade, Y. Wu, D. R. Spring, *Chem. Soc. Rev.* **2015**, *44*, 91-102.

[70] M. Kajino, K. Fujimoto, M. Inouye, *J. Am. Chem. Soc.* **2011**, *133*, 656-659.

[71] M. Zhang, B. Wu, J. Baum and J. W. Taylor, *J. Peptide Res.* **2000**, *55*, 398-408.

[72] M. Zhang, B. Wu, H. Zhao and J. W. Taylor, *J. Peptide Sci.* **2002**, *8*, 125-136.

Not long so, Madder´s group modified the DNA recognition region of the GCN4 to contain two optimally positioned cysteines which were stapled with bis electrophilic reagents. These resulted peptides exhibited an α-helical conformation which improved the DNA binding and the cellular uptake,[73] albeit the affinity constants were not presented.

Schepartz group described an alternative strategy for stabilizing the α-helix of the GCN4 basic region which is based in *residue grafting* on a pre-folded alpha helical scaffold. The tactic consisted in selecting the GGN4 residues that contact DNA and grafting them into the α-helix of the avian pancreatic polypeptide (aPP), a folded peptide consisting of a single α-helix stabilized by hydrophobic interactions with a type II polyproline helix (Figure 16). Using this strategy, they were able to construct a 42 amino acid peptide that exhibited extremely tight DNA affinity and selectivity: association constants of 1.5 nM at 4 °C.[74] Further refinements of this initial binder with use of phage libraries, resulted in a new peptide with extraordinary DNA affinity at 4 °C (K_D ≈ 23 pM), which even retained a high affinity at 25 °C (K_D ≈ 1.6 nM),[75] exhibiting, therefore, the same range of affinities as natural transcription factors but without the need to dimerize.

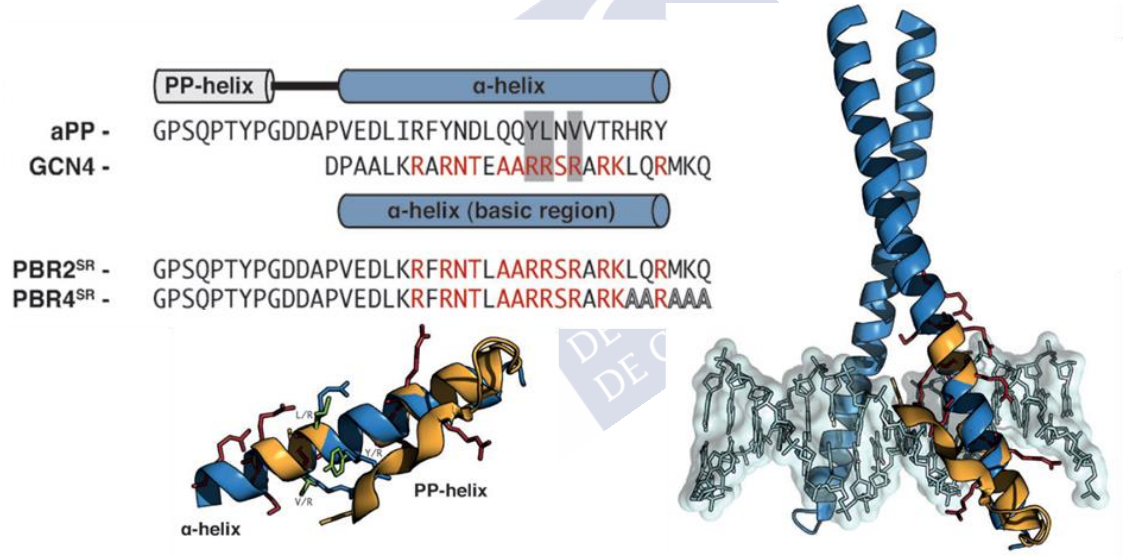


Figure 16. *Left top:* Sequence alignment of aPP and the CGN4 basic region. *Left bottom*: Structure of the aPP peptide (orange) and the basic region (blue) aligned, showing conflicting side chains (green and blue). *Right:* Structure of the DNA-binding domain of the CGN4 transcription factor showing the key side chains as red sticks (PDB 1YSA). Superimposed to the GCN4 basic region is the aPP miniprotein (orange, PDB 1PPT) (Source: adapted from[54a]).

In a related approach, Morii used a small compact domain of the F-actin bundling protein villin as a tertiary structure scaffold to graft the residues used by GCN4, necessary for

[73] A. Iyer, D. Van Lysebetten, Y. Ruiz García, B. Louage, B. G. De Geest, A. Madder, *Org. Biomol. Chem.* **2015**, *13*, 3856-3862.

[74] N. J. Zondlo, A. Schepartz, *J. Am. Chem. Soc.* **1999**, *121*, 6938-6939.

[75] J. W. Chin, A. Schepartz, *J. Am. Chem. Soc.* **2001**, *123*, 2929-2930.

DNA binding. Unfortunately, the monomeric construct did not show better DNA binding than the basic-region helix itself.[76]

Overall, there have been very important advances, but the DNA binding of short (less than 30 aa) monomeric peptide fragments made only from natural amino acids has not been achieved.

5.3 Conjugates between Peptides and Small Molecules

DNA recognition is often a cooperative process involving the formation of multimeric complexes of transcription factors that cooperatively bind to neighbour sequences in the DNA. Several groups have investigated whether combining DNA-binding domains (or their fragments) with other synthetic binders could lead to conjugates exhibiting good affinity and selectivity. Therefore, in 1990s, Barton described conjugates between short 13-mer peptides (derived from α_3-helix of the phage P22 repressor) and photoactive $[Rh(phi)_2(phen')]^{3+}$ complexes.[77] The coordination compounds intercalate through the DNA major groove with high affinity and upon photoactivation, induce DNA strand scission (Figure 17). Later examples demonstrated the general applicability of this strategy using different peptides with alternative sequence preferences.[78]

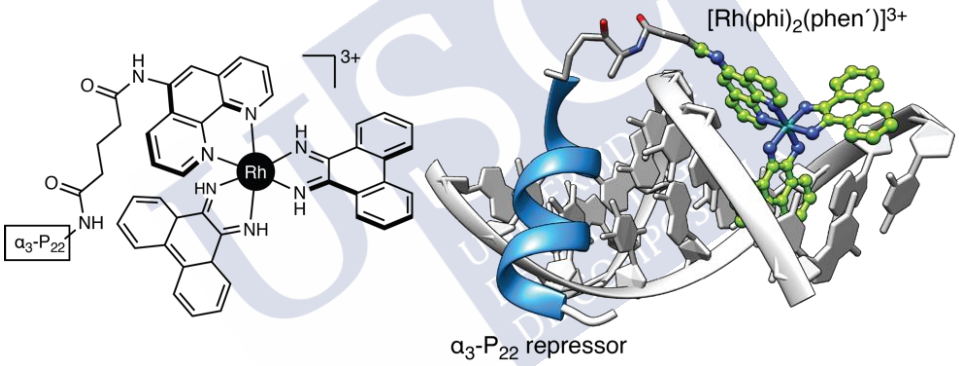


Figure 17. *Left:* Structure of the $[Rh(ph)_2(phen')]^{3+}$-α_3-P22 conjugate. *Right*: Model of the DNA interaction of the conjugate. The complex is shown in green, and the α_3-helix in blue.

In 1995 the Verdine's group demonstrated that it was possible to drive an specific DNA recognition by GCN4 basic regions if the peptide was delivered into the major groove in an intramolecular fashion.[79] Attaching the GCN4 BR to the DNA significantly increased its effective concentration, and reduced the entropic loss associated with intermolecular binding, thereby allowing the formation of the desired (intramolecular) complex. The design required a Gly–Gly–Cys linker to connect the C-terminus of the GCN4 br to an adenine nucleotide located next to the 5′ end of the AP1/GCRE half-site [5′-A(**c**)TCAT-3′, modified base in bold] through a disulfide bond.

[76] T. Morii, S. I. Sato, M. Hagihara, Y. Mori, K. Imoto, K. Makino, *Biochemistry* **2002**, *41*, 2177-2183.
[77] N. Y. Sardesai, K. Zimmermann, J. K. Barton, *J. Am. Chem. Soc.* **1994**, *116*, 7502-7508.
[78] N. Y. Sardesai, S. C. Lin, K. Zimmerman, J. K. Barton, *Bioconjug. Chem.* **1995**, *6*, 302-312.
[79] D. Stanojevic, G. L. Verdine, *Nat. Struct. Biol.* **1995**, *2*, 450-457.

Taking the concept of intramolecular delivery for specific DNA binding one step further, in 2001 our group reported the synthesis of conjugates between the GCN4 BR and a small minor-groove binding molecule (distamycin derivate) that was capable of sequence-specific DNA recognition.[80] While monomeric GCN4 basic regions are unable to display significant DNA-binding affinity, the conjugates are able to recognize specific DNA sequences that feature the AT-recognition site of the minor-groove binder adjacent to the GCN4 recognition site. The minor groove binder acts as an anchor that assists the delivery of the basic region into the major groove (Figure 18a).

After that initial report, the group demonstrated that the conjugation strategy could be generally applied for promoting the DNA binding of other transcription factor fragments that by themselves are not capable of interacting with their target sites. For instance, it is known that a 61-residue minimized version of the GAGA transcription factor, containing only the zinc finger recognition unit and the N-terminal basic regions can achieve the DNA recognition with decent affinity.[81] However, removal of 27 residues from the basic region led to suppression of the DNA affinity. Our group demonstrated that tethering a distamycin derivative to only the zinc finger region it is possible to interact with good affinity and selectivity with the composite sequence 5′-TTTT-GAGAG-3′ containing an A/T-rich site for distamycin binding (5′-TTTT-3′) next to the GAGA TF consensus binding sequence (5′-GAGAG-3′)[82] (Figure 18b).

Similar conjugates with GCN4 and GAGA fragments (Figure 18c-d),[83] and even with homeodomain fragments, based on the use of bis(benzamidines) as minor-groove binding agents, were also described (Figure 18e).[84] These results confirm the modular nature and general applicability of this conjugation strategy for obtaining sequence-selective DNA-binding peptides derived from transcription factors.

[80] M. E. Vázquez, A. M. Caamaño, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2001**, *40*, 4723-4725; *Angew. Chem*. **2001**, *113*, 4859-4861.

[81] P. V. Pedone, R. Ghirlando, G. M. Clore, A. M. Gronenborn, G. Felsenfeld, J. G. Omichinski, *Proc. Nat. Acad. Sci. USA.* **1996**, *93*, 2822-2826.

[82] O. Vázquez, M. E. Vázquez, J. B. Blanco, L. Castedo, J. L. Mascareñas*, Angew. Chem. Int. Ed*. **2007**, *46*, 6886-6890; *Angew. Chem.* **2007**, *119*, 7010-7014.

[83] M. I. Sánchez, O. Vázquez, M. E. Vázquez, J. L. Mascareñas, *Chem. Eur. J.* **2013**, *19*, 9923-9929.

[84] J. Mosquera, J. Rodriguez, M. E. Vazquez, J. L. Mascareñas, *ChemBioChem.* **2014**, *15*, 1092–1095.

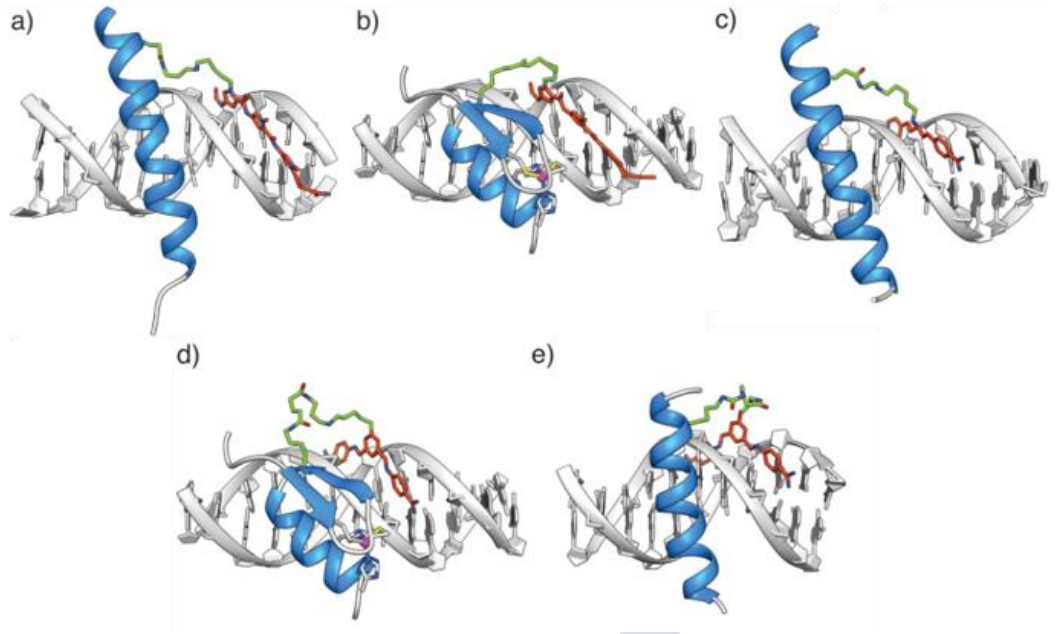


Figure 18. TF fragments stabilized by small minor-groove binding agents. a) Original GCN4 basic region conjugated with distamycin. b) GAGA-distamycin conjugate. c) GCN4-bis(benzamidine) hybrid. d) GAGA-bis(benzamidine) hybrid. f) Engrailed homeodomain–bis(benzamidine) conjugate. (Source: adapted from *Eur. J. Org. Chem*. **2018**, *2018*, 249-261).

This conjugation strategy does not need to be restricted to the small DNA binding molecules like those shown above. Indeed, our group lately informed the use of an AT-hooks as a minor-groove binding anchors, either to make conjugates with bZIP basic regions or with monomeric zinc finger modules.[85] The resulting peptide conjugates displayed very good DNA affinities and excellent sequence selectivity. Furthermore, the peptidic nature of the AT-hook allowed the straightforward synthesis of fully peptide chimeras using solid-phase peptide synthesis methods.

An example recently published by Roy's group in 2017, further demonstrates that the combination of short minor-groove anchors and TF fragments constitutes a robust approach to developing sequence-specific DNA-binding peptides. In their work, Roy's group used a short minor-groove binding tail from the Serum Response Factor (SRF) in combination with the DNA-binding helix of Elk1 to yield a dominant negative of the oncogene cFos.[86]

All these examples of binary artificial constructs allow the specific recognition of relatively short stretches of DNA (up to of 9 base pairs), far from the typical extended DNA sites covered by natural counterparts. Longer DNA sites had been successfully targeted by zinc fingers proteins in the form of recombinant oligomers, not as synthetic peptide constructs.[87]

[85] J. Rodríguez, J. Mosquera, J. R. Couceiro, M. E. Vázquez, J. L. Mascareñas, *Chem. Sci.* **2015**, *6*, 4767-4771.
[86] M. Chakraborty, S. Roy, *Chem. Commun.* **2017**, *53*, 376-379.
[87] C. A. Gersbach, T. Gaj, C. F. Barbas, *Acc. Chem. Res.* **2014**, *47*, 2309-2318.

Our group has developed the first engineered synthetic DNA-binding motif consisting of two different fragments of natural TFs (GCN4 and GAGA), connected via an AT-hook linker, capable to selectively recognize extended DNA sequences (up to 13 bp). The AT-hook does not only passively connect both TF fragment, but also provides stabilizing contacts with the DNA in the minor groove, while ensuring the correct delivery of the peptides to their consensus sites, into two consecutive major grooves.[88]

Figure 19. Model obtained using MM calculations of the interaction between the tripartite construct and the target composite DNA sequence. (Source: adapted from[88])

Biological processes, and in particular gene activation processes, are strictly regulated, and in many cases, they only take place in response to certain cellular signals. In this context, there is a great interest in designing synthetic molecules whose interaction with the DNA can be regulated at will in response to external stimuli.

With this objective, our group demonstrated in 2014 the nickel-promoted assembly of the peptide derived from GCN4 TF (with two histidines) and a bisbenzamidine equipped with a bipyridine unit. Key for the success of the approach is the dual role of the metal as an α-helix-nucleating factor and as a heterodimerization staple (Figure 20, left). The multicomponent nature of the system and the kinetic lability of the metal coordination facilitated the disassembly of the supramolecular structure upon addition of external agents that sequester the nickel cation.[89] A follow-up paper in 2016 applied this same strategy for the construction of a nickel-mediated trimeric DNA-binding metallopeptide composed of a basic region/AT-hook chimera equipped with a bipyridine ligand and the peptide of basic region modified with two histidines.[90] The metallopeptide complex assembled in the presence of Ni(II) salts, was capable of self-assembling at specific DNA sequences of twelve base pairs (Figure 20, right).

[88] J. Rodríguez, J. Mosquera, R. García-Fandiño, M. Eugenio Vázquez, J. L. Mascareñas, *Chem. Sci.* **2016**, *7*, 3298-3303.

[89] M. I. Sánchez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2014**, *53*, 9917-9921.

[90] J. Rodríguez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *Chem. Eur. J.* **2016**, *22*, 13474-13477.

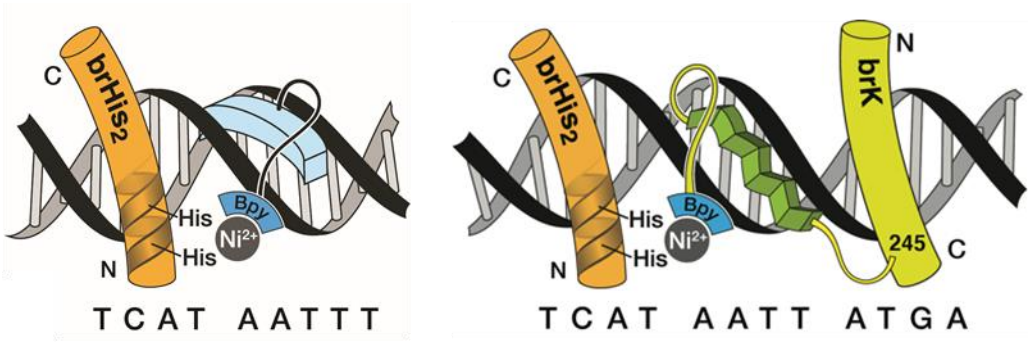


Figure 20. *Left:* Scheme of the metal-induced cooperative assembly of dimeric DNA-binding metallopeptides.[89] *Right:* Scheme of the metal-trimeric DNA binding.[90]

All these advances suggest that playing with synthetic and metal coordination chemistry, it is possible to bind DNA at specific sites using modular strategies. This is especially interesting because of the ensuing versatility and the ready availability of the synthetic modules.

However, there are yet many challenges ahead, such as demonstrating the generality of the non-covalent assembly strategies, the development of easy to build conjugates, the recognition of long sites or the implementation of more practical switches, among others.

General Objectives

Considering all precedents and the current challenges in the field of DNA recognition, part of this doctoral studies aim at the design and synthesis of a new generation of DNA-binding peptides and miniproteins, capable of recognizing designed DNA sites with high affinity and selectivity. Specifically, we have focused on:

1) Targeting long DNA sites with fully peptidic platforms: Our previous work has shown that it is possible to recognize tripartite DNA sites combining a minor groove binder and two transcription factor fragments (Figure 19).[88] Despite the strategy was successful, the synthesis of the conjugates is cumbersome and long, requiring the use of orthogonal protecting groups and the introduction of elaborate synthetic linkers. Moreover, the non-peptidic nature of these hybrid systems hampers the future possibility of biological engineering and genetic encoding.
These limitations have raised the question of whether it would be possible to assemble analogues of this multipartite DNA binders relying exclusively on natural amino acids and peptide linkers. We have approached this goal, and the results are discussed in the section 1 of this chapter.

2) As we described before, our group has explored the application of metal coordination as a potentially powerful approach to obtaining multimeric DNA binders (Figure 20). This is an excellent strategy, owing to its modularity, and therefore there is a great interest in demonstrating that it can be further expanded. In the section 2 of this chapter, we describe our attempts to bind DNA using tripartite systems in which the modules are all connected via metal coordination strategies.

3) Switchable DNA binding by isolated b-ZIP basic regions with native amino acids: The development of stimuli-responsible sequence-specific DNA recognition agents is a topic of fundamental scientific interest with potential far-reaching impact in biomedical sciences. We have mentioned before that monomeric GCN4 peptides display very low DNA-binding affinity, probably because the high entropic cost associated with the folding of the peptide chain into the α-helical conformation. In the section 3, we describe our work towards the design of monomeric bZIP basic regions whose interaction with DNA can be controlled by metals.[91]

[91] In order to facilitate reading and improve understanding, the necessary and corresponding references will be added to each section (some of them may be previously referenced in the general introduction of the chapter).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Section 1: DNA-Binding Miniproteins Based on Zinc Fingers

This chapter includes work published in *Chemical Science* as:

J. Rodríguez, S. Learte-Aymamí, J. Mosquera, G. Celaya, D. Rodríguez-Larrea, M. E. Vázquez, J. L. Mascareñas, *Chem. Sci.* **2018**, *9*, 4118-4123.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Introduction

Zinc fingers (ZFs), the most abundant eukaryotic transcription factors (TFs), regulate the expression of multiple genes involved in cell development, differentiation and tumor suppression.[92] Cys_2–His_2 ZFs are composed of small peptide domains, of about 30 amino acids, which fold into a simple ββα-fold, stabilized by hydrophobic interactions and by chelation of Zn(II) ions by Cys and His residues in their sequence. Cys_2–His_2 ZFs are a versatile platform for engineering genetically encoded transcriptional regulators and gene editing tools, some of which have even reached clinical trials.[93]

ZFs do not interact with their DNA target sites as isolated monomers, so specific binding to DNA typically requires the **cooperative interaction** of at least three zinc finger domains, that wind around the DNA while inserting their recognition α-helices in the major groove, each of them specifically recognizing three base pairs.[94] Despite the prevalence of oligomeric ZFs, there are also some examples of **monomeric** DNA binding ZFs, such as the prokaryotic **GAGA** protein which does not require tandem zinc finger repeats. This factor is 519 residue-long, comprises a single classical Cys_2-His_2 zinc finger core and an N-terminal extension containing two basic regions. Previously, it has been demonstrated that the 61-residue minimized version of the GAGA transcription factor, containing only the zinc finger recognition unit and the N-terminal basic regions was able to achieve the DNA recognition.[95] Nevertheless, removal of 27 residues from the basic region led to suppression of the DNA affinity.

However, even though there is high interest in these recombinant constructs, the use of only one class of DNA binding motifs limits the modes of interaction that can be achieved. Therefore, it would be of great value to generate alternative DNA binding agents that can combine different type of DNA interacting units.[96,97]

As mentioned before, our group has demonstrated that, while isolated peptides derived from the GAGA Cys_2–His_2 ZF fail to bind DNA, their covalent tethering to minor groove

[92] C. O. Pabo, E. Peisach, R. A. Grant, *Annu. Rev. Biochem.* **2001**, *70*, 313-340.

[93] a) A. Klug, *Annu. Rev. Biochem.* **2010**, *79*, 213-231; b) F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, *Nat. Rev. Genet.* **2010**, *11*, 636-646; c) S. A. Wolfe, L. Nekludova, C. O. Pabo, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183-212; d) C. A. Gersbach, T. Gaj, C. F. Barbas, *Acc. Chem. Res.* **2014**, *47*, 2309-2318.

[94] M. Suzuki, S. E. Brenner, M. Gerstein, N. Yagi, *Protein Eng.* **1995**, *8*, 319-28.

[95] P. V. Pedone, R. Ghirlando, G. M. Clore, A. M. Gronenborn, G. Felsenfeld, J. G. Omichinski, *Proc. Nat. Acad. Sci. USA.* **1996**, *93*, 2822-2826.

[96] a) P. B. Dervan, *Bioorg. Med. Chem*. **2001**, *9*, 2215-2235; b) A. Paul, A. Kumar, R. Nanjunda, A. A. Farahat, D. W. Boykin, W. D. Wilson, *Org. Biomol. Chem*. **2017**, *15*, 827-835; c) P. Guo, A. Paul, A. Kumar, N. K. Harika, S. Wang, A. A. Farahat, D. W. Boykin, W. D. Wilson, *Chem. Commun*. **2017**, *53*, 10406-10409; d) S. Boga, D. Bouzada, D. García Peña, M. Vázquez López, M. E. Vázquez, *Eur. J. Org. Chem*. **2018**, *2018*, 249-261.

[97] M. E. Vázquez, A. M. Caamaño, J. L. Mascareñas, *Chem. Soc. Rev*. **2003**, *32*, 338-349.

binders such as polypyrroles,[98] bisbenzamidines[99] or AT-hook peptide domains[100] restores their DNA binding.

Unfortunately, the synthesis of these conjugates is far from trivial, requiring the use of orthogonal protecting groups and the introduction of elaborate synthetic linkers. Moreover, their non-peptidic nature prevents the future possibility of biological engineering and genetic encoding. These limitations have raised the question of whether it would be possible to assemble analogues of this multipartite DNA binders relying exclusively on **natural amino acids** and peptide linkers.

[98] O. Vázquez, M. E. Vázquez, J. B. Blanco, L. Castedo, J. L. Mascareñas*, Angew. Chem. Int. Ed*. **2007**, *46*, 6886-6890; *Angew. Chem.* **2007**, *119*, 7010-7014.

[99] a) M. I. Sánchez, O. Vázquez, M. E. Vázquez, J. L. Mascareñas, *Chem. Eur. J*, **2013**, *19*, 9923-9929; b) J. Rodríguez, J. Mosquera, O. Vázquez, M. E. Vázquez, J. L. Mascareñas *Chem. Commun.* **2014**, *50*, 2258-2260.

[100] a) J. Rodríguez, J. Mosquera, J. R. Couceiro, M. E. Vázquez, J. L. Mascareñas *Chem. Sci*., **2015**, *6*, 4767-4771; b) J. Rodríguez, J. Mosquera, R. García-Fandiño, M. E. Vázquez, J. L. Mascareñas *Chem. Sci*., **2016***, 7*, 3298-3203.

Objective

We propose the synthesis of fully peptidic, ZF-based miniproteins, capable of interacting with specific DNA sites. In contrast to classical ZFs, which only establish contact with the DNA major groove, our designed constructs will combine interactions in the major and the minor grooves. The synthesis of these systems should be feasible in one shot, by using a solid synthesis protocol and without the need for orthogonally-protected amino acids to achieve the required chemoselective conjugations. Importantly, this system might be expressed *in vivo*, and therefore could be generated directly inside the cell by designing appropriate transcriptional plasmids. Furthermore, we suggest for the first time the use of single-molecule nanopore techniques for characterizing the interaction between synthetic peptides and DNA.

Results and discussion

Design and synthesis

The newly designed chimeras are composed by one AT-hook sequence connected to two Cys_2-His_2 replicas of the DNA-binding domain of the GAGA TF (Ser^{28} to Phe^{58} in the reference PDB structure).[101] Remarkably, neither of the components is capable of interacting with their respective DNA sites with appreciable affinity as isolated monomers.[102] Taking as starting points the experimental structures of the DNA complexes of GAGA and the third AT-hook of HMG-I(Y),[103] we built a hypothetical model for a simultaneous interaction of the AT-hook motif inserted into a central AATT minor groove site, flanked by two Cys_2-His_2 GAGA fragments bound to adjacent major grooves.

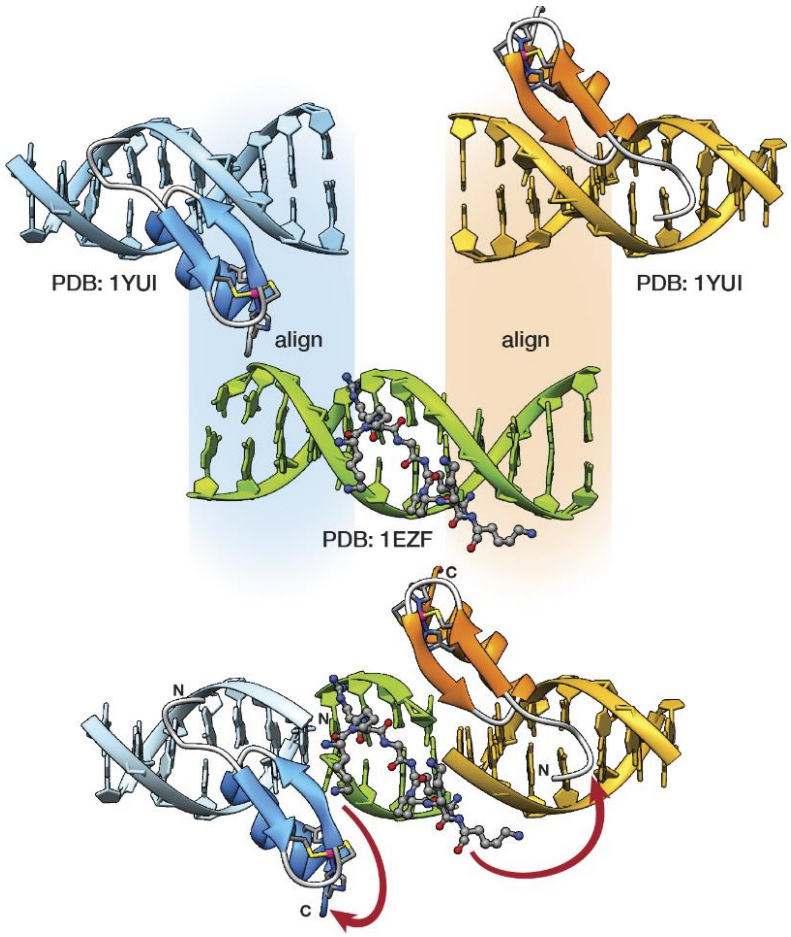


Figure 21. Outline of the structured-guided design consisting of the superposition of the DNA chains of the structures involved in the chimera, followed by clean-up of the overlapping DNA strands and indication of the conjugation scheme between the different DNA binding modules.

[101] J. G. Omichinski, P. V. Pedone, G. Felsenfeld, A. M. Gronenborn, G. M. Clore, *Nat. Struct. Biol.* **1997**, *4*, 122-132 (PDB ID: 1YUI).

[102] This 31-residue GAGA fragment, lacking the accessory basic regions that stabilize its complex with the DNA, does not show any significant DNA affinity. See ref 95.

[103] J. R. Huth, C. A. Bewley, M. S. Nissen, J. N. Evans, R. Reeves, A. M. Gronenborn, G. M. Clore, *Nat. Struct. Biol.* **1997**, *4*, 657-665 (PDB ID: 2EZF).

Inspection of this qualitative model suggested that a Gly_4 linker between the C-terminal end of the Cys_2-His_2 GAGA fragment and the N-terminal arm of the AT-hook might span the required distance. This model also revealed a potentially damaging steric clash involving the side chains in the β-hairpin of the second Cys_2-His_2 GAGA domain with the C-terminal Lys^{40} of the AT-hook, said Lys was therefore replaced by a glycine (**Hk^G**). In order to maintain the total positive charge of the AT-hook in the final conjugate and favor electrostatic contacts with the phosphate backbone, we introduced a lysine residue in the linker connecting the C-terminus of the AT-hook peptide and the N-terminal side of the GAGA fragment.

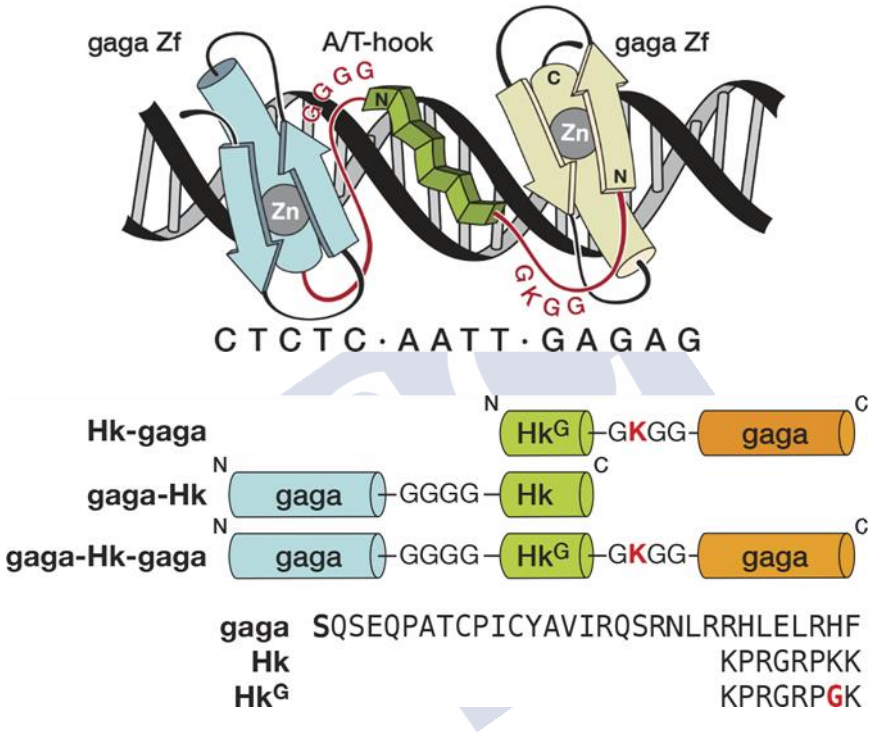


Figure 22. *Top:* Schematic illustration of the hypothetical tripartite major-minor-major groove recognition by a GAGA/AT-hook/GAGA chimera. *Bottom:* Schematic representation of the synthetic hybrids and sequences of the modules. Note: Molecular modelling considerations suggested that in the case of the C-terminal GAGA domains (in orange) is better to skip the N-terminal Ser residue (bolded in the sequence), in order to direct the linker towards the C-terminus of the AT-hook.

We validated our approach by designing three chimeras, **Hk-gaga**, **gaga-Hk**, and **gaga-Hk-gaga**, which were synthesized in good yields following standard Fmoc/tBu SPPS protocols.[104] It is important to note that whereas in the first hybrid, **Hk-gaga**, the connection involves the N-terminus of the zinc finger and the C-terminus of the AT-Hook, in **gaga-Hk** the linkage is between the AT-hook N-terminal side and the zinc finger C-terminus. In all peptides the C-terminus is amidated due to *Rink amide* resin used, and the N-terminal side does not have any modification. Remarkably, these synthetic procedures are straightforward, and the peptides can be assembled using an automatic synthesizer in just one working day (per peptide), which is an important advantage with

[104] I. Coin, M. Beyermann, M. Bienert, *Nat. Protoc*. **2007**, *2*, 3247-3256.

respect to previous approaches to conjugates containing non-peptide linkers and binders, as explained in the introduction.

Studies of the conjugate **gaga-Hk**

Having at hand the desired bivalent conjugates, we studied their DNA binding properties using non-denaturing electrophoresis mobility shift assays (EMSA) in polyacrylamide gels.[105] Thus, a double stranded (ds) oligonucleotide ***AT•GAG*** containing the AT-hook and GAGA binding sites in tandem, was mixed with increasing concentrations of the conjugate **gaga-Hk**. The gel showed concentration-dependent slow-migrating bands, which are consistent with the formation of the desired **gaga-Hk/AT•GAG** complex (Figure 23, Panel a). Importantly, no new bands were observed when the conjugate **gaga-Hk** was incubated with a dsDNA lacking the GAGA binding site (Figure 23, Panel c), demonstrating that the ZF peptide must be bound to its target site for the obtention of high-affinity binding. Curiously, incubation with a control oligonucleotide lacking the A/T-rich site also led to retarded bands, albeit in this case the interaction appears to be weaker (Figure 23, Panel b and lane 5 in panel d). As expected, in absence of the AT-hook unit, the zinc finger module of GAGA (**gaga**) by itself, does not give rise to slow-migrating bands, neither with ***AT•GAG*** or with ***cg•GAG*** (Figure 23, Panel d, lanes 3 and 6).

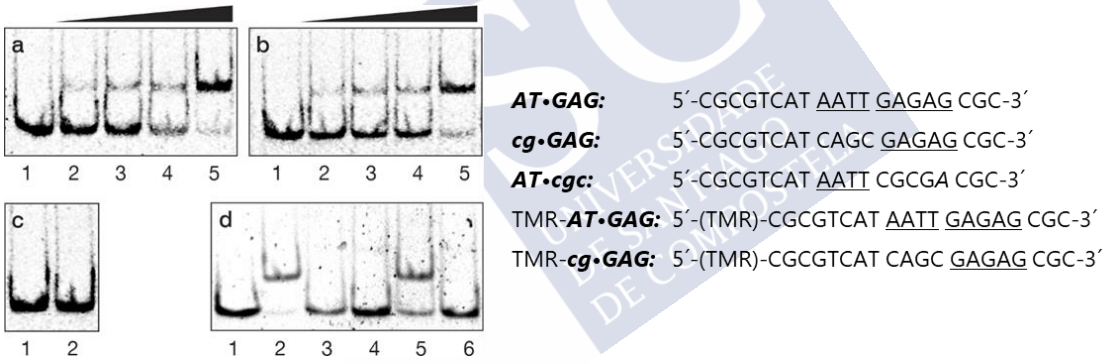


Figure 23. *Left***:** DNA binding of **gaga-Hk** by polyacrylamide gel electrophoresis. Panel a, lanes 1-5: [**gaga-Hk**] = 0, 300, 500, 700, 1000 nM with 75 nM of the ds-oligonucleotide ***AT•GAG***. Panel b, lanes 1-5: [**gaga-Hk**] = 0, 300, 500, 700, 1000 nM with 75 nM of DNA ***cg•GAG***. Panel c, lanes 1-2: [**gaga-Hk**] = 700, 1000 nM with 75 nM of the DNA ***AT•cgc***. Panel d, lanes 1-3: 75 nM of ***AT•GAG***; lane 2: 1000 nM of **gaga-Hk** and lane 3: 1000 nM of **gaga;** lanes 4-6: 75 nM of ***cg•GAG;*** lane 5: 1000 nM of **gaga-Hk** and lane 6: 1000 nM of **gaga.** *Right*: Oligonucleotide sequences (only one strand shown, consensus sites underlined). Incubations were carried out in 18 mM Tris-HCl buffer (pH 7.5), 90 mM KCl, 1.8 mM $MgCl_2$, 0.2 mM TCEP, 9% glycerol, 0.11 mg/mL BSA, 2.2% NP-40 and 0.02 mM of $ZnCl_2$. After 30 min at 20 ºC, products were resolved by PAGE on a 10% non-denaturing polyacrylamide gel and 0.5× TBE buffer over 40 min at 20 °C and analyzed by staining with SyBrGold.

Fluorescence anisotropy titrations with a rhodamine (TMR)-labeled dsDNA containing the target consensus sequence (AATT-GAGAG), confirmed that **gaga-Hk** binds with high affinity to its target site, with an apparent K_D of 58 ± 4 nM in the presence of competing

[105] a) L. M. Hellman, M. G. Fried, *Nat. Protoc***. 2007**, 2, 1849-1861; b) D. Lane, P. Prentki, M. Chandler, *Microbiol. Rev*. **1992**, *56*, 509-528; c) J. B. Blanco, M. E. Vázquez, J. Martinez-Costas, L. Castedo, J. L. Mascareñas, *Chem. Biol.* **2003**, *10*, 713-722.

calf thymus DNA (41 ± 7 nM in the absence of calf thymus) at 20 °C (Figure 24, left). Highly, fluorescence anisotropy titrations showed that in presence of excess of competing calf thymus DNA, **gaga-Hk** binds very weakly to the mutated dsDNA lacking the A/T-rich tract (Figure 24, right). This result indicates that the retarded band observed in the EMSA with this mutated DNA (Figure 23b) arises from relatively weak and less specific interactions in which the AT-hook peptide is most probably not inserted in the minor groove, but making electrostatic contacts with the DNA backbone.[106] Taken together, these results support the formation of a cooperative, bivalent DNA binding complex at specific composite DNA sites of nine base pairs (AATT-GAGAG), in which the GAGA peptide fragment binds in the major groove of its target sequence (GAGAG) while the AT-hook peptide is inserted in the adjacent minor groove (AATT).

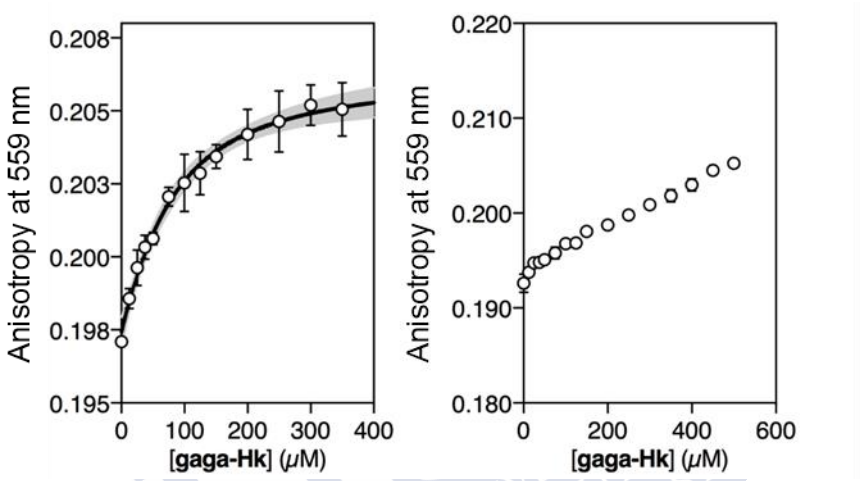


Figure 24. *Left:* Fluorescence anisotropy titration of a 25 nM solution of TMR-***AT•GAG*** in the presence of competing non-specific calf-thymus DNA (50 μM in base pairs) and with increasing concentrations of **gaga-Hk**. The best fit to a 1:1 binding model is also shown, including 95% confidence of the fit shown in light grey. *Right:* Fluorescence anisotropy titration of a 25 nM solution of TMR-***cg•GAG*** in the presence of competing non-specific calf-thymus DNA (50 μM in base pairs) and with increasing concentrations of **gaga-Hk**. Data correspond to the mean of three independent experiments.

Studies of the conjugate **Hk-gaga**

The inverted chimera **Hk-gaga** also targets the same composite DNA site. Thus, addition of increasing amounts of **Hk-gaga** to the dsDNA ***AT•GAG*** under standard conditions led to a new, slow-migrating band (Figure 25, Panel a). This new band is consistent with the expected specific peptide-DNA complex. As previously observed for the conjugate **gaga-Hk**, **Hk-gaga** does not elicit new retarded bands when incubated with a non-target sequence lacking the GAGAG site (Figure 25, Panel b), and shows residual binding with a control oligonucleotide featuring the GAGAG site but lacking the A/T-rich site (Figure 25,

[106] It must be noted that purely electrostatic and disordered complexes can also display high affinity: A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R. B. Best, B. Schuler, *Nature*. **2018**, *555*, 61–66.

Panel c). Therefore, the inverse arrangement of DNA binding moieties allowed an excellent selectivity.

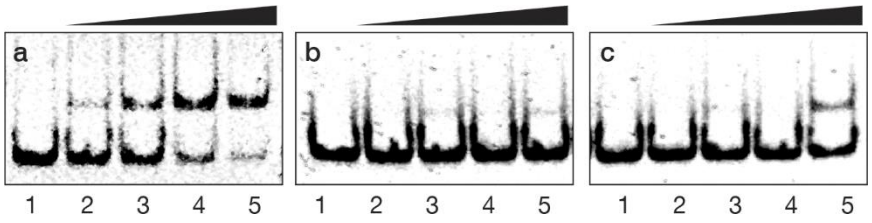


Figure 25. EMSA results for **Hk-gaga**. Panel a, lanes 1-5: [**Hk-gaga**] = 0, 300, 500, 700, 1000 nM with 75 nM of DNA ***AT•GAG***. Panel b, lanes 1-5: [**Hk-gaga**] = 0, 300, 500, 700, 1000 nM with 75 nM of ***AT•cgc***. Panel c, lanes 1-5: [**Hk-gaga**] hr= 0, 300, 500, 700, 1000 nM with 75 nM of ***cg•GAG***. Experimental conditions as indicated in the caption of Figure **4**23.

Using fluorescence anisotropy, we calculated an apparent K_D for its target site of 92 ± 11 nM at 20 °C (Figure 26, left), in the presence of competing calf thymus DNA (44 ± 6 nM in the absence of calf thymus). As with **gaga-Hk**, in presence of calf thymus, the interaction of **Hk-gaga** with the DNA featuring the A/T-hook mutated site is very weak (Figure 26, right). Taken together, these results confirm the formation of the expected bivalent complex at the specific composite DNA site of nine base pairs (complex **Hk-gaga/*AT•GAG***).

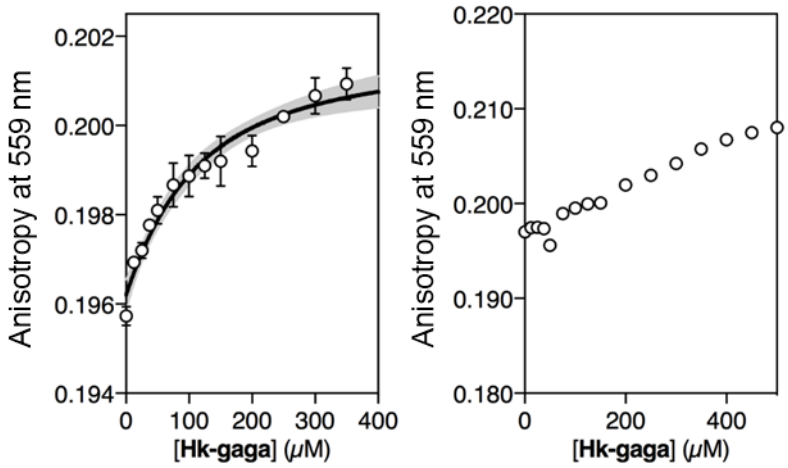
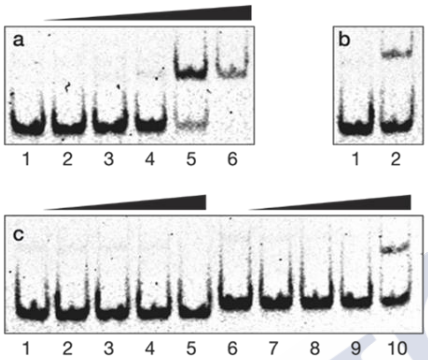


Figure 26. *Left:* Fluorescence anisotropy titration of a 25 nM solution of TMR-***AT•GAG*** in the presence of competing non-specific calf-thymus DNA (50 µM in base pairs) and with increasing concentrations of **Hk-gaga**. The best fit to a 1:1 binding model is also shown, including 95% confidence of the fit shown in light grey. *Right:* Fluorescence anisotropy titration of a 25 nM solution of TMR-***cg•GAG*** in the presence of competing non-specific calf-thymus DNA (50 µM in base pairs) and with increasing concentrations of **Hk-gaga**. Data correspond to the mean of three independent titrations.

Characterization of trimeric conjugate **gaga-Hk-gaga**

We then moved to the more challenging ternary "major-minor-major" groove interaction. In this case, the AT-hook plays the role of a central minor groove anchor that delivers the two GAGA DNA binding domains to the adjacent major groove sites. Gratifyingly, addition of increasing concentrations of the ternary chimera **gaga-Hk-gaga**

to a dsDNA containing the palindromic target composite binding site (***CTC•AT•GAG***), in standard conditions produced a new, slower migrating, band in the EMSA (Figure 27, Panel a), consistent with the formation of the expected specific ternary miniprotein/DNA complex. With the mutated dsDNA ***cat•AT•GAG***, which lacks the first GAGAG site, the gel shows a faint, slowest migrating band, that might correspond to a lower-affinity peptide/DNA complex involving a specific bivalent interaction with the target AATT-GAGAG sequence (Figure 27, Panel b). Significantly, the synthetic miniprotein does not elicit retarded bands when incubated with a non-target sequence lacking both GAGA binding sites, ***cat•AT•cgc*** (Figure 27, Panel c, lanes 1-5). Again, a control oligonucleotide lacking the A/T-rich site, ***CTC•gc•GAG***, showed only residual binding (Figure 27, Panel c, lanes 6-10). This result highlights the important role of the interaction between the AT-hook moiety of the conjugate with its target site, for obtaining high affinity complexes.



CTC•AT•GAG: 5′-GAG CTCTC AATT GAGAG CGCG-3′
cat•AT•GAG: 5′-CGC GTCAT AATT GAGAG CGC-3′
cat•AT•cgc: 5′-CGC GTCAT AATT CGCGA CGC-3′
CTC•gc•GAG: 5′-GGTT CTCTC GACC GAGAG TTGG-3′
TMR-***CTC•AT•GAG:*** 5′-(TMR)-GAG CTCTC AATT GAGAG CGCG-3′

Figure 27. *Left:* EMSA results for **gaga-Hk-gaga**. Panel a, lanes 1-6: [**gaga-Hk-gaga**] = 0, 300, 500, 700, 1000, 1500 nM with 75 nM of ***CTC•AT•GAG***. Panel b, lanes 1-2: [**gaga-Hk-gaga**] = 0, 1000 nM with 75 nM of ***cat•AT•GAG***. Panel c, lanes 1-5: [**gaga-Hk-gaga**] = 0, 300, 500, 700, 1000 nM with 75 nM of ***cat•AT•cgc***. Panel c, lanes 6-10: [**gaga-Hk-gaga**] = 0, 300, 500, 700, 1000 nM with 75 nM of ***CTC•gc•GAG***. *Right:* Oligonucleotide sequences (only one strand shown, consensus sites underlined). Experimental conditions as indicated in the caption of Figure 23.

Fluorescence anisotropy titrations using a TMR-labeled dsDNA (TMR-***CTC•AT•GAG***) confirmed the high affinity binding of the trivalent peptide chimera with the DNA (apparent K_D= 35 ± 4 nM at 20 °C, Figure 28). We could not calculate a reliable K_D in presence of excess calf-thymus DNA due to the formation of aggregates. Any case, these results support the formation of a trivalent DNA complex at the specific composite consensus DNA site of 14 base pairs. The lack of large enhancement in the binding affinity of the ternary chimera (**gaga-Hk-gaga**) versus the bivalent systems (**Hk-gaga** and **gaga-HK**) is likely due to the use of a suboptimal linker that does not allow to take full energetic advantage of the simultaneous interaction of the three binding modules.

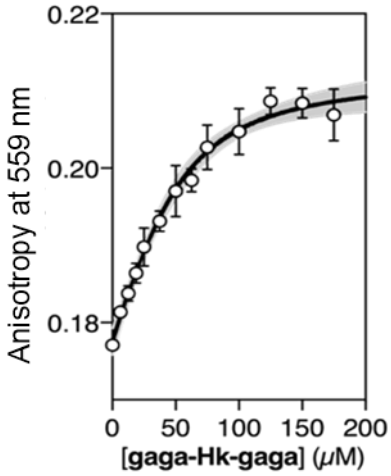


Figure 28. Fluorescence anisotropy titration of a 25 nM solution of TMR-***CTC•AT•GAG*** with increasing concentrations of **gaga-Hk-gaga** and best fit to a 1:1 binding model. Data shown corresponds to the mean of three titrations.

Nanopore force spectroscopy

A collaboration with Dr. David Rodríguez Larrea and Dra. Garbiñe Celaya of Universidad del País Vasco/Euskal Herriko Unibertsitatea, experts in nanopore analysis technology, allowed to test the interaction of the trivalent chimera with DNA using nanopores.[107] This single molecule method has shown utility for the determination of thermodynamic and kinetic parameters in the formation of protein-DNA complexes,[108] and to our knowledge, up to that moment it had not been used to characterize the DNA recognition of synthetic peptide binders. Briefly, it works by stochastically examining DNA states in presence of a given amount of the peptide binder, as described in the next figure.

[107] a) M. M. Marshall, J. Ruzicka, O. K. Zahid, V. C. Henrich, E. W. Taylor, A. R. Hall, *Langmuir* **2015**, *31*, 4582-4588; b) A. Squires, E. Atas, A. Meller, *Sci. Rep.* **2015**, *5*, 11643-11654; c) G. Celaya, J. Perales-Calvo, A. Muga, F. Moro, D. Rodriguez-Larrea, *ACS Nano*. **2017**, *11*, 5815-5825; d) G. Maglia, A. J. Heron, D. Stoddart, D. Japrung, H. Bayley, *Methods Enzymol.* **2010**, *475*, 591-623.

[108] B. Hornblower, A. Coombs, R. D. Whitaker, A. Kolomeisky, S. J. Picone, A. Meller, M. Akeson, *Nat. Methods.* **2007**, *4*, 315-317.

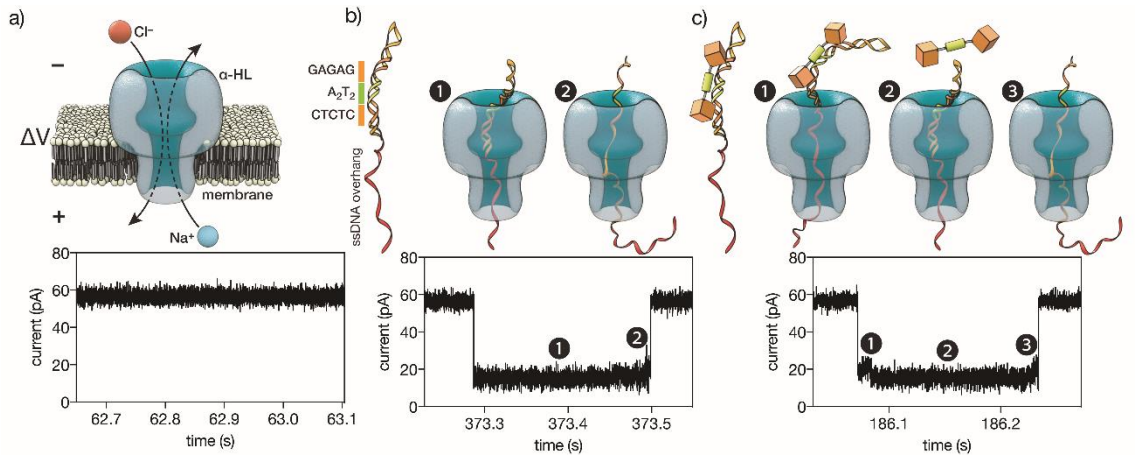


Figure 29. a) A single α-HL pore inserted in a lipid membrane allows the flow of ions in response to an applied voltage. The ionic current of the open pore is shown below. b) A double stranded DNA containing the target CTCTC-A_2T_2-GAGAG binding site inserted in a hairpin followed by a ssDNA overhang (see Experimental Data). This DNA is driven to the pore by the electric field. The threading, unzipping and translocation of the DNA causes the characteristic signal shown below. c) When the trivalent peptide is bound to the DNA there is an additional step with the DNA-protein complex atop of the pore, which causes a new high-conductance level (c1). Once the protein is detached, the reaction proceeds as with free DNA (c2-3). Below the ionic current signal when a protein-DNA complex is analyzed.

Typically, by analyzing up to a hundred DNA molecules, the fraction of complexes can be determined (Figure 30, left), and a K_D deduced (Figure 30, right). When analyzing the interaction of the trivalent chimera with its target ternary binding site (CTCTC-A_2T_2-GAGAG) we calculated a K_D of 120 ± 10 nM. This result is in reasonable agreement with the K_D obtained by fluorescence anisotropy, considering the differences in the assay conditions and in the characteristics of each technique. As expected, mutation of the first GAGAG site (*GTCAT*-A_2T_2-GAGAG) or of the second GAGAG site (CTCTC-A_2T_2-*CTGGG*) led to weaker affinities (K_D of 193 ± 66 nM and 269 ± 39 nM, respectively), in agreement with the trend in binding affinity observed in the EMSA experiments. Interestingly, the nanopore technique allows to measure the time required for a protein to detach from its DNA complex (dwell time in level 1 in

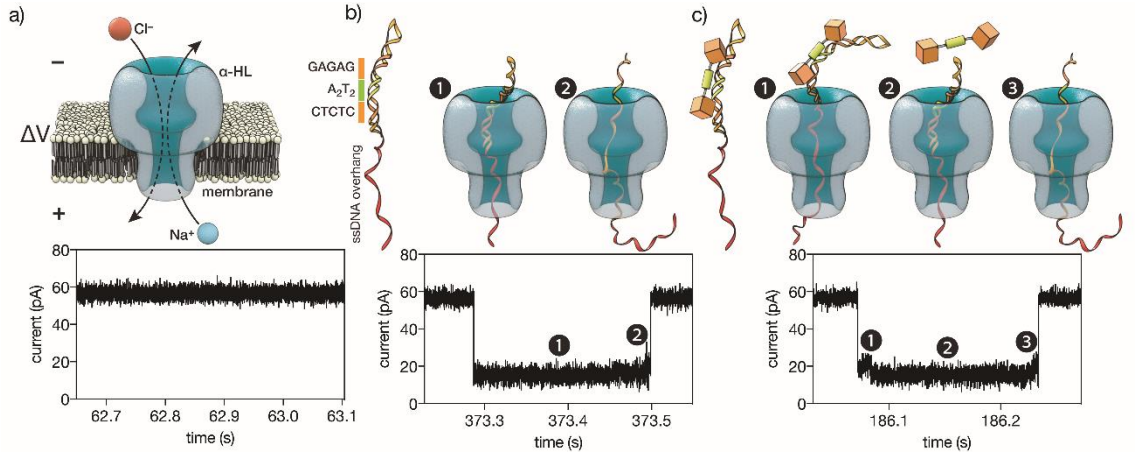


Figure 29c). This information is related to the kinetics in the presence of an applied force.[107c,108]

In order to compare the data obtained on the three different DNAs we fitted each dwell time distribution to a single exponential distribution (see Experimental Data). The values obtained from the fit should be taken cautiously because between 10 to 20% of the molecules did not fit the single exponential distribution. Overall, it was observed that higher voltages induced faster dissociation, likely because of the increased force applied to detach the protein under those conditions (see Experimental Data). The effect of the applied force was larger when the protein was bound to the DNA with the consensus tripartite site (CTCTC-A_2T_2-GAGAG). Within the voltage range studied (from +90 to +120 mV), the slower dissociation was also observed for this DNA (apparent $k_{off,110mV}$ = 90 [100-79] s^{-1} in brackets the 95% confidence interval, CI; n=284). For the DNAs with one mutated site the dissociation was faster (for DNA with GTCAT-A_2T_2-GAGAG, apparent $k_{off,110mV}$ = 165 [197-132] s^{-1} ([95% CI], n=100) and for that with CTCTC-A_2T_2-CTGGG, apparent $k_{off,110mV}$ = 253 [298-208] s^{-1} ([95% CI], n=121).

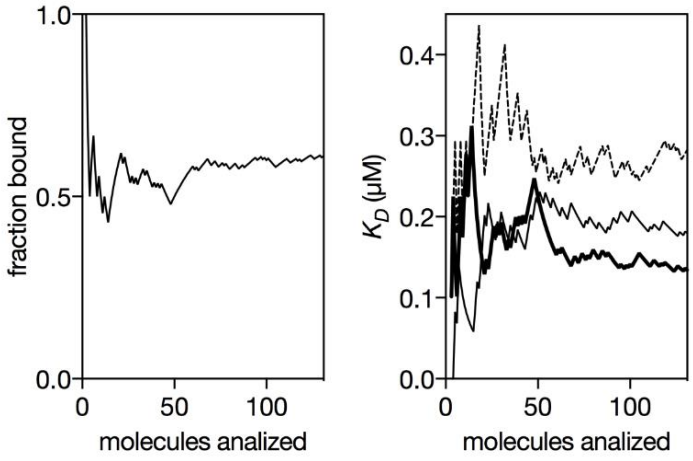


Figure 30. *Left*: Representative experiment where DNA molecules are stochastically analyzed by the pore. The fraction of peptide-DNA complexes in a mixture of 0.29 μM of peptide and 0.13 μM of DNA is calculated counting the number of molecules that produced the signal shown in Figure 29c divided by the total number of molecules that were analyzed (Figure 29b and 29c). Using a reversible one-to-one binding model, K_D = [Free DNA]×[Free Protein]/[Complex], with [DNA] = total concentration of DNA, [Prot] = total concentration of protein and f_b = fraction of DNA molecules observed in the bound state, we estimated an apparent K_D = [DNA]×(1-f_b)×([Prot]-[DNA]×f_b)/ ([DNA]×f_b) (black thick line in right panel). *Right*: Representative experiments for the K_D estimation for a DNA with the target ternary binding site (thick line, 0.29 μM of peptide and 0.13 μM of DNA), for a DNA with the 1st GAGAG site mutated (thin line, 0.52 μM of peptide and 0.13 μM of DNA) and for a DNA with the second GAGAG site mutated (dashed line, 0.19 μM of peptide and 0.13 μM of DNA). The experiments were carried out in 100 mM NaCl, 0.02 mM of ZnCl2, 20 mM Tris-HCl pH 7.5, at 22 ºC.

Mammalian Cell expression

The linear and fully peptidic structure of the DNA recognition framework makes it amenable to genetic encoding, therefore opening new opportunities for artificial gene regulation. With this purpose, two types of plasmids were required to perform the

peptide-DNA recognition inside the cell, in a productive and experimentally detectable manner.

In the first type of constructs, the DNA recognition peptides was cloned into expression plasmids for mammalian cells. Since the DNA sequence for these peptides is non-native, the oligos had to be synthesized "de novo" by a specialized company and subsequently subcloned into suitable vectors (see Experimental Part). These vectors possess viral promoters that ensure high expression of the peptides. In addition, with the aim of converting them into peptides with transcriptional activity, they were expressed as fusion proteins with the transcription factors VP16 and VPR (ATFs, artificial transcription factor).

On the other hand, we had the plasmids called "reporters", where the DNA sequences recognized by our DNA-binding peptides (ATFs) were cloned upstream of reporter genes expressing luciferase or green fluorescent protein, and with a minimal promoter capable of recognize the transcriptional activators VP16 or VPR. Thus, recognition of specific DNA sequences by our designed peptides would bring the activators into contact with the minimal promoter, triggering the expression of the reporter gene (Figure 31).

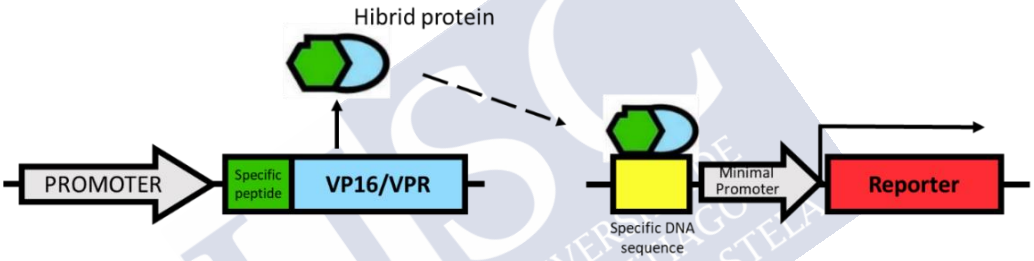


Figure 31. Scheme of transcriptional activation system in cells. Plasmid on the left was engineered to express a conjugate miniprotein of our artificial DNA binding peptide and an activation region (ATFs). Once expressed, it will bind a designed DNA sequence of other plasmid and promote the expression of a reporter.

Because it is highly sensitive and commercially available, we used the luciferase gene as a "reporter" for our assays. The expression of this gene produces the enzyme luciferase, capable of producing light in the presence of its substrate, luciferin. Therefore, the amount of light obtained by the cells would be a direct measure of the interaction of the recognition peptides and the specific DNA sequences. We cloned upstream of the luciferase the specific binding sites applying 1 or 4 copies of each sequence, to improve the signal (Figure 32). Besides, we also cloned sequences missing some recognition sites as negative controls (Figure 32, *Mut*, left).

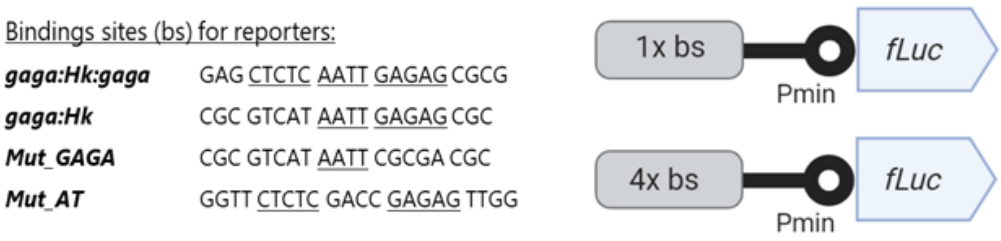


Figure 32. Reporter plasmids. *Left:* Sequences cloned into the reporter plasmids, which have the binding sites for each peptide conjugates or the mutated sequences, as controls. *Right:* Schematic representation of reporter plasmids. For each binding sites (bs) were cloned two different constructs, one with only 1 binding site (1x bs) and other with 4 copies (4x bs), upstream of the minimal promoter (Pmin) and luciferase gen (fLuc) (see Experimental Part).

As DNA recognition peptides, we designed constructs for the trimeric conjugate **gaga-Hk-gaga** and also for the **gaga-Hk** conjugate, both with transcription factors VP16 and VPR in order to compare their activation capacity. The peptides sequences are the same as employed previously, maintaining the required linkers. The resulted constructs are represented in the next figure.[109]

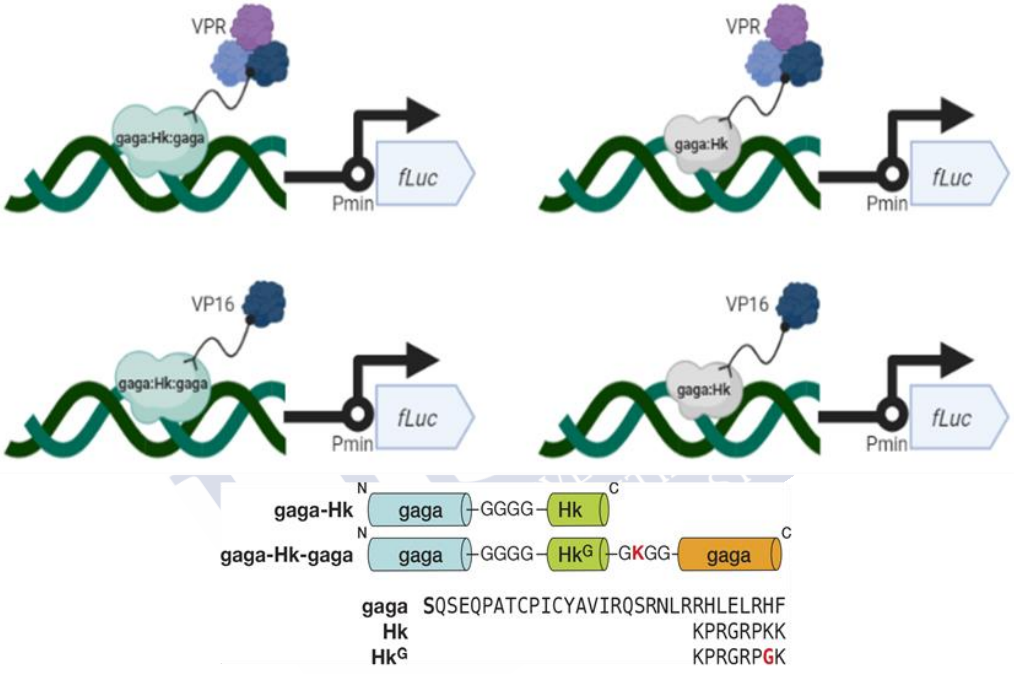


Figure 33. *Top:* Schematic representation of activation of Luciferase by the different recognition peptides expressed as a fused protein with VPR or VP16. *Bottom:* Peptides sequences and linkers.

The different constructs were transfected into Hek293t cell line using polyethylenimine (PEI), a stable cationic polymer. The amounts of transfected plasmids are indicated in the figures. To normalize reporter values to transfection efficiency, we used a constitutively expressed control plasmid, Renilla luciferase encoding plasmid (10 ng). Following 48h after transfection the cells were lysed. Firefly luciferase and Renilla luciferase expression were measured using the dual luciferase assay (Promega) on a microplate reader. Relative luciferase units (RLU) were calculated by normalizing each sample's firefly luciferase activity to the constitutive Renilla luciferase activity.

[109] These experiments were performed in the National Institute of Chemistry in Ljubljana (Slovenia) in the group of Prof. Roman Jerala, during a research stay.

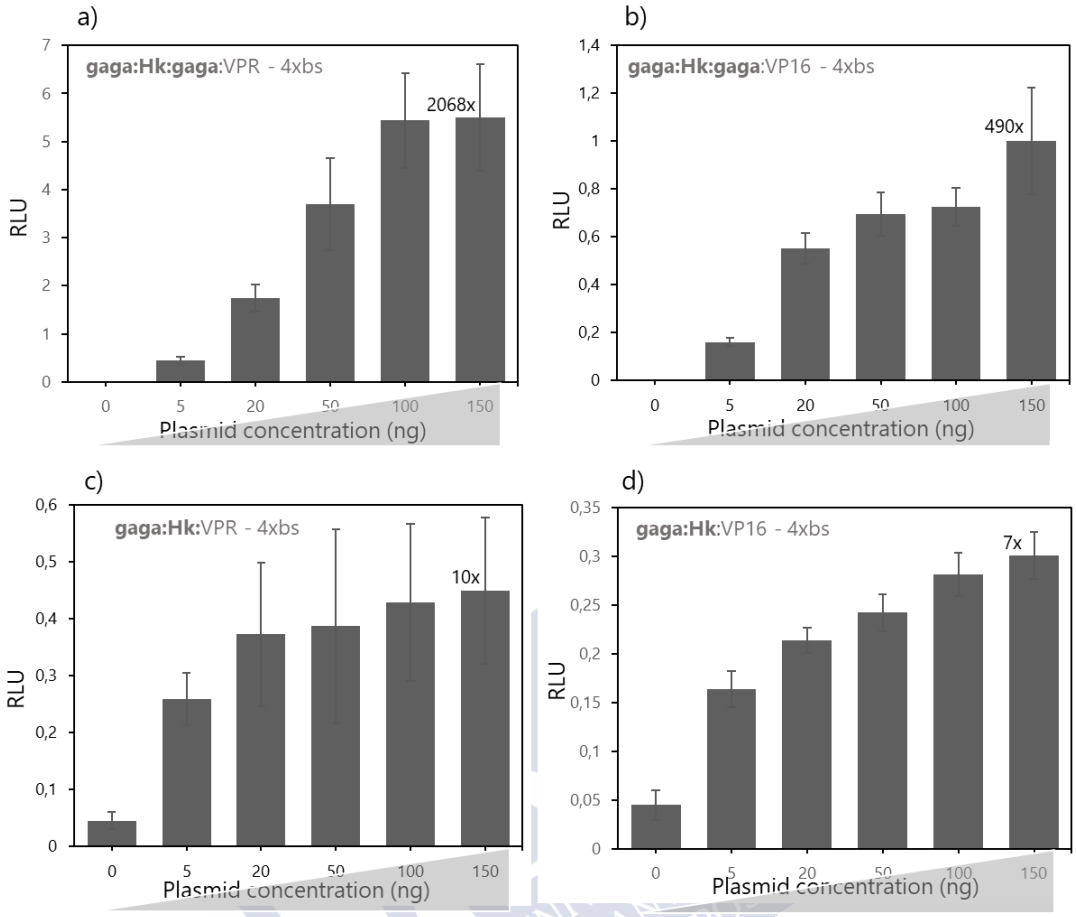


Figure 34. Dual luciferase assay. HEK293T cells were cotransfected with 50 ng of the corresponding Pmin_fLuc reporter plasmid, with different concentration of the encoding plasmids and 10 ng of Renilla Plasmid. The values represent the mean and standard deviation of replicates within three independent experiments (n=4). Fold activation was calculated by normalizing the RLU values of each sample to the RLU value of the reporter only control within the same experiment. a) Activation of its reporter gene expression (4xbs) for **gaga:Hk:gaga** encoding plasmid, fused with VPR factor, and fused with VP16 (b); c) Activation of its reporter gene expression (4xbs) for **gaga:hk** encoding plasmid, fused with VPR factor and fused with VP16 (d).

As we can infer from the data in Figure 34, with the conjugate **gaga:Hk:gaga** we detected the expression of luciferase gen, showed that the peptide is able to recognize its target DNA also in cells environments. Although the activation is higher in the assays based on VPR (2000-fold increase), with VP16 we also detected a good signal (500-fold), using in both cases the reporter with 4 copies of its target biding site (Figure 34, a-b). With only one copy of the target sequence the signal detected is much lower (3-fold). When using as DNA-binding agent the peptide **gaga:hk**, we also detected the luciferase, however with this shorter peptide the activation is lower, around 10-fold increase in the activation employing the transcription factor VPR (Figure 34c).

In addition, to confirm that the gene activation is associated to the selective DNA interaction of the designed miniprotein **gaga:HK:gaga**, we also made control gene

expression assays using reporter plasmids (using in all cases the reporter with 4x bs) containing mutations in the target binding sites (both, for the gaga or Hk binding regions). As can be deduced from the next figure, we could not detect any signal with this new reporter plasmids, which can be explained in terms of the inability of the peptide to recognize the mutated DNA sequences.

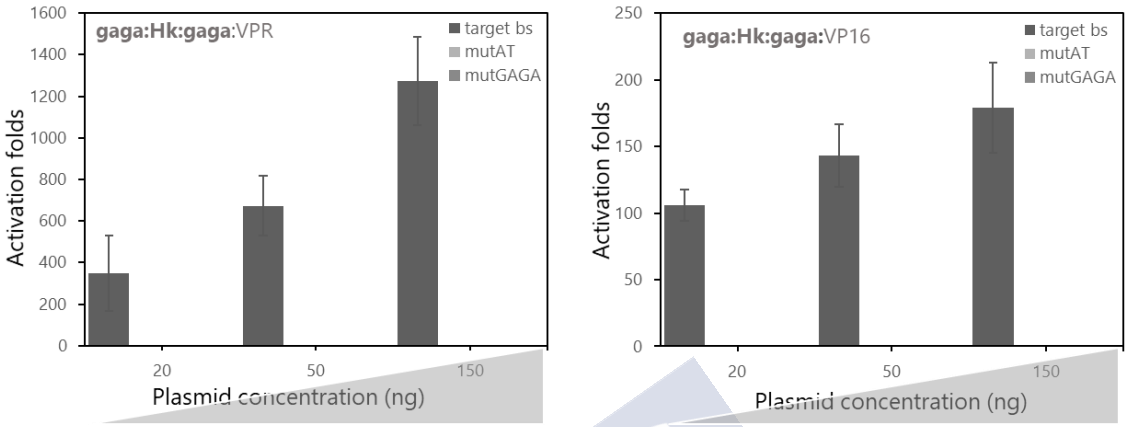


Figure 35. Dual luciferase assay with negative controls, representation of activation folds. HEK293T cells were cotransfected with 50 ng of the different Pmin_fLuc reporter plasmids, with different concentration of the encoding plasmids and 10 ng of Renilla Plasmid. The values represent the mean and standard deviation of replicates within three independent experiments (n=4). Fold activation was calculated by normalizing the RLU values of each sample to the RLU value of the reporter only control within the same experiment. *Left*: Activation of reporter genes expression for **gaga:Hk:gaga** encoding plasmid, fused with VPR. *Right*: Activation of reporter genes expression for **gaga:Hk:gaga** encoding plasmid, fused with VP16.

In order to improve the activation obtained with the shorter conjugate **gaga:hk**, we decided to clone a new expression plasmid containing an additional **Hk** peptide, **Hk(G):gaga:Hk,** and a new reporter plasmid with its corresponding binding site.

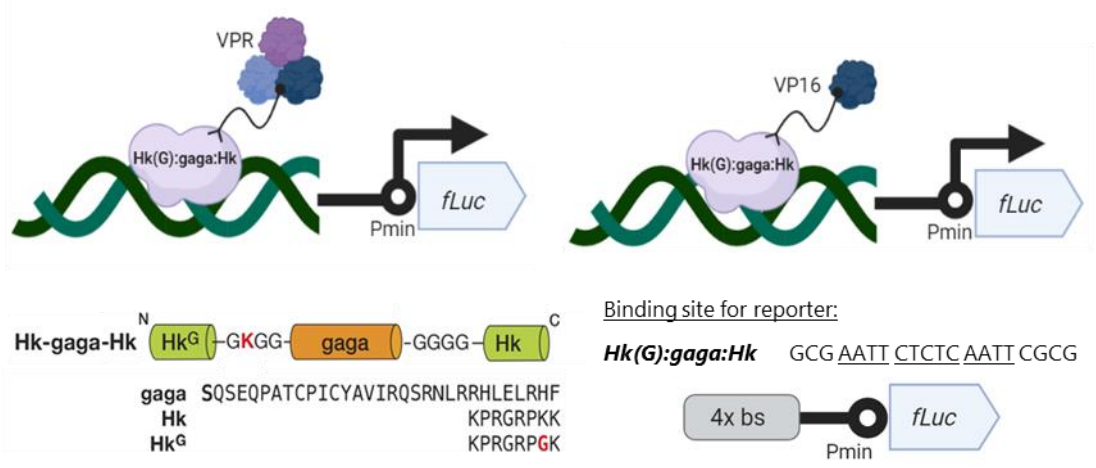


Figure 36. *Top:* Schematic representation of activation of Luciferase by new peptide Hk(G):gaga:Hk expressed as a fused protein with VPR or VP16. *Bottom Left:* Peptides sequences and linkers. *Bottom right:* Sequence cloned into the reporter plasmid and schematic representation of reporter plasmid. We cloned only one construct with 4 copies of this bs (see Experimental Part).

The experiment was performed following the standard protocol. After 48h of the transfection of the constructs and consequent lysate of the cells, firefly luciferase and Renilla luciferase expression were measured using the dual luciferase assay on a microplate reader. As shown in the next figure, the new conjugate is able to recognize its binding site improving the activation more than 10-times compared with short peptide **gaga:hk**. This additional Hk in the sequence helps the system to recognize its binding sites, stabilizing and improving its characteristics.

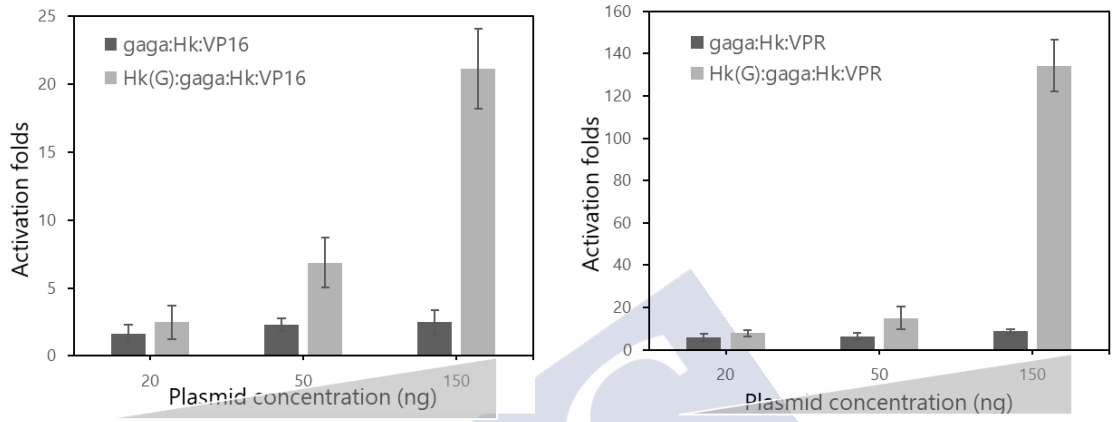


Figure 37. Dual luciferase assay comparing **gaga:hk** and **Hk(G):gaga:Hk** constructs. HEK293T cells were cotransfected with 50 ng of the 10_Pmin-fLuc reporter plasmids, with different concentration of the encoding plasmids and 10 ng of Renilla Plasmid. The values represent the mean and standard deviation of replicates within three independent experiments (n=4). Fold activation was calculated by normalizing the RLU values of each sample to the RLU value of the reporter only control within the same experiment. *Left*: Activation of reporter genes expression for each plasmid encoding **gaga:hk** or **Hk(G):gaga:Hk,** fused both with VPR. *Right*: Activation of reporter genes expression for each plasmid encoding **gaga:hk** or **Hk(G):gaga:Hk,** fused both with VP16.

These results demonstrated that our non-natural DNA recognition peptides, when engineered in appropriate expression vectors, can increase gene expression processes in mammalian cells that contain designed sites in their genome. These results open the door to generation of new genetic tools based in these types of miniaturized, artificial versions of transcription factors, shorter than the zinc fingers described until now.

Conclusions

In summary, we have devised new type of "non-natural" DNA binding motives that exhibit a fully peptidic nature, and therefore could be considered as a new family of DNA binding miniproteins. The bivalent and trivalent constructs can be prepared in a straightforward manner owing to their peptidic constitution and display excellent DNA recognition properties in terms of affinity and selectivity. In addition, we have shown that nanopore technologies allow to obtain biophysical information, in particular kinetic information, that complements that obtained using more standard ensemble techniques. The proteinogenic nature of these artificial DNA binders has allowed their expression in mammalian cells, with good activation results of the reporter genes employed, opening the possibility to the development of genetic tools other than those based on polydactyl zinc fingers.

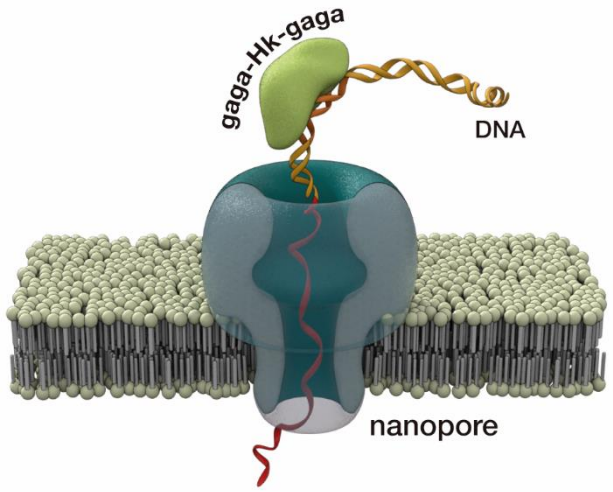


Figure 38. Schematic illustration of DNA recognition by a gaga/AT-hook/gaga chimera using nanopore technique.

Section 2: Assembly of a Ternary Metallopeptide Complex at Specific DNA Sites Mediated by an AT-Hook Adaptor

This chapter includes work published in *Chemical European Journal* as:

S. Learte-Aymamí, J. Rodríguez, M. E. Vazquez, J.L. Mascareñas, *Chem. Eur. J.* **2020**, *26*, 8875-8878.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Introduction

As previously discussed, protein expression is mainly regulated at the transcription stage by the coordinated action of specialized proteins called transcription factors (TFs), which bind to specific DNA sequences.[110] Typically, TFs interact with specific DNA sequences as non-covalent **homo- or hetero-oligomers**, which ensures high affinity, while maximizing the number of accessible DNA sites through combinatorial targeting.[111] In some cases, the DNA binding proteins also use the coordination to metal centers to acquire the tridimensional shape required for the formation of the DNA **supramolecular complexes**. This is the case of zinc-finger proteins, in which the α-helical folding necessary for DNA recognition, is mediated by coordination of a metal ion (Zn^{2+}) to amino acid side chains, typically Cys and His residues.

As chemists, we are not limited to the supramolecular models dictated by nature, and with a judicious introduction of unnatural modifications into the protein scaffolds we might generate new type of DNA binding strategies. Therefore, along recent years, a good number of synthetic DNA binders have been described in the literature. However, most of them are monomeric compounds or hybrids resulting from the covalent modification of DNA binding modules, which complicates the synthesis, and restricts the versatility of the strategy.[112,113]

Nevertheless, supramolecular DNA binders that exploit the self-assembly of different binding modules has been largely overlooked. In this context, we have explored the application of **metal coordination** as a potentially powerful approach to obtaining **multimeric DNA binders**. In particular, previously our group has demonstrated a nickel-promoted DNA interaction of bipartite systems consisting of metal-chelating peptides and minor groove DNA binders equipped with bipyridine moieties.[114] But, so far, this has been limited to bipartite systems with one-metal coordination site, and therefore can only recognize a relatively short sequence.

---

[110] C. W. Garvie, C. Wolberger, *Mol. Cell*. **2001**, *8*, 937-946.

[111] a) H. C. Nelson, *Curr. Opin. Genet. Dev*. **1995**, *5*, 180-189; b) R. Moretti, A. Z. Ansari, *Biochimie* **2008**, *90*, 1015–1025; c) L. Chen, *Curr. Opin. Struct. Biol.* **1999**, *9*, 48-55.

[112] a) C. R. Palmer, S. S. Sloan, J. C. Adrian, B. Cuenoud, D. N. Paolella, A. Schepartz, *J. Am. Chem. Soc.* **1995**, *117*, 8899-8907; b) E. Pazos, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *ChemBioChem*. **2011**, *12*, 1958-1973; c) C. Y. Majmudar, A. K. Mapp, Curr. *Opin. Chem. Biol*. **2005**, *9*, 467-474; d) O. Vazquez, M. E. Vazquez, J. B. Blanco-Canosa, L. Castedo, J. L. Mascareñas, *Angew. Chem. Int. Ed*. **2007**, *46*, 6886-6890; e) A. Iyer, D. Van Lysebetten, Y. Ruiz García, B. Louage, B. G. De Geest, A. Madder, *Org. Biomol. Chem.* **2015**, *13***,** 3856-3862; f) O. Vazquez, M. I. Sánchez, J. Martínez-Costas, M. E. Vazquez, J.L. Mascareñas, *Org. Lett.* **2010**, *12*, 216-219.

[113] a) M. Ueno, A. Murakami, K. Makino, T. Morii, *J. Am. Chem. Soc*. **1993**, *115*, 12575-12576; b) J. B. Blanco, V. I. Dodero, M. E. Vázquez, M. Mosquera, L. Castedo, J. L. Mascareñas, *Angew. Chem. Int. Ed*. **2006**, *118*, 8390-8394.

[114] a) M. I. Sánchez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas *Angew. Chem. Int. Ed*. **2014**, *53*, 9917-9921; b) J. Rodríguez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *Chem. Eur. J*. **2016**, *22*, 13474-13477.

Objective

We propose a Ni(II)-dependent self-assembly of three peptides at DNA sites of up to 13 bp, involving the simultaneous formation of two metal coordination complexes. A key element in our design is the use of an AT-hook peptidic domain equipped with two bipyridine chelating units. In the final complex, this hook will be inserted in the minor groove at the middle of the target DNA site and will recruit two bZIP basic regions to adjacent major grooves in response to the addition of Ni(II) complexes. Moreover, the kinetic lability of the metal coordination should facilitate the disassembly of the supramolecular structure upon addition of external agents that sequester the nickel cation.

Results and discussion

Design and synthesis

For our design reference, we selected the bZIP yeast transcription factor GCN4, which specifically binds to the ATF/CREB (5'–ATGA(c/g)TCAT–3') or AP-1 (5'–ATGA(c)TCAT–3') sites as a leucine zipper-mediated dimer of α-helices.[115] The contacts with the DNA mainly take place through the N-terminal basic region (**br**), which folds into the α-helix upon binding to its target DNA sequence.[116] As we commented previously, isolated basic regions show very weak affinities for the DNA. However, our group has described that both their covalent conjugation to an AT-Hook anchor[117] and non-covalent nickel-promoted assembly,[118] can provide chimeric systems that fold and specifically bind with high affinity to composite sites containing the basic region target sequences.

With this purpose, we have synthesized the fragment of the GCN4 basic region comprising residues Asp^{226} to Arg^{249}, which has been identified as the shortest peptide that, as a disulfide dimer, retains the specific DNA binding properties of the full GCN4 DNA-binding domain.[119] Inspection of the X-ray structure of the GCN4/DNA complex revealed that residues Leu^{230} and Arg^{234} are oriented towards the solvent on the outer face of the α-helix, therefore their substitution by His residues should not affect the DNA contacting surface of the peptide.[116]

We then reasoned that equipping the AT-Hook with two bipyridine ligands (**$HkBpy_2$**) might allow to recruit this modified basic region peptides equipped with two His residues (**$brHis_2$**) to adjacent DNA major grooves in response to a metal clip (Nickel(II)).[120] The Ni(II) mediates the self-assembly of the peptide binders and also acts as α-helix nucleating agent. Interestingly, the whole supramolecular assembly would bind designed long 13bp sites with high selectivity.

---

[115] P. König, T. J. Richmond, *J. Mol. Biol*. **1993**, *233*, 139-154; b) W. Keller, P. König, T. J. Richmond, *J. Mol. Biol*. **1995**, *254*, 657-667.

[116] M. A. Weiss, T. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison, K. Struhl, *Nature* **1990**, *347*, 575-578.

[117] a) J. Rodríguez, J. Mosquera, J. R. Couceiro, M. E. Vázquez, J. L. Mascareñas, *Chem. Sci.* **2015**, *6***,** 4767-4771; b) J. Rodríguez, J. Mosquera, R. García-Fandiña, M. E. Vázquez, J. L. Mascareñas, *Chem. Sci.* **2016**, *7*, 3298-3308.

[118] a) M. I. Sánchez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2014**, *53*, 9917-9921; b) J. Rodríguez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *Chem. Eur. J.* **2016**, *22*, 13474-13477.

[119] R. V. Talanian, C. J. McKnight, R. Rutkowski, P. S. Kim, *Biochemistry* **1992**, *31*, 6871-6875.

[120] The design was based on the structures of the third AT-Hook of HMG-I(Y), bound to the IFN-β promoter (PDB: 3UXW) and the dimeric GCN4 DNA-binding domain bound to the AP1 (5´-ATGA(c)TCAT-3´) site (PDB: 1YSA). Manual overlap of the DNA backbone in the complexes allowed us to construct a rough model of the simultaneous interaction of the AT-Hook and the two GCN4 peptides bound to contiguous sequences of the DNA. Inspection of this model suggested that N- and C-terminal extension of the core AT-Hook sequence with 3-oxapentanoic acid units would allow crossing over the phosphate backbone.

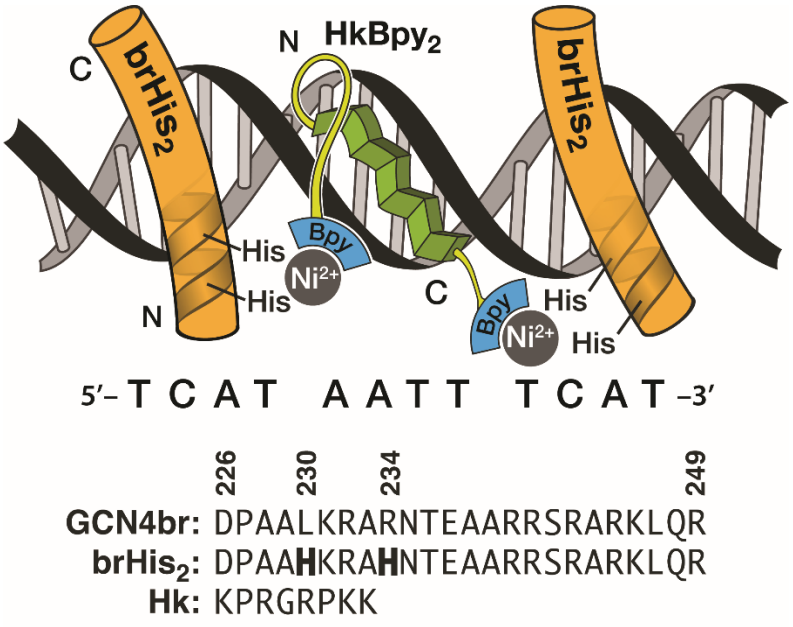


Figure 39. *Top*: Schematic structure of the supramolecular assembly of a bifunctional AT-Hook derivative equipped with bipyridine moieties (**HkBpy$_2$**) and two modified GCN4 basic regions (**brHis$_2$**). His^{230} and His^{234} residues coordinate Ni(II) and connect each of the two **brHis$_2$** to the central AT-hook anchor. *Bottom*: Sequences of the natural GCN4 basic region and AT-hook peptides, as well as the **brHis$_2$**. Mutated residues with respect to the natural GCN4 are in bold.

The peptide **brHis$_2$** and the AT-Hook derivative **HkBpy$_2$** were synthesized following standard solid-phase peptide synthesis procedures.[121] The C-terminal bipyridine was derivatized as a Fmoc-protected amino acid with 5-amino-3-oxapentanoicacid (Fmoc–O1PenBpy–OH, (1), Figure 40a).[122] The N-terminal bipyridine ligand was incorporated while the peptide was still attached to the solid support. All peptides have the C-terminus amidated due to the use of a *Rink amide* resin.

[121] I. Coin, M. Beyermann, M. Bienert, *Nat. Protoc.* **2007**, *2*, 3247-3256.

[122] a) G. Rama, A. Ardá, J. D. Maréchal, I. Gamba, H. Ishida, J. Jiménez-Barbero, M. E. Vazquez, M. Vázquez López, *Chem. Eur. J.* **2012**, *18*, 7030-7035; b) I. Gamba, G. Rama, E. Ortega-Carrasco, J. D. Maréchal, J. Martinez-Costas, M. Eugenio Vázquez, M. V. López, *Chem. Commun.* **2014**, *50*, 11097-11100.

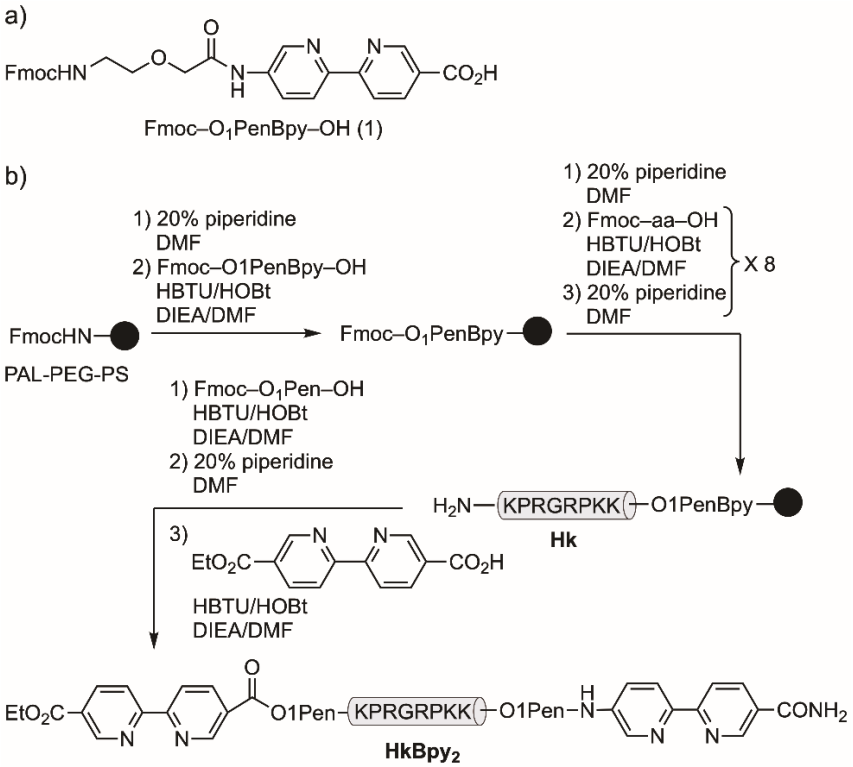


Figure 40. a) Structure of amino acid Fmoc–O1PenBpy–OH. b) Synthetic route to the AT-Hook derivative **HkBpy$_2$**.

DNA binding studies

After the synthesis, purification and characterization of the desired peptides, we studied their DNA binding properties by electrophoretic mobility assays (EMSA) in polyacrylamide gel under non-denaturing conditions,[123] and staining the DNA with SYBR Gold. As expected, incubation of the peptide **brHis$_2$** with the double-stranded (ds) oligonucleotide ***CAT•AT•CAT*** that contains the composite target site, did not induce the formation of retarded bands in the gel (Figure 41a, lane 2). Addition of 20 equiv. of $Ni(ClO_4)_2$ to this mixture also failed to give rise to new retarded bands (Figure 41a, lane 3), but subsequent addition of **HkBpy$_2$** provided a new (more retarded) band that is consistent with the formation of the desired complex [(**brHis$_2$**)$_2$(**HkBpy$_2$**)Ni(II)$_2$]/DNA (Figure 41a, lane 4).

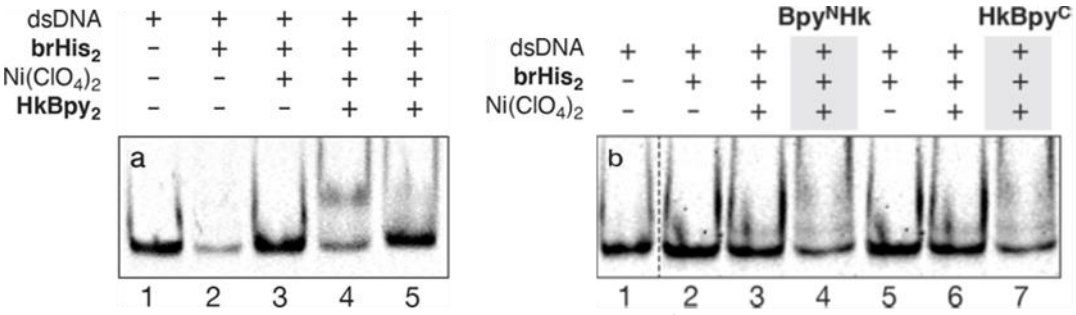
The key role of the nickel ion in mediating the assembly of the supramolecular assembly led us to examine the possibility of dismantling the DNA complex by adding a competitive metal chelator.[124] Gratifyingly, addition of EDTA to the mixture giving rise to the [(**brHis$_2$**)$_2$(**HkBpy$_2$**)Ni(II)$_2$]/DNA complex, led to the disappearance of the slow

[123] L. M. Hellman, M. G. Fried, *Nat. Protoc*. **2007**, *2*, 1849-1861.

[124] For other stimuli responsive-DNA binding tactics, see: a) A. M. Caamaño, M. E. Vázquez, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2000**, *39*, 3104-3107; b) A. Jiménez-Balsa, E. Pazos, B. Martínez-Albardonedo, J. L.Mascareñas, M. E. Vázquez, *Angew. Chem. Int. Ed*. **2012**, *51*, 8825-8829; d) Y. Azuma, M. Imanishi, T. Yoshimura, T. Kawabata, S. Futaki, Angew. *Chem. Int. Ed*. **2009**, *48*, 6853-6856; e) A. Onoda, N. Arai, N. Shimazu, H. Yamamoto, T. Yamamura, *J. Am. Chem. Soc*. **2005**, *127*, 16535-16540.

migrating band in the gel, which is in agreement with the expected disassembly of the supramolecular DNA complex (Figure 41a, Lane 5).

Importantly, control experiments with AT-Hook derivatives featuring a single bipyridine unit, either N-terminal (**Bpy^{N}Hk**) or C-terminal (**HkBpy^{C}**), did not produce slower-migrating bands under the same conditions in the presence of the target dsDNA, ***CAT•AT•CAT*** (Figure 41b).

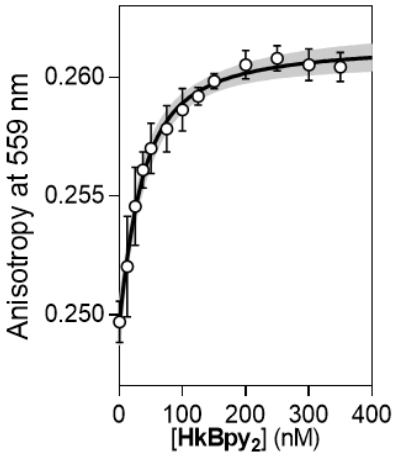


CAT•AT•CAT: 5′-GCGAG TCAT AATT TCAT AGGCG-3′

Figure 41. *Top:* a) DNA binding studies for [(**brHis$_2$**)$_2$(**HkBpy$_2$**)Ni(II)$_2$] with its target ds-oligonucleotide. Concentrations of the components are, when present: 75 nM dsDNA, 2 μM **brHis$_2$**, 20 μM Ni(ClO$_4$)$_2$, and 1 μM **HkBpy$_2$**. Lane 5a is after addition of 50 equiv of EDTA to the mixture in 4a to disassemble the system. b) EMSA of the N-terminal (**Bpy^{N}Hk**) and C-terminal (**HkBpy^{C}**) mono-bipyridine AT-Hook derivatives with the target oligonucleotide (lanes 4 and 7, respectively). *Bottom:* Oligonucleotide doble strand sequence (only one strand shown, consensus sites underlined). EMSA experiments were resolved on a 10% nondenaturing polyacrylamide gel and 0.5X TBE buffer over 40 min at 4 °C and analyzed by staining with SYBR Gold (5 μL in 50 mL of 1X TBE) for 10 min, followed by fluorescence visualization. Note: The slight smearing of the bands comes from the absence of EDTA, commonly used in these assays, and possibly also by the formation binary assemblies [(**brHis$_2$**)(**HkBpy$_2$**)Ni(II)] forming low-affinity complexes with the dsDNA.

The EMSA conditions require the use of relatively large amounts of DNA and peptides, which induces some precipitation, and could also favor the kinetic disassembly of the complex, and thus the bands in the EMSA were smearing. We therefore decided to quantify the interaction spectroscopically by running fluorescence anisotropy titrations. These experiments were carried out by adding increasing concentrations of the peptide **HkBpy$_2$** to a nanomolar solution containing a tetramethylrhodamine-labeled dsDNA with the target sequence, the peptide **brHis$_2$**, and 20 μM Ni(ClO$_4$)$_2$. As shown below (Figure 42), titration with **HkBpy$_2$** led to a progressive increase in the anisotropy, in agreement with the formation of a larger complex with decreased rotational diffusion.[125] Considering a simplified 1:1 binding mode in which the [(**brHis$_2$**)$_2$(**HkBpy$_2$**)Ni(II)$_2$] complex act as a single species binding to the DNA, we calculated an apparent dissociation constant of 22.3 ± 0.4 nM at 4 °C. Taken together, these results confirm the need of the metal-chelating **HkBpy$_2$** anchor for the formation of the DNA peptide complex.

[125] V. J. LiCata, A. J. Wowor, *Methods in Cell Biology* **2008,** Academic Press, 243-262.



*TMR-**CAT•AT•CAT:*** 5´-(TMR)-GCGAG TCAT AATT TCAT AGGCG-3´

Figure 42. *Top:* Fluorescence anisotropy titration at 559 nm of a 25 nM solution of a tetramethylrhodamine-labeled ds-oligonucleotide (TMR-***CAT•AT•CAT***) in the presence of 500 nM of **brHis$_2$** and 20 μM $Ni(ClO_4)_2$ with increasing concentrations of **HkBpy$_2$**. The best fit to a 1:1 binding model is also shown. 95% confidence of the fit in light grey. Data correspond to the mean of three independent titration experiments. All experiments carried out at 4°C. *Bottom*: Labeled oligonucleotide target sequence (only one strand shown, consensus sites underlined).

Additional controls

Additional control experiments with a basic region peptide containing a single His residue, confirmed that both histidines are required to detect any DNA binding (Figure 43a). As expected, the AT-Hook peptide **HkBpy$_2$** by itself does not elicit new EMSA bands upon incubation with the oligo ***CAT•AT•CAT*** that contain a target sequence A_2T_2 (Figure 43b),[126] demonstrated that the system assembles dynamically regardless of the order in which the components are mixed.

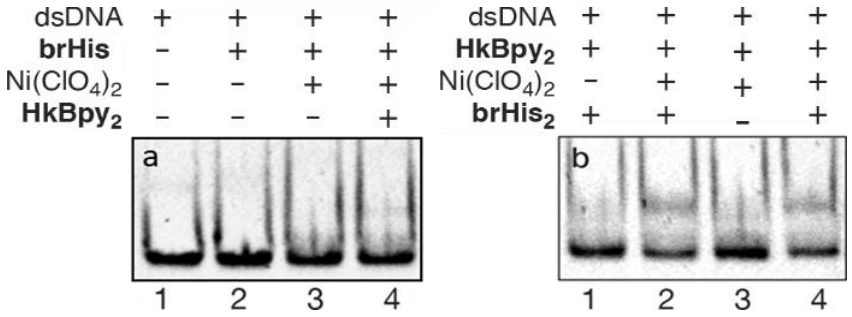
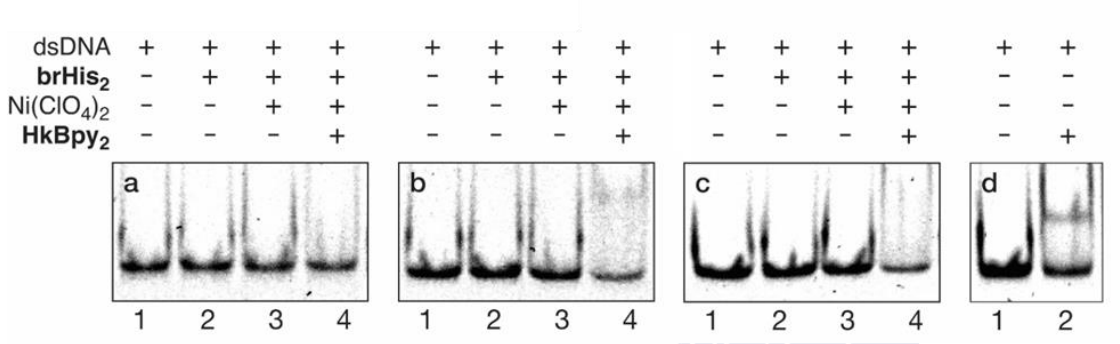


Figure 43. a) The DNA binding profile was further confirmed using a mutated peptide, **brHis**, with its target oligonucleotide ***CAT•AT•CAT.*** Concentrations of the components are, when present: 75 nM dsDNA, 1 μM **HkBpy$_2$,** 20 μM $Ni(ClO_4)_2$, and 2 μM **brHis**. b) Different order of addition to confirm the formation of the system only when all components are present (Lanes 2 and 4). Experimental conditions as indicated in the caption of Figure 41.

[126] a) C. Crane-Robinson, A. I. Dragan, P. L. Privalov, *Trends. Biochem. Sci.* **2006**, *31*, 547-552; b) L. Aravind, D. Landsman, *Nucleic Acids Res.* **1998**, *26*, 4413-4421; c) R. Reeves M. S. Nissen, *J. Biol. Chem*. **1990**, *265*, 8573-8582; d) T. Lund, K. H. Dahl, E. Mork, J. Holtlund. S. G. Laland, *Biochem. Biophys. Res. Commun.* **1987**, *146*, 725-730.

We next studied the selectivity of the recognition process by using other dsDNAs with specific mutations. Not surprisingly, we did not observe the formation of retarded bands in the presence of mutated dsDNAs in both AP1hs sites (***ctc•AT•gag***, Figure 44a), or just in one of them (***CAT•AT•gag***, Figure 44b), which confirms the cooperativity of the coordination-mediated assembly. Furthermore, we did not observe the formation of retarded bands in the presence of a DNA lacking the A/T-rich tract required for AT-Hook binding (***CAT•cg•CAT***, Figure 44c). This result contrasts with previously studied ternary covalent derivatives that shows binding to the DNAs lacking the AT-rich tract,[117] and highlights the advantages of using non-covalent, dynamic supramolecular interactions to achieve high-affinity and selective DNA interactions. Not surprisingly, the flexibility of the supramolecular system allows some mobility in the minor groove, and thus tolerates a slightly larger A/T spacer sequence between the basic region consensus sites (TA_2T_2), albeit the interaction in that case appears less intense (Figure 44d).



ctc•AT•gag: 5′- CGCGC TCTC AATT GAGAG CTC-3′

CAT•AT•gag: 5′- CGCAG TCAT AATT GAGAGCGC -3′

CAT•cg•CAT: 5′- GCGA GTCAT CAGG TCAT AGGCG -3′

CAT•TAT•CAT: 5′- GCGA GTCAT TAATT TCAT AGGCG -3′

Figure 44. *Top:* EMSA studies with [(**brHis$_2$**)$_2$(**HkBpy$_2$**)Ni(II)$_2$]. a) ***ctc•AT•gag*** (both peptide sites mutated); b) ***CAT•AT•gag*** (only the binding site of one basic region mutated); c) ***CAT•cg•CAT*** (AT-hook site mutated); d) ***CAT•TAT•CAT*** (extended AT-hook site). Concentrations of the components are, when present: 75 nM dsDNA, 1 μM **HkBpy$_2$**, 2 μM **brHis$_2$**, and 20 μM $Ni(ClO_4)_2$. Experimental conditions as indicated in the caption of Figure 41. *Bottom*: Oligonucleotide sequences (only one strand shown, consensus sites underlined).

We also studied the selectivity of the system with regard to the metal ion (Co(II), Ni(II), Zn(II), Cu(II), and Fe(II) 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 presents the best combination of properties—i.e., geometry, affinity and lability—to promote an effective trimeric assembly in the presence of the cognate DNA.[127]

[127] For examples of nickel(II)-promoted biomolecular folding or assembly, see: a) *R*. J. Radford, F. A. Tezcan, *J. Am. Chem. Soc*. **2009**, *131*, 9136-9137; b) M. Kalek, A. *S*. Madsen, J. Wengel, *J. Am. Chem. Soc.* **2007**, *129*, 9392-9400; c) M. R. Ghadiri, C. Choi, *J. Am. Chem. Soc.* **1990**, *112*, 1630-1632.

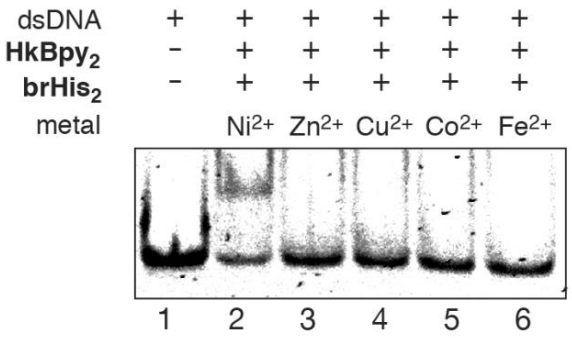


Figure 45. Metal screening with target oligonucleotide ***CAT•AT•CAT***. Concentrations of the components are, when present: 75 nM dsDNA, 1 μM **HkBpy$_2$,** 20 μM metal complexes and 2 μM **brHis$_2$**. Lane 2: $Ni(ClO_4)_2$; Lane 3: $ZnSO_4$; Lane 4: $CuSO_4$; Lane 5: $Co(ClO_4)_2$; Lane 6: $Fe(NH_4)_2(SO_4)_2$. Experimental conditions as indicated in the caption of Figure 41.

Spectroscopic characterization

As expected for a poorly structured peptide, the circular dichroism spectrum of **brHis$_2$** presents a relatively weak negative signal at 222 nm, even in the presence of **HkBpy$_2$** and the consensus DNA ***CAT•AT•CAT*** (Figure 46, solid line); the addition of $Ni(ClO_4)_2$ to the mixture promoted a considerable increase in the negative ellipticity intensity at 222 nm, which is consistent with the folding of the peptide chain into an α-helix (Figure 46, thick line).[128] In accordance with the results obtained by EMSA, the addition of EDTA to the supramolecular complex that resulted from mixing **brHis$_2$**, **HkBpy$_2$**, the Ni(II) salt, and consensus DNA promoted a drastic decrease in the helicity of the peptide (Figure 46, dotted line), which correlates with the disruption of the DNA complex.

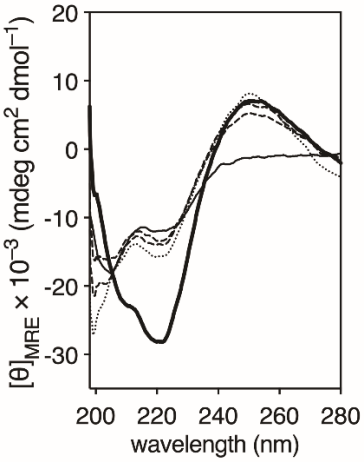


Figure 46. Circular dichroism of a 5 μM solution of **brHis$_2$** in 10 mM phosphate buffer pH 7.5 and 100 mM of NaCl (solid line). Spectra after the subsequent addition of 1 equivalent of the target dsDNA and 1 equivalent of **HkBpy$_2$** (dashed lines), and after the addition of $Ni(ClO_4)_2$ (thick solid line). Addition of EDTA (dotted line) promotes the disassembly of the complex and demonstrates reversibility. The contribution of the DNA to the CD spectrum have been subtracted for clarity and the MRE is calculated with respect to the 24-mer brHis$_2$. All experiments carried out at 4 °C.

[128] J. T. Pelton, L. R. McLean, *Anal. Biochem*. **2000**, *277*, 167-176.

HkBpy$_2$ is not an innocent linker

Finally, we tested if the supramolecular complex could be assembled by a linker that, unlike the AT-Hook, does not bind to the DNA minor groove. With this aim in mind we synthesized flexible oligoglycine and rigid oligoproline linkers equipped with the two bipyridines (**Gly$_7$Bpy$_2$** and **Pro$_7$Bpy$_2$**, respectively). Both peptides are long enough to span the distance of approximately 32 Å between the coordinating histidine residues in the two **brHis$_2$** modules. Curiously, neither **Gly$_7$bpy$_2$** nor **Pro$_7$bpy$_2$** gave rise to stable peptide/DNA complexes upon incubation with DNAs containing composite consensus sequences (Figure 47). These results confirm that the AT-Hook is not a passive connector, but its interaction with the DNA and insertion in the A/T rich minor groove are essential to obtain the desired peptide/DNA assembly.

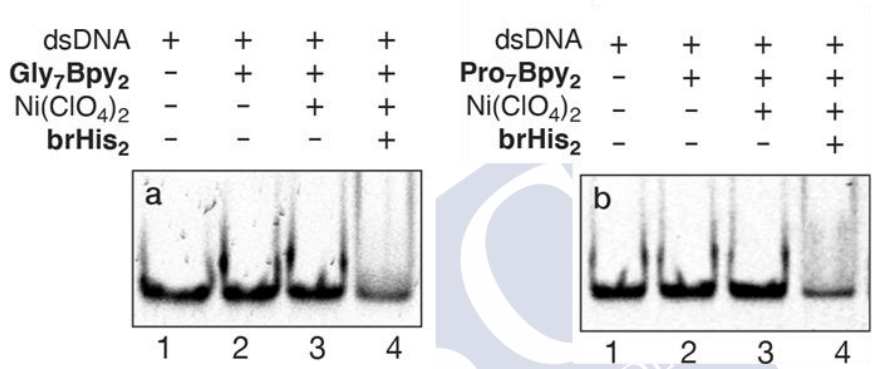


Figure 47. DNA binding studies for [(**brHis$_2$**)$_2$(**Gly$_7$bpy$_2$/Pro$_7$bpy$_2$**)Ni(II)$_2$] with ***CAT•AT•CAT***. Concentrations of the components are, when present: 75 nM dsDNA, 1 μM **Gly$_7$bpy$_2$ / Pro$_7$bpy$_2$**, 20 μM Ni(ClO$_4$)$_2$, and 2 μM **brHis$_2$.** Experimental conditions as indicated in the caption of Figure 41.

Conclusions

In summary, we have introduced a new approach for achieving a highly selective trivalent recognition of specific, relatively long sequences of DNA, using non-covalent assemblies. The recognition systems engage three individual peptide modules. Key for the success of the approach is the dual role of the metal as an α-helix-nucleating factor and heterodimerization clip. The multicomponent nature of the system and the kinetic lability of the metal coordination facilitate the disassembly of the supramolecular structure upon addition of external agents that sequester the nickel cation. The reported system features some of the emergent properties exhibited by naturally occurring DNA-binding proteins, such as multivalence, selectivity, responsiveness to external stimuli and reversibility. The system represents an infrequent case of self-assembly, as it involves four different components: a metal, two peptides, and a nucleic acid.

Figure 48. Schematic illustration of tripartite major-minor-major groove recognition by a bZIP/AT-hook/BZIP chimera.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Section 3:
Metal-Dependent DNA Recognition and Cell Internalization of Designed, Basic Peptides

This chapter includes work published in *Journal of the American Chemical Society* as:

S. Learte-Aymamí, N. Curado, J. Rodríguez, M. E. Vázquez, J. L. Mascareñas, *J. Am. Chem. Soc*. **2017**, *139*, 16188-16193.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Introduction

Protein expression depends on the concerted action of transcription factors (TFs).[129] Over the last few years, there has been a great interest in the development of **miniaturized models of TFs** capable of reproducing their DNA-binding properties, as they might open interesting biomedical opportunities.[130] However the design and preparation of this minimalistic peptides has proven extremely challenging.

The most successful examples have been inspired by the DNA binding domain of bZIP proteins such as GCN4, an archetypal member of this family of TFs, studied previously in Section II, that specifically binds to palindromic ATF/CREB (5'-ATGAC-GTCAT-3') or AP1 (5'-ATGA(c)TCAT-3') sites as a leucine zipper-mediated dimer of uninterrupted α-helices.[131] The DNA interaction occurs through the N-terminal basic region (br), which folds into an α-helix upon insertion in the major groove of the target DNA site.[132]

Following the seminal work by Kim and coworkers in 1990,[133] a number of groups have demonstrated that the complete leucine zipper of the GCN4 DNA binding domain could be replaced by artificial dimerizing elements without significantly compromising the DNA recognition properties of the system.[134] In contrast, **monomeric GCN4** peptides display very low DNA-binding affinity,[135] probably because the high entropic cost associated with the peptide chain folding into the α-helical conformation, required for DNA binding, is not compensated by the enthalpic gain of a monomeric interaction.[136,137] A couple of reports have shown that introducing lactam bridges or covalent staples can promote DNA binding, albeit the affinity and selectivity of these modified systems is rather modest.[138] Better results have been obtained by grafting key DNA binding residues from the bZIP basic regions into stable peptide scaffolds that support the α-helical conformation (see General Introduction, section 5.2).[139]

Despite these important advances, an efficient and selective major groove DNA interaction by simple, minimalist peptides, featuring **natural amino acids**, had not been demonstrated.

---

[129] a) D. S. Latchman, *Eukaryotic Transcription Factors*, Elsevier, London, 5th edn, **2011**, ch. 1, p. 1; b) A.L. Todeschini, A. Georges, R. A. Veitia, *Trends Genet*. **2014**, *30*, 211-219.

[130] a) L. Strekowski, B. Wilson, *Mutat. Res.* **2007**, *623*, 3-13; b) P.S. Arora, A. Z. Ansari, T.P. Best, M. Ptashne, P.B. Dervan. *J. Am. Chem. Soc.*, **2002**, *124*, 13067-13071.

[131] a) T.E. Ellenberger, C.J. Brandl, K. Struhl, S.C. Harrison, *Cell* **1992**, *71*, 1223-1237; b) P. Konig, T. J. Richmond, *J. Mol. Biol*. **1993**, *233*, 139-154; b) W. Keller, P. Konig, T. J. Richmond, *J. Mol. Biol*. **1995**, *254*, 657-667.

[132] M. A. Weiss, T. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison, K. Struhl, *Nature* **1990**, *347*, 575-578.

[133] R. V. Talanian, C. J. McKnight, P. S. Kim, *Science* **1990**, *249*, 769-771.

[134] See Introduction: section 5.1.

[135] X. Wang, W. Cao, A. Cao, L. Lai, *Biophys. J.* **2003**, *84*, 1867-1875.

[136] M. I. N. Zhang, B. Wu, H. Zhao, J. W. Taylor, *Pept. Sci*. **2002**, *136*, 125-136.

[137] C. Park, J. L. Campbell, W. A. Goddard, J. Am. Chem. Soc. **1996**, 118, 4235-4239.

[138] A. Iyer, D. Van Lysebetten, Y. Ruiz García, B. Louage, B. G. De Geest, A. Madder, *Org. Biomol. Chem*. **2015**, *13*, 3856-3862.

[139] a) J. k. Montclare, A. Schepartz, *J. Am. Chem. Soc.* **2003**, *125*, 3416-3417; b) N. J. Zondlo, A. Schepartz, *J. Am. Chem. Soc*. **1999**, *121*, 6938-6939.

Objective

We propose the design of bis-histidine grafted peptides featuring 23 amino acids of the basic region of GCN4 that might bind their consensus DNA site (5'-GATGA-3') only in the presence of specific metal complexes. The metal should work as a stapling agent that favors the required α-helical folding, and ultimately, the insertion of the peptide into the major groove of the target DNA. The peptide-DNA complex could be disassembled by addition of a metal chelator. Importantly, we suggest the possibility that the metallopeptide could translocate across cell membrane as others stapled peptides.

Results and discussion

Design and synthesis

As in previous cases, as reference for our design we used a fragment of the GCN4 basic region substituting Leu^{230} and Arg^{234} by two His residues. Thus, we synthesized the peptide **brHis$_2$**, featuring two His residues, as well as the natural basic region peptide (**br**), and the control peptide **brHis**, in which only one residue is replaced by His (Arg^{234}→ His).

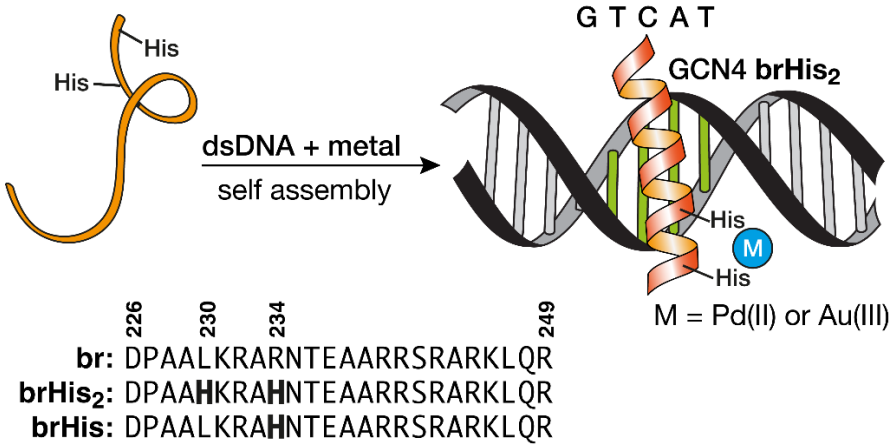


Figure 49. *Top:* Schematic structure of the supramolecular assembly into specific DNA sites of a modified GCN4 basic region featuring two His residues (His^{230} and His^{234}, **brHis$_2$**) in presence of a metal clip. *Bottom:* Sequences of the natural GCN4 basic region and the mutated peptides **brHis$_2$** and **brHis**. Mutated residues are in bold.

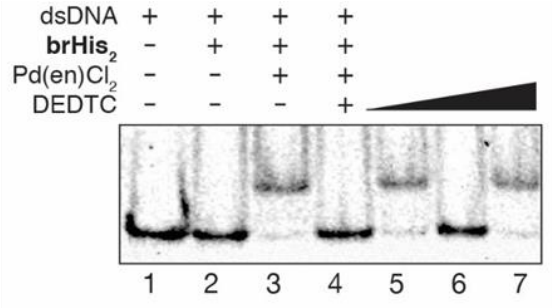
DNA binding of the peptide **brHis$_2$**

The DNA binding properties were studied by electrophoretic mobility assays (EMSA) in polyacrylamide gel under non-denaturing conditions, using SybrGold as the DNA stain.[140] As expected, incubation of the peptide **brHis$_2$** with the double-stranded (ds) oligonucleotide ***GTCAT*** containing the consensus binding site (5'-GTCAT-3'), does not induce the formation of retarded bands (Figure 50, lane 2), even when using a large excess of the peptide. This is in consonance with the poor DNA affinity of such monomeric basic regions. However, addition of 20 equiv of *cis*-$PdCl_2$(en) (en, ethylenediamine) to this mixture gives rise to a neat, new (more retarded) band that is consistent with the formation of the desired peptide-DNA complex ***GTCAT***/[(**brHis$_2$**)Pd(en)]$^{2+}$ (Figure 50, lane 3).

Remarkably, the complex can be dismantled by the addition of a small excess of an external Pd chelator (diethyldithiocarbamate, DEDTC, Figure 50, lane 4), a result that supports the role of the palladium reagent in promoting the DNA binding. Furthermore, the DNA complex can be reversibly assembled or dismounted by controlling the relative amounts of the Pd complex and the DEDTC (Figure 50, lane 4-7). To our knowledge,

[140] a) W. Hendrickson, *BioTechniques*, **1985**, *3*, 198-207; b) R. S. Tuma, M. P. Beaudet, X. Jin, L. J. Jones, C. Y. Cheung, S. Yue, V. L. Singer, *Anal. Biochem.* **1999**, *268*, 278–288.

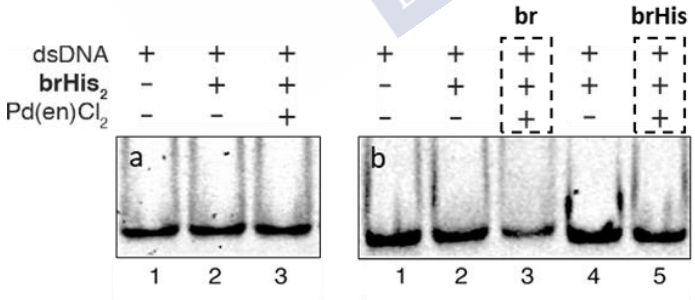
reversible cyclic DNA binding of such minimalist synthetic peptides lacks any precedent.[141,142]



GTCAT: 5′-CGC GTCAT AATTGAGAGCGC-3′

Figure 50. *Top*: DNA binding studies of **brHis$_2$** with its target ds-oligonucleotide. Concentration of the components when present: 75 nM dsDNA, 2 μM **brHis$_2$**, 20 μM *cis*-$PdCl_2$(en). Lane 1: dsDNA; Lane 2: dsDNA and **brHis$_2$**; Lane 3: previous mixture and *cis*- $PdCl_2$(en); Lane 4: mixture in lane 3 after addition of 50 equiv DEDTC; Lane 5: addition of 60 equiv *cis*- $PdCl_2$(en) to the mixture in lane 4; Lane 6: mixture in lane 5 after addition of 150 equiv DEDTC; Lane 7: addition of 200 equiv *cis*-$PdCl_2$(en) to the mixture in lane 6. All the equiv are expressed relative to **brHis$_2$**. *Bottom:* Oligonucleotide double strand sequence (only one strand shown, consensus sites underlined). Samples were resolved on a 10% nondenaturing polyacrylamide gel and 0.5X TBE buffer over 40 min at 25 °C and stained with SyBrGold (5 μL in 50 mL of 1X TBE) for 10 min, followed by fluorescence visualization.

The metal-promoted interaction is highly selective; therefore, mutation in a single position of the target site was enough to completely abolish the DNA binding (Figure 51a). It is also important to note that neither the natural GCN4 basic region (**br**), nor the peptide containing a single His mutation (**brHis**) give rise to new bands in the presence of palladium (Figure 51b), which confirm the requirement of the two coordinating histidines.



MUT: 5′-CGC GTGAT AATTGAGAG CGC-3′

Figure 51. *Top:* a) Incubation with mutated dsDNA ***MUT.*** b) The DNA binding profile was further confirmed using two different peptides: **br** (no mutations) and **brHis** (single His mutated) with target ds-oligonucleotide ***GTCAT***. Concentration of the components are, when present: 75 nM dsDNA, 2 μM peptides, 20 μM *cis*- $PdCl_2$(en). Experimental conditions as indicated in the caption of Figure 50. *Bottom:* Oligonucleotide sequence (only one strand shown, consensus sites underlined).

[141] A reversible binding of peptide dimers was reported by our group: J. Mosquera, A. Jiménez, V. I. Dodero, M. E. Vázquez, J. L. Mascareñas, *Nat. Commun*. **2013**, *4*, 1874-1881.

[142] There is an earlier report of a Co(II)-responsive GCN4 dimer featuring iminodiacetic acid residues: Y. Azuma, M. Imanishi, T. Yoshimura, T. Kawabata, S. Futaki, Angew. *Chem. Int. Ed*. **2009**, *48*, 6853–6856.

The above data are consistent with the proposed DNA binding by a peptide monomer. Binding by a dimeric species is unlikely, as it would require a specific arrangement of two basic regions that is difficult to envision. Nevertheless, to fully discard the presence of such dimeric species binding to the DNA target, we carried out control experiments with a disulfide dimer of the basic region peptide [**br$_2$(SS)**], similar to that used by Kim *et al.*[143] As shown in the next figure, the interaction of this dipeptide with a DNA featuring either its consensus dimeric target site (***AP-1***) or the monomeric half site (***GTCAT***) leads to slower migrating bands than those observed for the DNA complex with [**brHis$_2$**(Pd(en))]$^{2+}$. The retarded band observed in lane 8 with ***AP1***, in presence of **brHis$_2$** and the Pd reagent, is consistent with the insertion of two monomeric peptides in their respective half-sites.

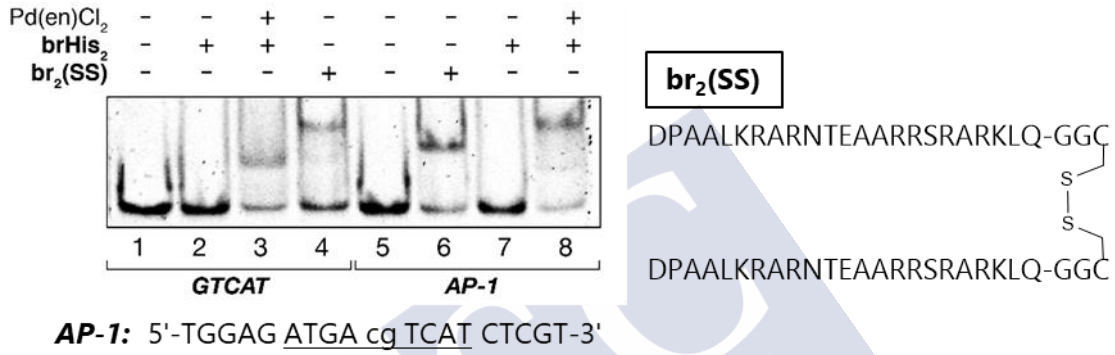


Figure 52. *Top Left:* Comparative EMSA experiment with monomeric **brHis$_2$** and **br$_2$(SS)**. Lanes 1-4: 75 nM ***GTCAT***; Lanes 2-3: 2 μM **brHis$_2$**; Lane 3: 20 μM $PdCl_2$(en); Lane 4: 1.2 μM **br$_2$(SS)**; Lane 5-8: 75 nM ***AP-1***; Lane 6: 1,2 μM **br$_2$(SS)**; Lane 7-8: 2 μM **brHis$_2$**; lane 8: 20 μM $PdCl_2$(en). *Bottom Left:* Oligonucleotide sequence (only one strand shown, consensus sites underlined). The dsDNA ***AP-1*** features two binding sites in neighbor sequences. *Right:* Sequence of **br$_2$(SS)** peptide. Each peptide of the disulfide features the sequence of peptide **br** (Figure 49, bottom). with the C-terminal Arg (249) substituted by Gly-Gly-Cys. Experimental conditions as indicated in the caption of Figure 50.

DNA binding is only observed in presence of specific metal complexes

At this point we checked whether other metal ions could also promote DNA binding, or if the induction of DNA interaction was exclusive for the palladium reagent. As shown in Figure 53a, mixing the target dsDNA ***GTCAT*** with **brHis$_2$** in the presence of several Ni^{2+}, Zn^{2+}, Cu^{2+}, Co^{2+}, or Fe^{2+} salts, failed to give raise to new bands in the EMSA. It is important to highlight the failure of $NiCl_2$, as this metal complex had been successfully used for promoting bivalent major-minor groove interactions in cooperation with bipyridine ligands, as have been described before.[144] However, it is not useful in the case of the more challenging monovalent peptide-DNA interactions.

Not surprisingly, *trans*-$PdCl_2(PPh_3)_2$ (PPh_3, Triphenylphosphine) also failed to induce DNA binding, while *cis*-$PdCl_2$(bpy) (bpy, bipyridine) was roughly as effective as *cis*-$PdCl_2$(en).

[143] R. V. Talanian, C. J. McKnight, R. Rutkowski, P. S. Kim, *Biochemistry* **1992**, *31*, 6871-6875.

[144] a) J. Rodríguez, J. Mosquera, R. García-Fandiño, M. Eugenio Vázquez, J. L. Mascareñas, *Chem. Sci.* **2016**, *7*, 3298-3303; b) Chapter I: Section 2.

We also proved that an analogous complex featuring nitrate instead of chloride ligands $Pd(bpy)(NO_3)_2$ is also a very efficient binding trigger (Figure 53b). Therefore, it appears that square planar *cis*-Pd(II) complexes present the ideal coordination properties to promote an effective binding to the target DNA.[145]

Notably, similar $PtCl_2$ complexes were ineffective, most probably because of the low reactivity and kinetic stability of the precursor salts (Figure 53c). However, electrophilic Au(III) complexes, particularly $[AuCl_2(bpy)]+$, which exhibit a square planar geometry similar to that of Pd(II), can also promote the DNA binding (Figure 53d). This result is quite appealing as gold complexes had never been used in related bio-supramolecular strategies.

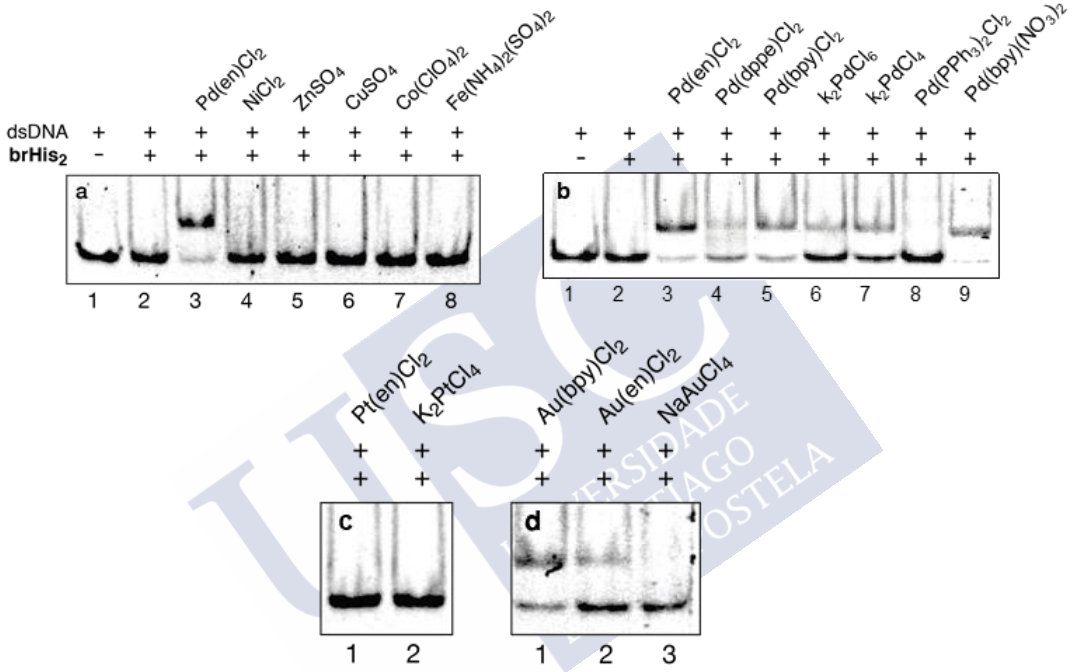


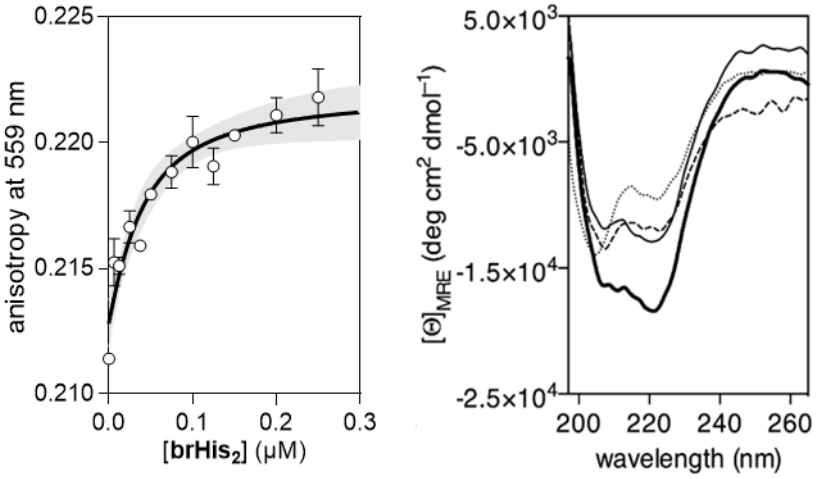
Figure 53. DNA properties of **brHis$_2$** with different metal salts. Lane 1a-b: 75 nM ***GTCAT***; Lane 2a-b: 75 nM ***GTCAT***, 2 μM **brHis$_2$**; All other lanes: 75 nM ***GTCAT***, 2 μM **brHis$_2$**, 20 μM metal complexes (same conditions in all the figures). Experimental conditions as indicated in the caption of Figure 50**.** Note: the gold dichloride complexes present a positive charge.

Spectroscopic characterization

In order to obtain more information on the DNA binding process, we performed fluorescence anisotropy titrations by adding **brHis$_2$** to a solution containing 20 μM *cis*-$PdCl_2$(en) and a tetramethylrhodamine (TMR)-labeled dsoligonucleotide containing the target sequence. The addition of **brHis$_2$** led to a progressive increase in the fluorescence anisotropy of the TMR label. The data could be fitted to a simplified 1:1

[145] For examples of metal-promoted related peptide folding, see: a) M. R. Ghadiri, A. K. Fernholz, *J. Am. Chem. Soc*. **1990**, *112*, 9633-9635; b) N. E. Shepherd, H. N. Hoang, G. Abbenante, D. P. Fairlie, *J. Am. Chem. Soc*. **2005**, *127*, 2974-2983; c) M. J. Kelso, H. N. Hoang, W. Oliver, N. Sokolenko, D. R. March, T. G. Appleton, D. P. Fairlie, *Angew. Chem. Int. Ed Engl.* **2003**, *42*, 421-424; d) S. J. Smith, R. J. Radford, R. H. Subramanian, B. R. Barnett, J. S. Figueroa, F. A. Tezcan, *Chem. Sci.* **2016**, *7*, 5453-5461.

binding model by formally considering the complex [(**brHis$_2$**)(Pd(en))]$^{2+}$ as a single DNA binding unit, with an apparent $K_D \approx 24 \pm 17$ nM (Figure 54, left).



*TMR-**GTCAT:*** 5′-(TMR)-CGC GTCAT AATTGAGAGCGC-3′

Figure 54. *Top Left*: Representative fluorescence anisotropy titration of a 25 nM solution of the TMR-labeled target ds-oligonucleotide TMR-***GTCAT*** in the presence of 20 μM cis-$PdCl_2$(en) and increasing concentrations of **brHis$_2$**. The best fit to a 1:1 binding model is also shown. Measurements were made in Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl. Data correspond to the mean of three independent titration experiments. *Top right*: CD of a 5 μM solution of **brHis$_2$** (dotted line), after the subsequent addition of 1 equiv. of the target dsDNA (dashed line), of the same solution after the addition of *cis*-$PdCl_2$(en) (thick solid line) and after addition of DEDTC demonstrating the reversibility (solid line). The experiments were carried out at 25 °C, in 10 mM phosphate buffer pH 7.5, 100 mM of NaCl. The contribution of DNA to the CD spectrum have been subtracted for clarity. MRE was calculated with respect to the 24-mer **brHis$_2$**. *Bottom*: Labeled oligonucleotide target sequence used in anisotropy assay (only one strand shown, consensus sites underlined).

As expected for a poorly structured peptide, the circular dichroism (CD) spectrum of **brHis$_2$** presents a relatively weak negative signal at 222 nm, even in the presence of the consensus DNA ***GTCAT*** (Figure 54 right, dashed line). Addition of *cis*-$PdCl_2$(en) to the mixture promoted a significant increase in the negative ellipticity at 222 nm, which is consistent with the folding of the peptide into an α-helix (Figure 54 right, thick solid line).[146] In accordance with the results obtained by EMSA, addition of DEDTC to the mixture resulted in a drastic decrease in the helicity of the peptide (Figure 54 right, solid line), which correlates with the disruption of the DNA complex. Not surprisingly, CD experiments in presence of ***MUT*** did not show any increase in the negative ellipticity even when peptide and $PdCl_2$(en) were present (Figure 55a). In addition, $NiCl_2$ (Figure 55b) or $PtCl_2$(en) (Figure 55c) complexes did not promote the peptide α-helical folding in the presence of the target DNA.

[146] a) X. Wang, W. Cao, A. Cao, L. Lai, *Biophys. J.* **2003**, *84*, 1867-1875.

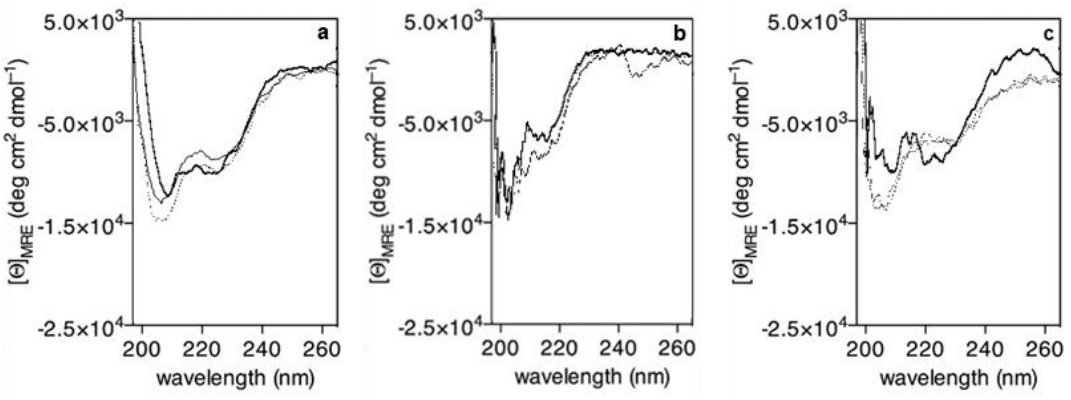


Figure 55. a) CD of a 5 μM solution of **brHis$_2$** (dotted line), after the subsequent addition of *cis*-$PdCl_2$(en) (solid line) and the same solution after the addition of 1 equiv. of dsDNA ***MUT*** (thick solid line). b) CD of a 5 μM solution of **brHis$_2$** (dotted line), after the subsequent addition of $NiCl_2$ (solid line) and the same solution after the addition of 1 equiv. of the target dsDNA (thick solid line). c) CD of a 5 μM solution of **brHis$_2$** (dotted line), after the subsequent addition of $PtCl_2$(en) (solid line) and the same solution after the addition of 1 equiv of the target dsDNA (thick solid line). All experiments were carried out at 25 °C, in 10 mM phosphate buffer pH 7.5, 100 mM of NaCl. The contribution of DNA to the CD spectrum have been subtracted for clarity. MRE was calculated with respect to the 24-mer **brHis$_2$**.

Interestingly, adding *cis*-$PdCl_2$(en) to **brHis$_2$**, in absence of the target DNA, gives rise only to a relatively small increase in helicity (Figure 56a). This result raised the question on whether the palladium is coordinating the peptide in the absence of the DNA template. UV-vis spectra and HPLC-MS analysis of the mixture of the peptide and the palladium complex, were fully consistent with the formation of the expected metal-clipped peptide (see Experimental Data). Furthermore, fluorescence anisotropy experiments carried out by adding *cis*-$PdCl_2$(en) to a solution of a TMR-labeled **brHis$_2$** (the same peptide with a TAMRA fluorophore in the N-terminal, attached by peptidic bond with the last amino acid) resulted in a progressive decrease in the anisotropy of the TMR fluorophore at 559 nm, which agrees with the peptide acquiring a more compact conformation (Figure 56b).

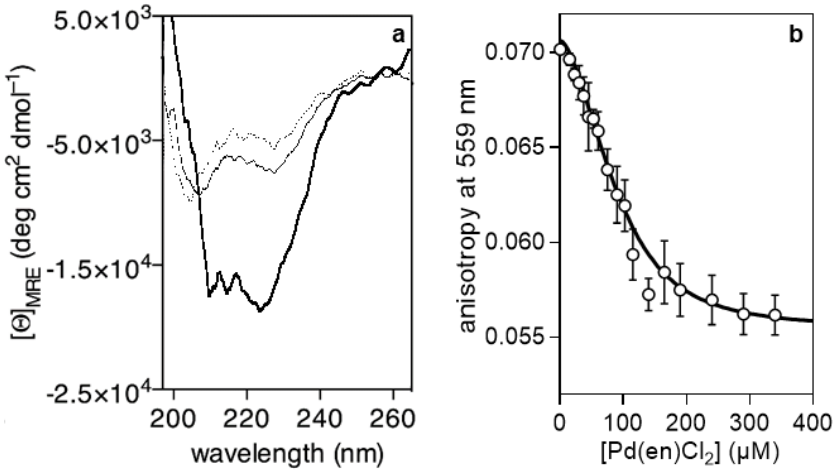


Figure 56. a) CD of a 5 μM solution of **brHis$_2$** (dotted line), after the subsequent addition of *cis*-$PdCl_2$(en) (solid line) and after the addition of 1 equiv. of the target dsDNA (thick solid line)**.** Experimental conditions as indicated in the caption of Figure 55. b) Representative fluorescence anisotropy titration of a 25 nM solution of the TMR-**brHis$_2$** with increasing concentrations of *cis*-$PdCl_2$(en). The best fit to a 1:1 binding model is also shown. We calculated an apparent dissociation constant of 3 μM at 25°C. Measurements were made in Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl. Data correspond to the mean of three independent titration experiments.

The above data indicate that the Pd(II) complex does coordinate the histidine residues, but the peptide only acquires a high proportion of helicity after addition of the target DNA. The palladium clip is not inducing a permanent helical conformation as traditional helix staples,[147,148] but rather it enhances the local helical propensity, which thus facilitates the subsequent folding in the presence of the DNA.[149] Therefore, the system maintains the folding-upon-binding behavior of the parent GCN4 transcription factor, which might be the reason why the [(**brHis$_2$**)(Pd(en))]$^{2+}$ complex shows such an excellent DNA selectivity.[150]

Internalization studies: **brHis$_2$** only translocates in the presence of Pd(II).

A key factor when considering potential biological applications of this type of peptides has to do with their ability to cross cell membranes. It is known that basic peptides derived from transcription factors, such as the Antennapedia can be efficiently internalized.[151]

Moreover, previous studies with covalently stapled peptides suggest that the conformational restriction associated to the stapling favors the cell uptake,[152] and

[147] M. Zhang, B. Wu, H. Zhao and J. W. Taylor, *J. Peptide Sci.* **2002**, *8*, 125-136.

[148] a) F. Bernal, A. F. Tyler, S. J. Korsmeyer, L. D. Walensky, G. L. Verdine, G. L. *J. Am. Chem. Soc.* **2007**, *129*, 2456-2457; b) Y. W. Kim, T. N. Grossmann, G. L. Verdine, *Nat. Protoc.* **2011**, *6*, 761-771.

[149] I. Drobnak, H. Gradišar, A. Ljubetič, E. Merljak, R. Jerala, *J. Am. Chem. Soc.* **2017**, *139*, 8229-8236.

[150] a) P. Tompa, *FEBS Lett.* **2005**, *579*, 3346-3354; b) V. N. Uversky, *J. Biol. Chem.* **2016**, *291*, 6681-6688.

[151] a) D. Derossi, A. H. Joliot, G. Chassaing, A. Prochiantz, *J. Biol. Chem.* **1994**, *269*, 10444-10450; b) D. Derossi, S. Calvet, A. Trembleau, A. Brunissen, G. Chassaing, A. Prochiantz, *J. Biol. Chem.* **1996**, *271*, 18188-18193.

[152] a) H. Zhao, Y. Jiang, Y. Tian, D. Yang, X. Qin, Z. Li, *Org. Biomol. Chem*. **2017**, *15*, 459-464; b) Y. A. Nagel, P. S. Raschle, H. Wennemers, *Angew. Chem. Int. Ed*. **2017**, *56*, 122-126; c) L. Peraro, Z. Zou, K. M. Makwana, A. E.

therefore we were curious to know if the palladium clipping might also influence the cell penetration ability of **brHis$_2$**.

Incubation of mammalian Vero or Hela cells with TMR-**brHis$_2$**, and twofold washing with PBS, revealed that despite the basic character of the peptide, it shows a rather poor internalization.[153] Gratifyingly, and also quite surprisingly, addition of 1 equiv. of *cis*-$PdCl_2$(en) led to the appearance of bright intracellular emission, which was mainly localized in endosomes.[154] Relevantly, neither the TMR-labeled basic region peptide (TMR-**br**), nor the single mutated basic region (TMR-**brHis**) showed appreciable cell translocations upon incubation using similar conditions.

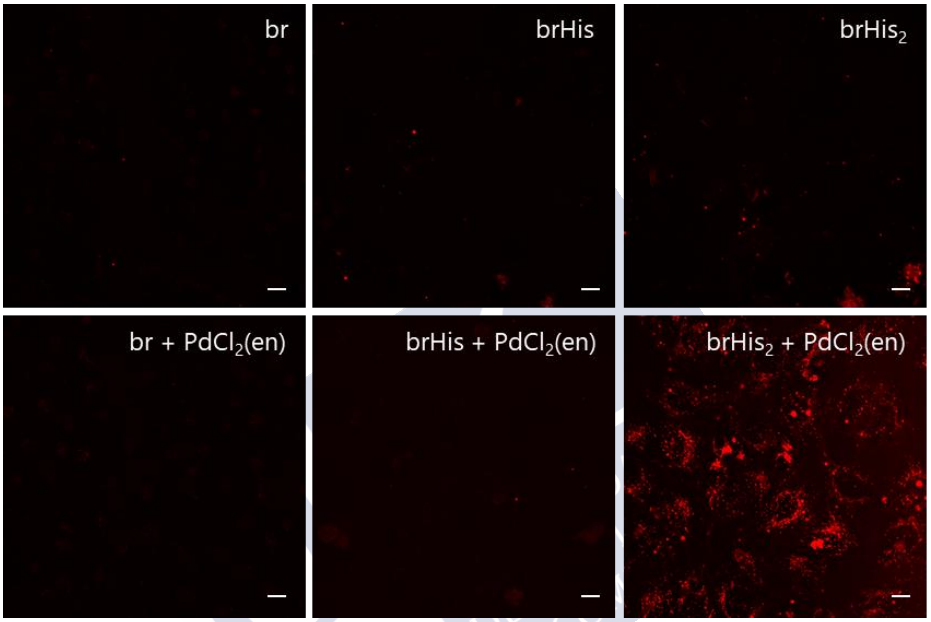


Figure 57. Fluorescence micrographies of Hela cells (confocal). *Top:* incubation with 5μM of peptides for 30 min at 37 ºC. *Bottom*: Peptides were pre-mixed with $PdCl_2$(en) (1:1 ratio) in water for 10 min and incubated with cells for 30 min at 37 ºC. The cells were washed twice with PBS before observed under the microscope. All Incubations were made in Dulbecco's Modified Eagle Medium (DMEM) completed with 5% Fetal Bovine Serum (FBS). Red channel: λexc = 561 nm, λem = 620/60 nm. Scale bar: 10 μm. For Vero Cells internalization experiments see Experimental Data.

It is important to note that cell viability tests (MTT in HeLa Cells, 24 h) demonstrated that none of the peptide**,** metallopeptide nor palladium complex were toxic below 50 μM (more than 90% viability after 24 h).

Cummings, H. L. Ball, H. Yu, Y. S. Lin, B. Levine, J. A. Kritzer, *J. Am. Chem. Soc.* **2017**, *139*, 7792-7802; d) G. Lättig-Tünnemann, M. Prinz, D. Hoffmann, J. Behlke, C. Palm-Apergi, I. Morano, H. D. Herce, M. C. Cardoso, *Nat. Commun.* **2011**, *2*, 453-459; e) A. Iyer, D. Van Lysebetten, Y. Ruiz García, B. Louage, B. G. De Geest, A. Madder, *Org. Biomol. Chem.* **2015**, *13*, 3856-3862.

[153] a) F. Milletti, *Drug Discov. Today.* **2012**, *17*, 850-860; b) B. Gupta, T. S. Levchenko, V. P. Torchilin, *Adv. Drug Deliv. Rev.* **2005**, *57*, 637-651; c) E. A. Goun, T. H. Pillow, L. R. Jones, J. B. Rothbard, P. A. Wender, *ChemBioChem* **2006**, *7*, 1497-1515; d) J. Mosquera, M. I. Sánchez, J. Valero, J. de Mendoza, M. E. Vázquez, J. L. Mascareñas, *Chem. Commun*. **2015**, *51*, 4811-4814.

[154] For strategies to promote endosomal escapes, see: a) A. Erazo-Oliveras, N. Muthukrishnan, R. Baker, T. Y. Wang, J. P. Pellois, *Pharmaceuticals* **2012**, *5*, 1177-1209.

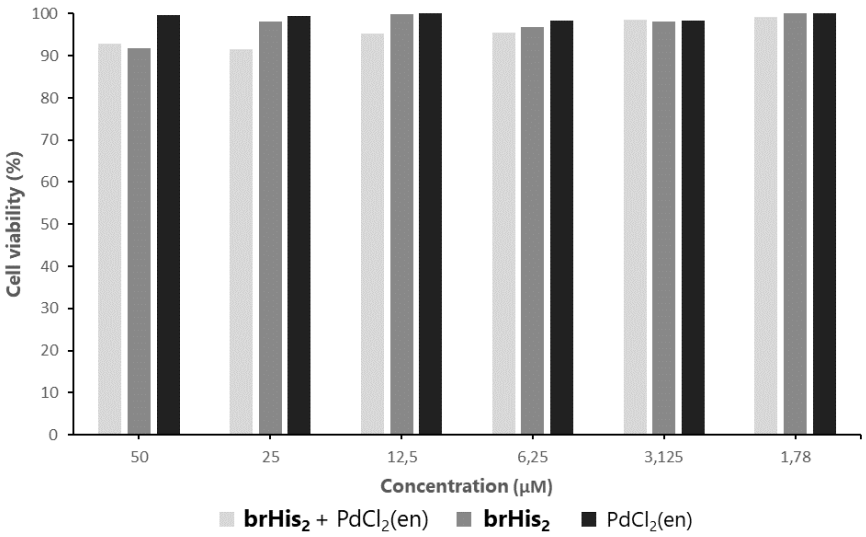


Figure 58. Toxicity assay. **brHis$_2$** was mixed with $PdCl_2$(en) (1:1 ratio) in water for 10 min before the addition to cells. Then, HeLa cells were incubated with this metallopeptide, the peptide or only with $PdCl_2$(en) at different concentrations, for 24 h at 37 ºC. Finally, the cells were washed two times with PBS before carrying out the MTT assays. Data are mean for experimental repeated in three independent times.

Although unraveling the exact structural requirements for the cell uptake switch requires further work, these early results indicate that the internalization requires the metal clip. The underlying internalization mechanism is also unknown, but it is likely that the clipping effect of the palladium promotes a change in the amphiphilicity of the system which favors the membrane crossing. Therefore, the above results represent the first examples of a metal-triggered cell internalization of designed basic peptides.

Conclusions

In summary, we have introduced a new approach for achieving a highly selective recognition of dsDNA with a minimal (23 amino acids long), synthetic peptide which is entirely made of natural residues. Key for the success of the approach is the introduction of histidines in suitable *i*, *i*+4 positions in the sequence, and the addition of square-planar, Pd(II) or Au(III) reagents. In contrast with other covalently-stapled α-helical peptides, our palladium-peptide complex is mostly unstructured in solution, and acquires a predominantly helical conformation only in the presence of the target DNA sequence. The kinetic lability of the metal coordination facilitates the disassembly of the supramolecular structure upon addition of external agents that sequester the palladium.

Importantly, the DNA binding and disassembly processes can be repeated several times, therefore providing a fully reversible switchable system. Additionally, we show that while the apo-peptide presents a very poor cellular uptake, the addition of the palladium reagent prompts an efficient internalization. This is the first demonstration of triggering the cell penetration of a short, basic peptide using a metal complex.

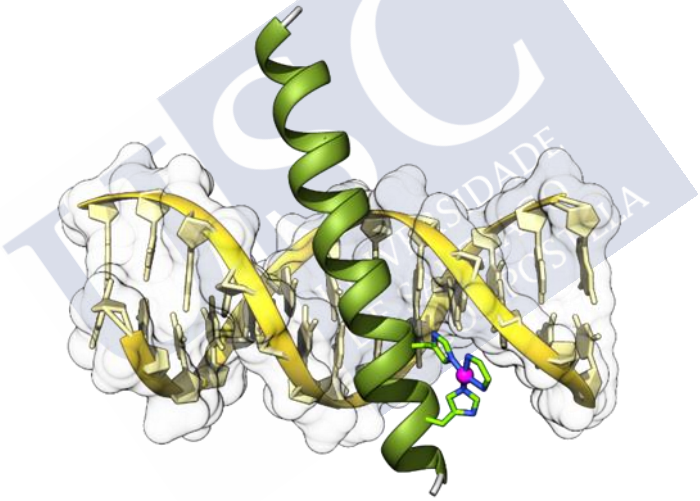


Figure 59. Schematic illustration of the hypothetical DNA recognition of GCN4 basic region, in presence of a metal clip.

General Conclusions

The design of non-native DNA binding agents that can specifically recognize extended sites in the genome and therefore mimic the interaction of naturally occurring transcription factors, is a long-standing goal in bio-supramolecular chemistry, and a major challenge in research at the frontier between Chemistry and Biology. Natural DNA binding proteins bind long sites by presenting several DNA binding modules connected in suitable, well-defined manners or involving **non-covalent** cooperative association. Unfortunately, **multimodular polypeptides** that can emulate these types of recognition are challenging to make and have not been truly developed.

Inspired by the modularity of natural transcription factors, we have designed synthetic miniproteins that combine zinc finger module of transcription factor GAGA with AT-hook peptide domains. These constructs are capable of binding to composite DNA sequences of up to 14 base pairs with high affinity and good selectivity. In particular, we have synthesized **three different chimeras** and characterized their DNA binding properties by electrophoresis and fluorescence anisotropy. Remarkably, we have demonstrated that these artificial DNA binders can be incorporated into expressed miniproteins as sequence specific **artificial gene activators**. We have also used, for the first time in the field, **nanopore force spectroscopy** to obtain further data on the DNA interaction.

In addition, we have also demonstrated for the first time the viability challenging **tripartite DNA binding**, involving the simultaneous formation of **two metal coordination** complexes. We reported a Ni(II)-dependent self-assembly of three peptides at DNA sites of up to 13 bp. A key element in our design is the use of an AT-hook peptidic domain equipped with two bipyridine chelating units. In the final complex, this hook is inserted in the minor groove at the center of the target DNA site, and recruits two bZIP basic regions to adjacent major grooves in response to the addition of Ni(II) complexes. Moreover, we also show that the supramolecular peptide assembly can be easily taken apart by adding a nickel chelating agent.

The development of **stimuli-responsible** sequence-specific DNA recognition agents is a topic of fundamental scientific interest with potential far-reaching impact in biomedical sciences. As a result, for many years there have been many attempts to developing miniature proteins that mimic the DNA properties of natural TFs. We have presented the **smallest peptide** so far developed that is capable of binding to specific sites of the DNA with affinity and selectivity like that of natural DNA binding proteins. More importantly, the DNA binding depends on the presence of a palladium complex that favors the α-helix nucleation required for the interaction. This property allows switching the interaction in a fully **reversible** and cyclic manner, by playing with controlled amounts of the palladium catalysts and a metal sequestering agent. We also demonstrate that gold(III) chlorides can also promote the folding-binding event, which represents the first application of gold in this type of metallo-supramolecular chemistry. In addition to these important novelties, we have demonstrated for the first time the possibility of triggering the **cell internalization** of designed basic peptides by adding a metal clipping agent.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter II: Controlling Oncogenic KRAS Signaling Pathways in Living Cells with a Coordination Staple

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Introduction

RAS proteins are oncogenic supervillains responsible of almost one third of all cancers; with mutations in the family of *RAS* genes found in almost half of all cancers.[155] These mutations drive some of the most lethal cancer types, including about 95% of pancreatic, 45% of colorectal cancers, and 33% of lung adenocarcinomas.[156] More specifically, KRAS, the most frequently mutated isoform, is responsible for most RAS cancers and generates the more aggressive phenotypes with the worse prognosis.[157] The relevance of KRAS makes of this protein a key target in the search of new antitumor agents and has stimulated an extensive search for small-molecule inhibitors for over three decades.[158] Unfortunately, RAS proteins have resisted most efforts aimed at the development of small-molecule inhibitors to the point that, until recently, they were considered undruggable targets.[159]

Physiologically, RAS proteins are binary molecular switches that oscillate between an inactive, GDP-bound state, and an active RAS-GTP complex. This switch is aided by specific guanine nucleotide exchange factor proteins (GEFs), such as SOS1, which stimulate the GDP/GTP exchange and activation, and GTPase-activating proteins (GAPs) that catalyze the hydrolysis of GTP and inactivation in the RAS-GDP state. Certain mutations compromise the GTPase activity of RAS proteins, locking them in the active state, inducing the aberrant activation of several signaling pathways, ultimately leading to uncontrolled cell growth and proliferation, invasiveness, angiogenesis, and metastasis (Figure 60).[160] Activation of RAS-GDP by SOS1 is induced by the insertion of the αH helix of the exchange factor, which displaces the switch I region of RAS, opening up the GTP binding pocket, and distorts the switch II region reducing the affinity for the nucleotide.[161] It has been shown that stabilized α-helical peptides based on the SOS1 αH helix bind KRAS with high affinity and competitively inhibit RAS activation by SOS1. These stabilized versions include hydrogen bond surrogates,[162] as well as classic stapled

[155] F. Sanchez-Vega, M. Mina, J. Armenia, W. K. Chatila, A. Luna, K. C. La, S. Dimitriadoy, D. L. Liu, H. S. Kantheti, S. Saghafinia, D. Chakravarty, F. Daian, Q. Gao, M. H. Bailey, W. W. Liang, S. M. Foltz, I. Shmulevich, L. Ding, Z. Heins, A. Ochoa, B. Gross, J. Gao, H. Zhang, R. Kundra, C. Kandoth, I. Bahceci, L. Dervishi, U. Dogrusoz, W. Zhou, H. Shen, P. W. Laird, G. P. Way, C. S. Greene, H. Liang, Y. Xiao, C. Wang, A. Iavarone, A. H. Berger, T. G. Bivona, A. J. Lazar, G. D. Hammer, T. Giordano, L. N. Kwong, G. McArthur, C. Huang, A. D. Tward, M. J. Frederick, F. McCormick, M. Meyerson, Cancer Genome Atlas Research Network, E. M. Van Allen, A. D. Cherniack, G. Ciriello, C. Sander, N. Schultz, *Cell* **2018**, *173*, 321-337.e10.

[156] S. A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, et al., *Nucleic Acids Res.* **2015**, *43*, D805-D811.

[157] G. A. Hobbs, C. J. Der, K. L. Rossman, *J. Cell Sci.* **2016**, *129*, 1287-1292.

[158] a) H. Ledford, Cancer: The Ras Renaissance. *Nature* **2015**, *520* (7547), 278-280; b) J. Spiegel, P. M. Cromm, G. Zimmermann, T. N. Grossmann, H. Waldmann, *Nat. Chem. Biol.* **2014**, *10*, 613-622.

[159] A. R. Moore, S. C. Rosenberg, F. McCormick, S. Malek, *Nat. Rev. Drug Discov.* **2020**, *19*, 533-552.

[160] a) A. Wittinghofer, H. Waldmann, *Angew. Chem. Int. Ed.* **2000**, *39*, 4192–4214; b) K. Rajalingam, R. Shreck, U. R. Rapp, S. Albert, *Biochim. Biophys. Acta* **2007**, *1773*, 1177-1195.

[161] S. M. Margarit, H. Sondermann, B. E. Hall, B. Nagar, A. Hoelz, M. Pirruccello, D. Bar-Sagi, J. Kuriyan, *Cell* **2003**, *112*, 685-695.

[162] a) A. Patgiri, K. K. Yadav, P. S. Arora, D. Bar-Sagi, *Nat. Chem. Biol.* **2011**, *7*, 585-587; b) D. Y. Yoo, A. D. Hauser, S. T. Joy, D. Bar-Sagi, P. S. Arora, *ACS Chem. Biol.* **2020**, *15*, 1604-1612.

peptides featuring hydrocarbon linkers between solvent-exposed *i,i+4* residues in the αH helix.[163]

Following our recent work on the application of metal chelation for modulating the properties of bioactive peptides,[164] we questioned whether this approach could be applied for the development of a stimuli-responsive KRAS peptide inhibitor. Specifically, we set out to explore whether a *i,i+4* His chelation with Pd(II) is an effective means of dynamically controlling the conformation and binding properties of short α-helical peptides that could inhibit RAS activation by SOS1 protein.

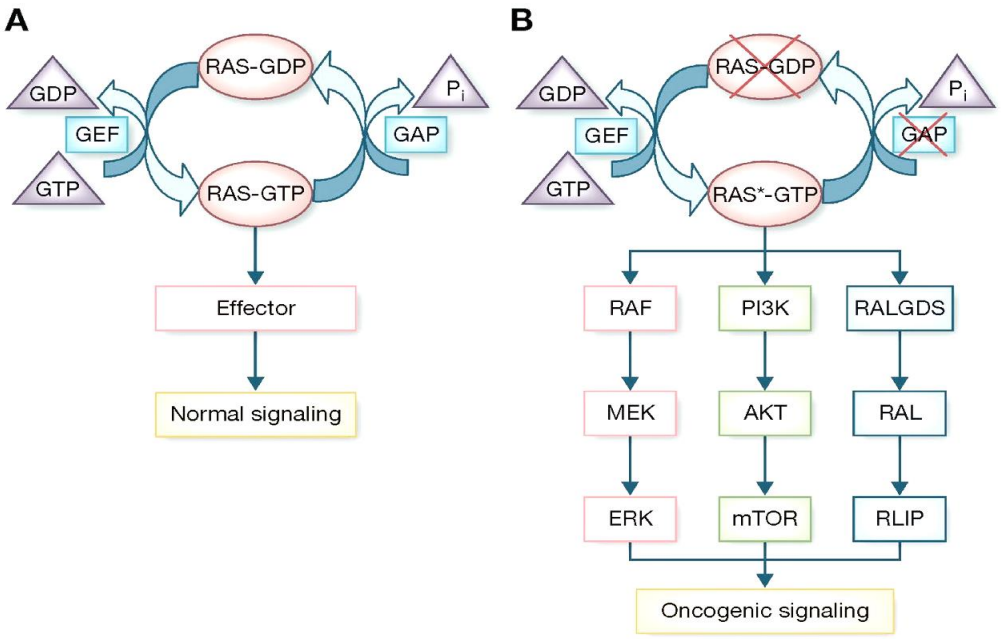


Figure 60. a) Normal Ras signaling. b) Oncogenic Ras signaling. When Ras is mutated, it is constitutively bound to GTP such that its GAP cannot bind. The activated Ras signals through a multitude of effectors and downstream signaling pathways, a subset of which is shown here. (Source: Adapted from [165]).

[163] E. S. Leshchiner, A. Parkhitko, G. H. Bird, J. Luccarelli, J. A. Bellairs, S. Escudero, K. Opoku-Nsiah, M. Godes, N. Perrimon, L. D. Walensky, *Proc. Natl. Acad. Sci.* **2015**, *112*, 1761-1766.

[164] a) J. Rodríguez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *Chem. Eur. J.* **2016**, *22*, 13474-13477; b) J. Rodríguez, J. Mosquera, R. García-Fandiño, M. Eugenio Vázquez, J. L. Mascareñas, *Chem. Sci.* **2016**, *7*, 3298-3303. c) C. Penas, J. L. Mascareñas, M. E. Vázquez, M. E. *Chem. Sci.* **2016**, *2016*, 2674-2678.

[165] N. Vasan, J. L. Boyer, R. S. Herbst, *Clin. Cancer Res.* **2014**, *20*, 3921- 3930.

Objective

We propose the design and synthesis of peptides based on SOS1 helix whose RAS inhibitory activity could be controlled with metals. For that, an equivalent SOS1 peptide featuring His residues in *i, i+4* positions should be able to coordinate to palladium (or other metal) complexes, forming a metallopeptide capable of binding the KRAS protein with high affinity. In contrast to the previous reports with hydrocarbon stapled peptides, our strategy could allow to dissemble the complex with a metal chelator, and consequently, control the KRAS activation/inhibition process in a switchable manner. In addition, considering that the stapled peptide might be able to cross cell membranes with good efficiency, the metallopeptide may interact with KRAS not only *in vitro* but also inside living cells.

Results and discussion

Design and synthesis

A number of SOS1 αH peptidomimetics,[163] as well as details on their interaction with KRAS have been described.[166] Based on this information, we chose the fragment comprising SOS1 residues Phe[929] to Asn[944], which spans the full αH helix. Inspection of an hypothetical αH/RAS complex revealed that positions Tyr[933] and Ile[937] are located in the solvent-exposed face of the αH helix fragment, and when replaced by His residues, the resulting side chains would be in a suitable register to chelate a Pd(II) ion (Figure 61). In addition to include these Y933H and I937H mutations, we also decided to extend the natural αH sequence at its N-terminus with two Arg residues in order to increase solubility, and cap the amino terminal group with 6-carboxytetramethylrhodamine to favor quantification and monitoring by using fluorescence (Figure 61, **αH-His$_2$**).

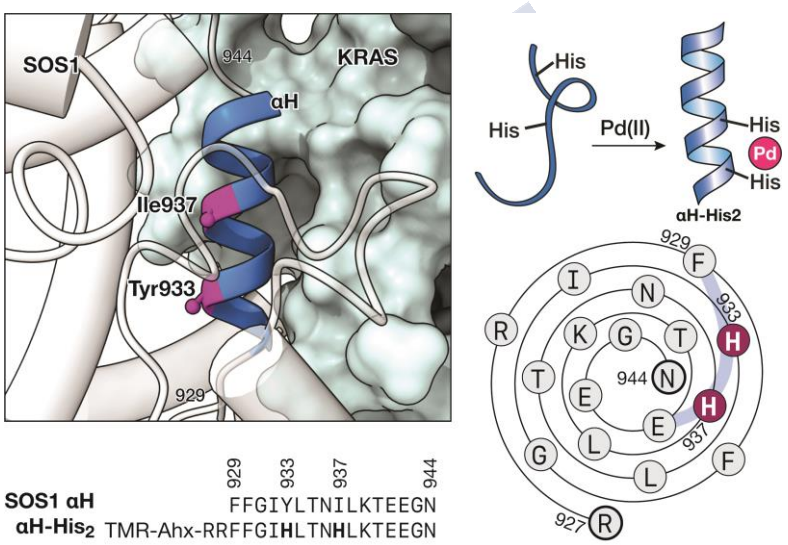


Figure 61. *Top left:* Detail of the αH helix of SOS1 complexed with RAS (PDB code 1BKD) indicating the residues Tyr[933] and Ile[937] that are replaced by the metal chelating His residues. *Bottom left*: Sequence of the natural αH helix of SOS1 and of the mutant peptide αH-His$_2$ used in this study, highlighting the two metal-chelating His residues. *Top right:* Helical stabilization by Pd(II) chelation. *Bottom right*: Wenxiang diagram of the **αH-His$_2$**,[167] indicating the residues aligned in the outer face of the αH helix, including the two His residues in consecutive turns this α-helix.

The chelating peptides were synthesized following standard Fmoc- solid-phase peptide synthesis methods on an automated microwave synthesizer.[168] The peptides have the C-terminus amidated due to the use of a *Rink amide* resin, and in the N- terminal have an

[166] a) S. M. Margarit, H. Sondermann, B. E. Hall, B. Nagar, A. Hoelz, M. Pirruccello, D. Bar-Sagi, J. Kuriyan, *Cell* **2003**, *112*, 685-695 (PDB code 1NVW); b) P. A. Boriack-Sjodin, S. M. Margarit, D. Bar-Sagi, J. Kuriyan, J. *Nature* **1998**, *394*, 337-343 (PDB code 1BKD).

[167] a) K. C. Chou, C. T. Zhang, G. M. Maggiora, *Proteins* **1997**, *28*, 99-108; b) K. C. Chou, W. Z. Lin, X. Xiao, *New Solut.* **2011**, *03*, 862-865.

[168] I. Coin, M. Beyermann, M. Bienert, *Nat. Protoc.* **2007**, *2*, 3247-3256.

amino hexanoic acid as a linker between the last arginine and the TAMRA fluorophore (see Experimental Data for further information).

Spectroscopic characterization

Having the desired peptides at hand, we performed circular dichroism studies as initial step. CD assays showed that metal chelation had a modest effect in the α-helical content of the peptide **αH-His$_2$**. Thus, the CD of a solution of **αH-His$_2$** displayed the signature of a random coil peptide and a low intensity band at 222 nm (≈ –2 200 deg cm^2 dmol^{-1}). Incubation with one equivalent of $PdCl_2$(en) for 30 min induced only a slight increase of the negative intensity of this band (≈ –6 000 deg cm^2 dmol^{-1}), which suggests a small α-helical content of about 22% in the metallopeptide **αH-His$_2$**[Pd] (Figure 62a).[169] Even when we performed the experiment adding increasing equivalents of metal complex (without incubation time) the helicity only slightly increased until ≈ –7 000 deg cm^2 dmol^{-1} (Figure 62b). Coordination of the Pd(II) ion was confirmed by MS (see Experimental Data).

Addition of a Pd(II) chelator, diethyldithiocarbamate (DEDTC), to the above metallopeptide complex (**αH-His$_2$**[Pd]), led to a significant reduction in helicity (Figure 62c), in consonance with the removal of the Pd-staple. A new addition of the Pd salt led to a recovery of the helicity. Therefore, the system seems to be easily switchable between the metallo-stapled and the non-chelated forms.

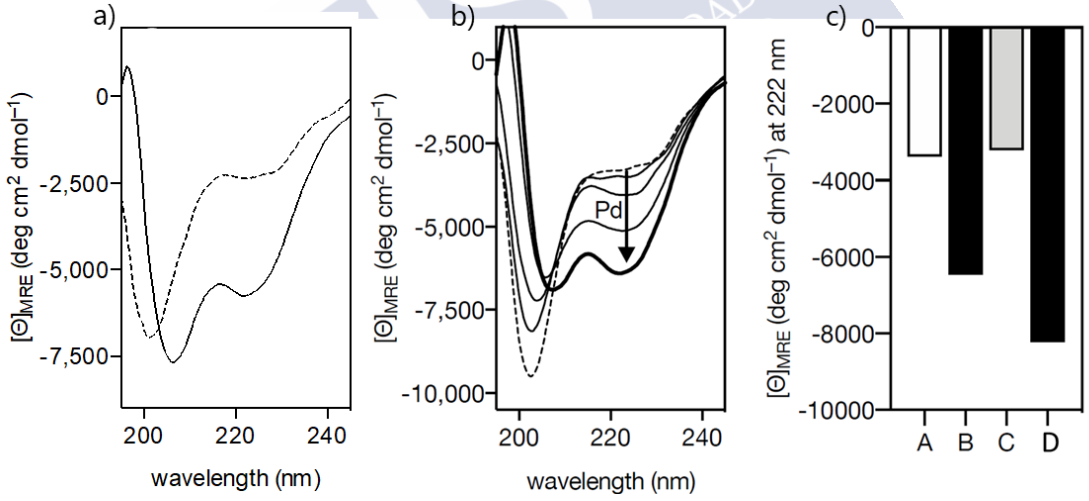


Figure 62. Circular dichroism experiments. a) Circular dichroism of a 20 μM solution of **αH-His$_2$** (dashed line) and the spectrum 30 min after addition of 1 eq. of $PdCl_2$(en). b) CD titration experiment. Circular dichroism of a 20 μM solution of **αH-His$_2$** (dashed line) and addition of 1, 2, 5 and 10 equivalents of $PdCl_2$(en), measured directly after the addition. c) CD switch experiment. Circular dichroism values at 222nm for: A) 20 μM solution of **αH-His$_2$**; B) 20 μM of metallopeptide **αH-His$_2$**[Pd]; C) same solution after addition of 5 eq. of DEDTC; D) solution in C after addition of 20 eq. of $PdCl_2$(en). The experiments were carried out in 10 mM phosphate buffer pH 7.5, 100 mM of NaCl ad 10% TFE, 25 ºC. Mean residue molar ellipticity ($[\Theta]_{MRE}$) was calculated considering an 18-mer.

[169] We have used the empirical *k* factor ≈ 4 that allows the reliable application of CD to characterize the α-helical content of short peptides as reported in N. E. Shepherd, H. N. Hoang, G. Abbenante, D. P. Fairlie, *J. Am. Chem. Soc.* **2005**, *127*, 2974-2983.

Although the coordination to Pd(II) ions apparently had a relatively little effect on the conformation of the peptide **αH-His$_2$** in solution, we hypothesized that it still might cooperatively favor its folding when coupled to the KRAS recognition, so we decided to compare the binding of **αH-His$_2$** and its complex **αH-His$_2$**[Pd] to KRAS using fluorescence anisotropy.

As anticipated, mixing a 15 nM solution of the unstructured peptide **αH-His$_2$** in absence of Pd(II), with increasing concentrations of KRAS^{wt} did not produce any increase in the fluorescence anisotropy of the TAMRA emission at 559 nm, which is consistent with the expected low affinity of **αH-His$_2$** for KRAS^{wt} (Figure 63a, white dots). In contrast, when the titration was repeated with the metallocomplex **αH-His$_2$**[Pd], we observed a marked increase in the anisotropy consistent with binding to KRAS^{wt} and formation of a larger complex with slower rotational diffusion. The increase in anisotropy could be fitted to a simple 1:1 binding model with an approximate dissociation constant of 240 nM (Figure 63a, black dots). Furthermore, we performed the same experiment with oncogenic mutants of KRAS protein. As we can see in the figure, similar titration profiles were obtained when **αH-His$_2$**[Pd] was incubated with the oncogenic mutants KRAS^{G12V} and KRAS^{G12C}, which displayed also tight binding with dissociation constants of 345 nM and 294 nM, for KRAS^{G12V} and KRAS^{G12C}, respectively (Figure 63b-c). As we saw before with the wild type KRAS, **αH-His$_2$** is not functional, as we did not see any increase in the fluorescence when mixed with the RAS proteins.

Importantly, once a maximum and stable anisotropy value was obtained after mixing **αH-His$_2$**[Pd] with KRAS^{wt}, we added DEDTC to the solution, which resulted in a marked decrease of the rhodamine anisotropy, consistent with the disassembly of the metallopeptide **αH-His$_2$**[Pd], and disengagement of the interaction of the peptide with the RAS protein (Figure 63d). Furthermore, addition of fresh $PdCl_2$(en) to that solution restored the original anisotropy, indicating that the released peptide **αH-His$_2$** was still able to chelate Pd(II) ions and bind once again as **αH-His$_2$**[Pd] to KRAS^{wt}, which proves once more the reversible switching capacity of the system.

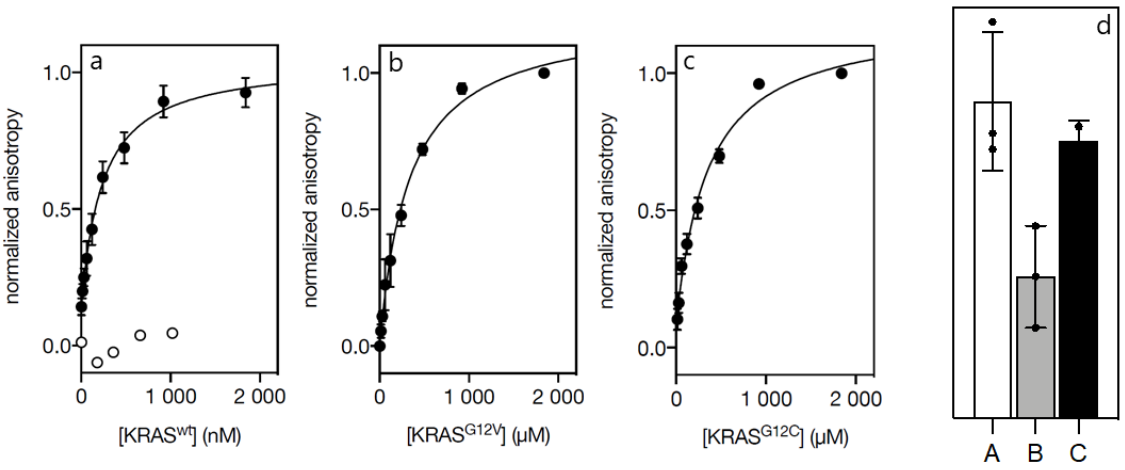


Figure 63. Anisotropy titration with KRAS^{wt} and oncogenic mutants KRAS^{G12C} and KRAS^{G12V}. Normalized emission anisotropy at 559 nm of a 15 nM solution of **αH-His$_2$**[Pd] with increasing concentrations of a) KRAS^{wt}, b) KRAS^{G12V} and c) KRAS^{G12C}. The best fit to a 1:1 binding model is also shown;[170] d) anisotropy values at 559 nm for: A, a saturated **αH-His$_2$**[Pd]/KRAS complex; B, same solution after addition of 50 eq. of DEDTC; C, solution in B after addition of 50 eq. of $PdCl_2$(en). Data are mean ± SEM of three independent experiments performed in triplicate, with independent preparations of recombinant KRAS^{wt}. All experiments carried out at 25 °C.

These data suggest that the palladium clip is working as a facilitator of the helical folding rather than as a true helix staple in which the covalent link permanently fixes the bioactive conformation.[171] In other words, the coordination to Pd(II) appears to provide just enough thermodynamic stabilization so that the peptide displays folding-upon-binding properties, typically observed for intrinsically disordered proteins.[172] Likely, the region close to the metal clip might be highly structured despite the low total helical content of the **αH-His$_2$**[Pd] metallopeptide detected by Circular dichroism, and this preorganization is critical to facilitate the interaction with the target protein.[173]

Nucleotide release process study

Having confirmed the high-affinity of the complex **αH-His$_2$**[Pd] for KRAS^{wt}, we studied its effect in the nucleotide release process required for KRAS^{wt} activation.[174] Therefore, we first measured the intrinsic nucleotide release rate by monitoring the fluorescence emission of the complex of KRAS^{wt} with the GTP analog mantGTP, 2′-/3′-*O*-(*N*-methylanthraniloyl) guanosine-5′-triphosphate, which displays low fluorescence in

[170] Experimental data were fitted with the DynaFit 4.0 software, which performs a numerical treatment of the system: a) P. Kuzmič, *Anal. Biochem.* **1996**, *237*, 260-273; P. Kuzmič, in *Methods Enzymol.*, Academic Press, **2009**, pp. 247-280.

[171] a) F. Bernal, A.F. Tyler, S. J. Korsmeyer, L. D. Walensky, G. L. Verdine, *J. Am. Chem. Soc.* **2007**, *129* (9), 2456-2457; b) P. M. Cromm, J. Spiegel, T. N. Grossmann, *ACS Chem. Biol.* **2015**, *10*, 1362-1375.

[172] a) A. Toto, F. Malagrinò, L. Visconti, F. Troilo, L. Pagano, M. Brunori, P. Jemth, S. J. Gianni, *Biol. Chem.* **2020**, *295*, 6586-6593.

[173] a) R. A. Kammerer, T. Schulthess, R. Landwehr, A. Lustig, J. Engel, U. Aebi, M. O. Steinmetz, *Proc. Natl. Acad. Sci.* **1998**, *95*, 13419-13424; b) D. L. Lee, P. Lavigne, R. S. Hodges, R. S. *J. Mol. Biol.* **2001**, *306*, 539-553.

[174] a) R. J. Rojas, R. J. Kimple, K. L. Rossman, D. P. Siderovski, J. Sondek, *Comb. Chem. High Throughput Screen.* **2003**, *6*, 409-418; b) S. Jeganathan, M. P. Müller, I. Ali, R. S. Goody, *Biochemistry* **2018**, *57*, 4690-4699.

solution, but becomes strongly emissive when associated with KRAS^{wt}.[175] As expected, the KRAS^{wt}/mantGPT complex was stable, and the fluorescent mantGTP was released very slowly, with a half-life of about 7 min (Figure 64). The half-life of this complex was somewhat decreased by the presence of the free peptide **αH-His$_2$** (Figure 64 right, $t_{1/2} \approx$ 5 min), but addition of metallopeptide **αH-His$_2$**[Pd] drastically increased the rate of nucleotide release, so that the measured half-time of the KRAS^{wt}/mantGPT complex was reduced to about 3 min under the same conditions (Figure 64, right). Thus, we find that the binding of the metallopeptide **αH-His$_2$**[Pd] to KRAS^{wt} induces the displacement of the bound mantGTP nucleotide from the preformed complex, affecting the GTPase activity of the protein.

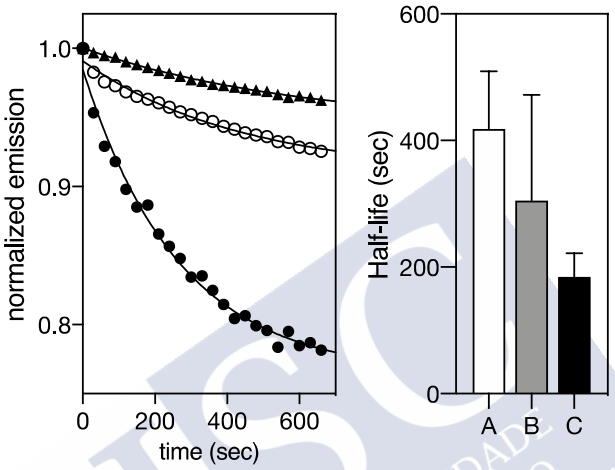


Figure 64. αH-His$_2$[Pd] increases the displacement of mantGTP from KRAS^{wt}. *Left*: normalized emission at 430 nm. Black triangles: intrinsic nucleotide displacement curve of a 1 μM solution of the preformed KRAS^{wt}/mantGTP complex; white circles, same solution in the presence of 10 μM **αH-His$_2$**; black circles: same KRAS^{wt}/mantGTP solution in the presence of 10 μM **αH-His$_2$**[Pd]. *Right*: half-life of the kinetic traces shown on the left. A: intrinsic KRAS^{wt}/mantGTP complex; B: with **αH-His$_2$**; C: with **αH-His$_2$**[Pd].

Mammalian Cells experiments

To assess the capacity of the metallopeptide to antagonize the KRAS activation inside living cells, we first monitored the cellular uptake by fluorescence microscopy. Incubation of lung carcinoma A549 cells with **αH-His$_2$** (10 μM) revealed a rather poor internalization (Figure 65, Top left). Gratifyingly, if this peptide was previously incubated with $PdCl_2$(en), the resulting metallopeptide **αH-His$_2$**[Pd] showed an intense intracellular fluorescence signal (Figure 65, Top right), validating the suitability of this metallopeptide for cell-based assays.

The activation of KRAS leads to stimulation of various signal transduction pathways. To determine whether the inhibitory effect of **αH-His$_2$**[Pd] on KRAS could modulate downstream signaling events, we studied the activation of the MAPK RAF-MEK-ERK cascade, a well-documented KRAS effector pathway implicated in cell proliferation,

[175] T. Hiratsuka, *Biochim. Biophys. Acta* **1983**, *742*, 496-508.

survival, and differentiation.[176] Thus, serum-starved A549 cells[177] were incubated with the peptide **αH-His$_2$** (in absence of palladium) and with the preformed **αH-His$_2$**[Pd] complex for 4 hours, using different concentrations of both (Figure 65, bottom). Then, after washing and lysing of the cells, we measured the levels of ERK phosphorylation by blotting the cell lysates with pERK specific antibodies. We performed a Western blot study, where the samples were first resolved in a SDS-PAGE and then transferred to a Western Blot membrane. The results (Figure 65, bottom) demonstrate that **αH-His$_2$**[Pd] inhibits ERK phosphorylation dose-responsively, while $PdCl_2$(en) or the uncomplexed **αH-His$_2$** does not show any effect in the levels of pERK. The use of an antibody against total ERK confirmed that the variation in pERK levels are due to changes in the phosphorylation state and not in the expression levels of the protein.

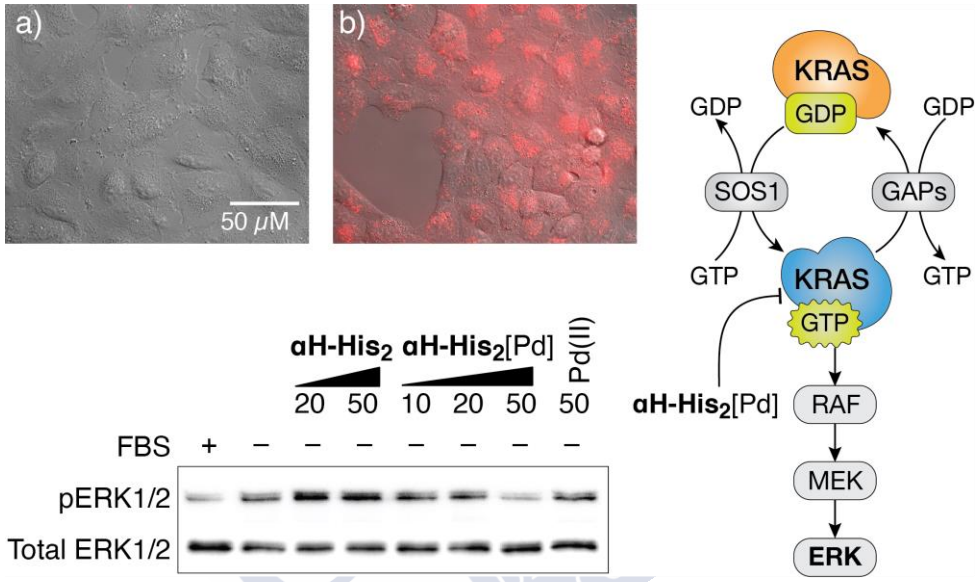


Figure 65. *Top left*: Internalization of **αH-His$_2$**[Pd] in A549 cells. a) Incubation with **αH-His$_2$** (10 μM, 30 min at 37 ºC); b) incubation with the **αH-His$_2$**[Pd] complex, prepared by mixing an equimolar mixture of **αH-His$_2$** and $PdCl_2$(en) in water for 10 min (10 μM, 30 min at 37 ºC). The cells were washed with PBS before being observed under the microscope. Incubations were made in Dulbecco´s modified Eagle medium (DMEM) completed with 5% of fetal bovine serum (FBS-DMEM). λ_{exc} = 550 nm, λ_{em} = 590-650 nm; *Bottom*: **αH-His$_2$**[Pd] inhibits the Raf-MEK-ERK mitogen-activated protein kinase (MAPK) pathway downstream of KRAS in living cells. Serum-starved A549 cells were incubated with **αH-His$_2$**, the metallopeptide **αH-His$_2$**[Pd], or with $PdCl_2$(en) at the indicated doses for 4 h. Cellular lysates were subjected to Western blot analysis by using antibodies to phospho- and total ERK1/2. *Right*: Scheme of the MAPK pathway and the inhibition of the SOS1-mediated activation of KRAS by **αH-His$_2$**[Pd] that leads to reduced phosphorylation of ERK.

[176] P. J. Roberts, C. J. Der, *Oncogene* **2007**, *26*, 3291-3310.

[177] Y. D. Jung, K. Nakano, W. Liu, G. E. Gallick, L. M. Ellis, *Cancer Res.* **1999**, *59*, 4804-4807.

Conclusions

In conclusion, we have shown that metal-based staples, by favouring the alpha-helical folding ability of designed bishistidine containing SOS1-based peptides, can drastically influence the interaction of these peptides with RAS proteins. This interaction can be switched in a reversible manner by using palladium chelating agents. Importantly, the resulting metallopeptides are functionally active, and behave as potent RAS inhibitors.

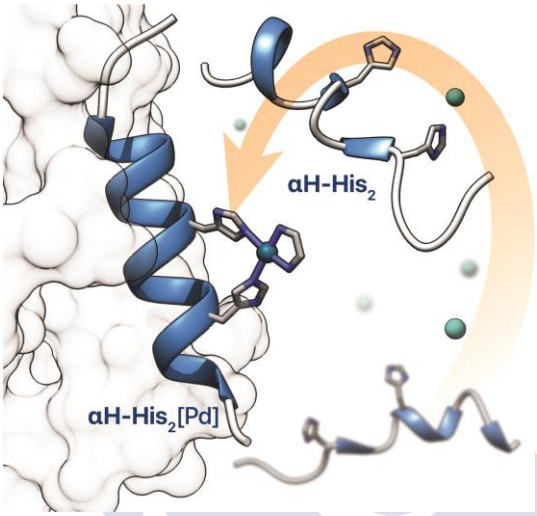


Figure 66. Schematic illustration of **αH-His$_2$**[Pd] binding to KRAS protein.

Chapter III: Intracellular Reactions Promoted by Bis-histidine Miniproteins Stapled with Pd(II) Complexes

This chapter includes work published in *Angewante Chemie International Edition* as:

S. Learte-Aymamí, C. Vidal, A. Gutiérrez-González, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2020**, *59*, 9149-9154.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Introduction

Metalloproteins make up a substantial portion of the human proteome (between one-third and one-half) and play critical structural and catalytic role in living cells and whole organisms.[178] Catalytic metalloproteins, *i.e. metalloenzymes*, are involved in many fundamental metabolic reactions, including, among others, redox processes, isomerizations, hydrolysis or condensations.[2] Interestingly, despite the large number of alternatives offered by the periodic table, Nature has evolved to use just a few metals as cofactors to build the palette of metalloenzymes required to sustain life. Indeed, most metal-based proteins feature either alkali and alkaline earth elements, or first row transition metals like Fe, Cu, Mn, Co or Zn.[179,180]

The recognition that introducing other transition metals like Ni, Ru, Rh, Pd or Ir at designed sites of proteins might provide for non-natural transformations with enzymatic characteristics has led to very significant advances in the field of "artificial metalloenzymes".[181] Therefore, a number of transition metal protein hybrids that promote otherwise difficult reactions have been recently reported.[182] Most successful results involve the use of prosthetic groups as metal coordinating units, albeit some examples in which the artificial metalloproteins result from direct metal coordination to amino acid side chains have also been described.[183]

Whereas these efforts are ascribed to the realm of synthetic chemistry, a challenging, remaining goal in the field consists of the introduction or generation of artificial metalloenzymes in living mammalian cells, which is the native environment of many natural enzymes.[184] Not surprisingly, progress in this matter has been extremely slow,

[178] a) D. H. Nies, S. Silver in *Molecular Microbiology of Heavy Metals*, Springer-Verlag, Berlin, **2007**; b) H. Zheng, M. Chruszc, P. Lasota, L. Lebioda, W. Minor, *J. Inorg. Biochem.* **2008**, *102*, 1765-1776; c) K. J. Waldron, N. J. Robinson, *Nat. Rev. Microbiol.* **2009**, *7*, 25-35; d) K. J. Waldron, J. C. Rutherford, D. Ford, N. J. Robinson, *Nature* **2009**, *460*, 823-830.

[179] a) C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton, *J. Biol. Inorg. Chem.* **2008**, *13*, 1205-1218; b) A. Pordea, *Curr. Opin. Chem. Biol.* **2015**, *25*, 124-132; c) H. Eom, W. J. Song, *J. Biol. Inorg. Chem.* **2019**, *24***,** 517-531.

[180] a) J. C. Lewis, *ACS Catal.* **2013**, *3*, 1-22; b) C. E. Valdez, Q. A. Smith, M. R. Nechay, A. N. Alexandrova, *Acc. Chem. Res.* **2014**, *47*, 3110-3117.

[181] a) P. J. Deuss, R. Denheetenm, W. Laan, P. C. J. Kamer, *Chem. Eur. J.* **2011**, *17*, 4680-4698; b) M. Dürrengberger, T. R. Ward, *Curr. Opin. Chem. Biol.* **2014**, *19*, 99-106; c) T. Heinisch, T. R. Ward, *Eur. J. Inorg. Chem.* **2015**, 2015, 3406–3418.; d) T. K. Hyster, T. R. Ward, *Angew. Chem. Int. Ed.* **2016**, *55*, 7344-7357; e) F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. KÖhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142-231; f) A. D. Liang, J. Serrano-Plana, R. L. Peterson, T. R. Ward, *Acc. Chem. Res.* **2019**, *52*, 585-595.

[182] a) H. J. Davis, T. R. Ward, *ACS Cent. Sci.* **2019**, *5*, 1120-1136; b) T. Matsuo, T. Miyake, S. Hirota, *Tetrahedron* **2019**, *60*, 151226-151233.

[183] a) F. Nastri, D. D´Alonzo, L. Leone, G. Zambrano, V. Pavone, A. Lombardi, *Trend. Biochem. Sci.* **2019**, *44*, 1022-1040; b) S. Studer, D. A. Hansen, Z. L. Pianowski, P. R. E. Mittl, A. Debon, S. L. Guffy, B. S. Der, B. Kuhlman, D. Hilvert, *Science* **2018**, *362*, 1285-1288.

[184] H. Renata, Z. J. Wang, F. H. Arnold, *Angew. Chem. Int. Ed.* **2015**, *54*, 3351-3367.

and essentially limited to one elegant report in the internalization of streptavidin-ruthenium constructs in HEK-293 cells.[185]

This paucity contrasts with the increasing number of reports on the development of bioorthogonal and intracellular reactions promoted by discrete transition metal complexes or metal nanoparticles.[186] However, the complexity of biological media, with many nucleophiles like glutathione (GSH), thiol-containing proteins, nucleobases, etc. that can coordinate the metal and decrease or inhibit the catalytic activity, makes their use also extremely challenging.

For the metal complexes to work properly in a biological environment it is essential to control the orthogonal selectivity, so that they can activate the designed substrate(s), and avoid reactions and even strong interactions with other concurrent functional groups (*e.g.,* nucleophiles and electrophiles) that could compromise the catalytic activity. In other words, the reaction should be *bioorthogonal*.[187] This term was coined by Bertozzi, and it refers to the properties that a chemical reaction must fulfil to be useful in living settings, that is, fast rate and inertness toward the myriad of functionalities present in living environments. In addition, the transition metal reagent has to be nontoxic, highly reactive under very dilute conditions and, ideally, cell-permeable.[188]

One of the earliest examples of a metal-promoted reaction in biological settings dates back to 1985, when a homogeneous water-soluble ruthenium complex incorporated into mesophyll protoplasts was shown to hydrogenate C-C double bonds of fatty acids in the presence of hydrogen gas.[189] However, there were few chemical details in the manuscript, and an extensive cell damage was observed after a moderate period of exposure (1 hour).

It was not until 2002 when this field started to grow, with the discovery of the copper catalyzed azide-alkyne cycloaddition (CuAAC).[190] This reaction had a tremendous impact in chemical biology, and in particular, in the field of bioconjugation, as it is bioorthogonal and compatible with water. In fact, it has been used for the labelling of proteins, nucleic acids, DNA and polysaccharides;[191] it has also been used, although with less success, for

[185] Y. Okamoto, R. Kojima, F. Schwizer, E. Bartolami, T. Heinisch, S. Matile, M. Fussenegger, T. R. Ward, *Nat. Commun.* **2018**, *9*, 1943-1949.

[186] For recent reviews, see: a) M. Martínez-Calvo, J. L. Mascareñas, *Coord. Chem. Rev.* **2018**, *359*, 57-79; b) A. H. Ngo, S. Bose, L. H. Do, *Chem. Eur. J.* **2018**, *24*, 10584-10594; c) Y. Bai, J. Chen, S. C. Zimmerman, *Chem. Soc. Rev.* **2018**, *47*, 1811-1821; d) J. J. Soldevilla-Barreda, N. Metzler-Nolte, *Chem. Rev.* **2019**, *119*, 829-869.

[187] a) J. A. Prescher, D. H. Dube, C. R. Bertozzi, *Nature*. **2004**, *430*, 873-877; b) J. A. Prescher, C. R. Bertozzi, *Nat. Chem. Biol.* **2005**, *1*, 13-21.

[188] Y. Bai, J. Chen, S. C. Zimmerman, *Chem. Soc. Rev.* **2018**, *47*, 1811-1821.

[189] L. Vigh, F. Joó, Á. Cséplö, *Eur. J. Biochem.* **1985**, *146*, 241-244.

[190] a) V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599; b) C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057-3064.

[191] J. E. Moses, A. D. Moorhouse, *Chem. Soc. Rev.* **2007**, *36*, 1249-1262.

protein modifications,[192] even within living bacteria [193] or in the surface of human cells.[194] Its use inside livings cells has been precluded in part by the well-known toxicity of copper and ascorbate, which can generate reactive oxygen species (ROS).

Another breakthrough in this field was reported in 2006 by Meggers and coworkers, and consisted of a Ru(II)-catalyzed cleavage of N-allylcarbamates in presence of thiols, a reaction that could be carried out in water, under air, and importantly, even in the presence of live mammalian cells.[195] The generation of a fluorescent product such as rhodamine 110, resulting from the allylcarbamate cleavage of a protected derivative, was used to confirm the reactivity of the catalyst under dilute conditions. Albeit the reaction required the use of high amounts of a co-solvent such as DMSO, the work can be seen as a starting point towards the rational development of transition metal catalyzed transformations that can be performed in biological media and living systems.

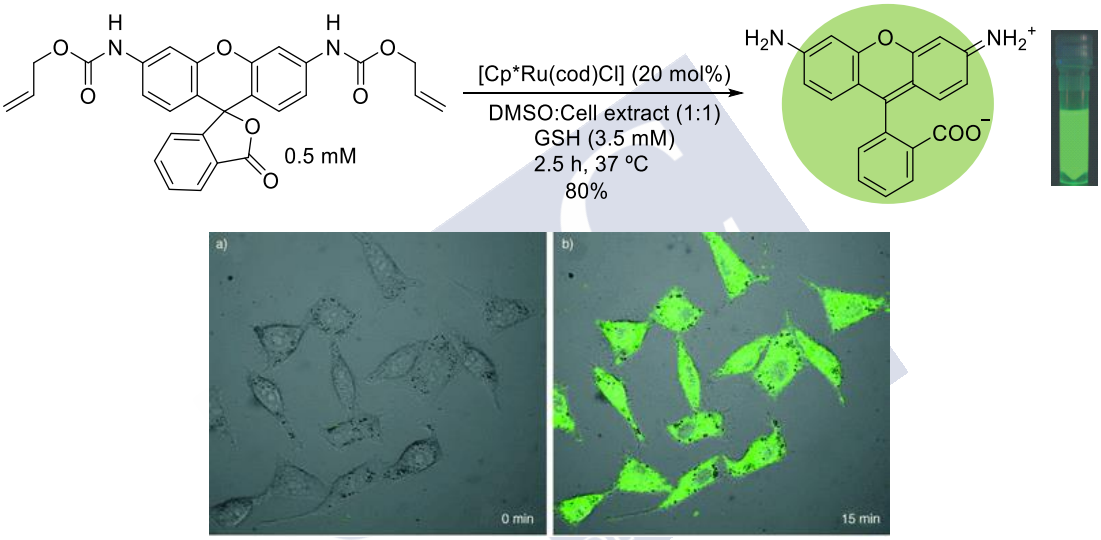


Figure 67. *Top:* Allyl carbamate cleavage in biological media. *Bottom:* Imaging of HeLa cells. a) before the addition of the Ru catalyst to cells preincubated with Rhodamine-alloc and PhSH. b) 15 minutes after the addition of ruthenium complex. (Source: Adapted from[195])

Bradley and Unciti-Broceta have pioneered the use of palladium nanostructured materials to promote chemical transformations in living settings. In 2011, they reported the use of polymer-supported Pd(0) nanoparticles to promote the first Suzuki Miyaura cross coupling reactions in cells,[196] even though the cells were fixed before the analysis, and the results seem difficult to reproduce. These authors also demonstrated that the palladium nanoparticles can be used to remove allylcarbamate groups in different type

[192] a) L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science*. **2001**, *292*, 498-500; b) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem*. **2010**, *79*, 413-444.

[193] M. Yang, A. S. Jalloh, W. Wei, J. Zhao, P. Wu, P. R. Chen, *Nat. Commun*. **2014**, *5*, 4981-5024.

[194] C. Uttamapinant, A. Tangpeerachaikul, S. Grecian, S. Clarke, U. Singh, P. Slade, K. R. Gee, A. Y. Ting, *Angew. Chem. Int. Ed*. **2012**, *51*, 5852-5856.

[195] C. Streu, E. Meggers, *Angew. Chem. Int. Ed*. **2006**, *45*, 5645.

[196] R. M. Yusop, A. Unciti-Broceta, E. M. Johansson, R. M. Sanchez-Martin, M. Bradley, *Nat. Chem*. **2011**, *3*, 239-243.

of substrates. After that, the group of P. R. Chen continued exploring the use of palladium in biological environments. In 2014, they reported the use of discrete Pd(II) catalysts for the activation of protein in living cells, using a depropargylation reaction of protected lysines.[197] This uncaging strategy was used for the chemical rescue of a non-overexpressed bacterial enzyme phosphothreonine lyase (OspF) within HeLa cells, using $Pd_2(allyl)_2Cl_2$ as precatalyst. Our group has also pioneered the use of intracellular ruthenium catalysis for the activation of DNA binders.[198]

Following their preliminary work, Meggers demonstrated in 2014 that using Kitamura's ruthenium catalysts, which are equipped with quinoline ligands, the addition of external nucleophilic thiol for activating the catalyst is not needed; the glutathione present in the cells is nucleophilic enough to ensure the catalytic process.[199] In 2016, our group demonstrated that by appropriate engineering of the ligands, the ruthenium complexes can be targeted to different cellular compartments.[200] In particular, we found that using a triphenylphosphonium derivative the ruthenium complex accumulates in the mitochondria of living cells, where it is still able to promote not only the uncaging of the alloc-protected rhodamine, but also the uncaging of a pro-drug that works as a very efficient uncoupler of the mitochondrial membrane potential.

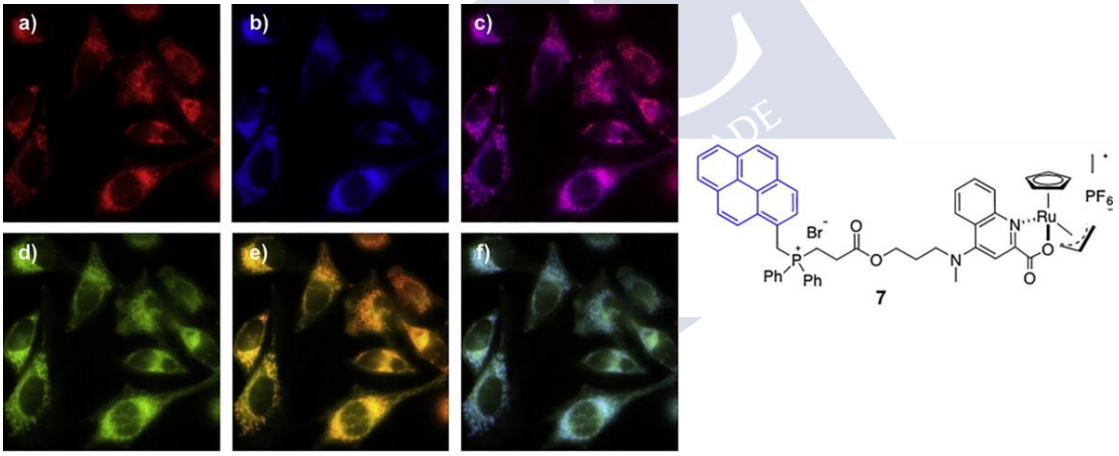


Figure 68. *Left:* Imaging of the subcellular localization and catalytic activity of ruthenium complex **7**. a) Mitochondrial labeling with tetramethylrhodamine ethyl ester (TMRE) (red). b) emission of cells incubated with the ruthenium complex **7** (blue). c) merging of (a) and (b). d) fluorescence in cells pre-incubated with **7** after addition of caged rhodamine. e) merging of (a) and (d). f) merging of (b) and (d). *Right:* Structure of ruthenium complex **7**. (Source: Adapted from[200]).

In 2017, Weissleder and co-workers reported another example of the applicability of palladium catalysts for deallylation reactions in living settings, despite they confirmed

[197] J. Li, J. Yu, J. Zhao, J. Wang, S. Zheng, S. Lin, L. Chen, M. Yang, S. Jia, X. Zhang, P. R. Chen, *Nat. Chem.* **2014**, *6*, 352-361.

[198] M. I. Sánchez, C. Penas, M. E. Vázquez, J. L. Mascareñas, *Chem. Sci.* **2014**, *5*, 1901-1907.

[199] T. Völker, F. Dempwolff, P. L. Graumann, E. Meggers, *Angew. Chem. Int. Ed.* **2014**, *53*, 10536-10540.

[200] M. Tomás-Gamasa, M. Martínez-Calvo, J. R. Couceiro, J. L. Mascareñas, *Nat. Commun.* **2016**, *7*, 12538-12548.

that the active palladium catalysts are ready deactivated in the complex cellular milieu.[201] Our group has also shown that discrete Pd(II) complexes with designed phosphine ligands present a better balance between reactivity and stability to achieve deallylation reactions in living Hela and Vero cells.[202]

In these last years, other, more complex reactions, have been added to the portfolio of transformations that can take place inside living mammalian cells. In 2018, our group demonstrated that water-soluble copper(I) complexes featuring designed ligands can tune the cell uptake and thereby promote intracellular CuAAC annulations of small, abiotic, and freely diffusible molecules.[203] Therefore, some of the previous limitations of the intracellular CuAAC chemistry, like the requirement of excess of ligands, or the use of cytotoxic sodium ascorbate were partially solved. In an alternative approach, we also reported a gold-mediated C–C bond formation that occurs in complex aqueous habitats and can be translated to living mammalian cells. Moreover, we demonstrated the viability of achieving the gold-promoted process in parallel with a ruthenium-mediated reaction, inside living cells, and in a bioorthogonal and mutually orthogonal manner.[204]

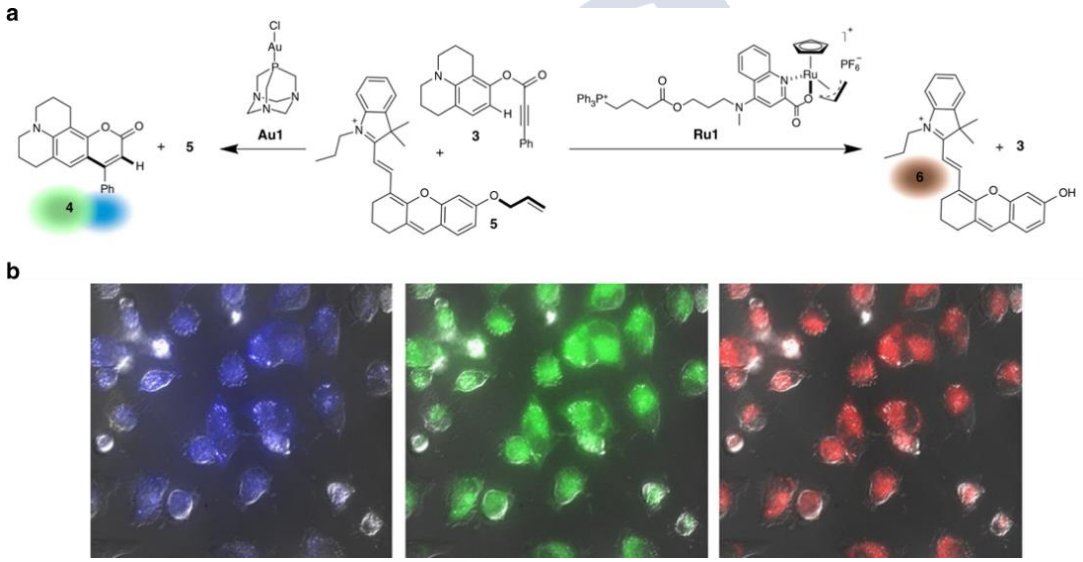


Figure 69. Concurrent Au(I) and Ru(II) catalysis inside HeLa cells. a) Schematic representation of the ruthenium and the gold-mediated reactions. b) Intracellular transformations. Blue, green and red fluorescence of cells incubated with **Ru1** and **Au1** followed by addition of substrates **3** and **5**. (Source: Adapted from[204])

Sadler and Do have independently reported the use of osmium or iridium complexes to alter NADH/NAD^+ cellular equilibriums or promote reductions of specific aldehydes or

[201] M. A. Miller, B. Askevold, H. Mikula, R. H. Kohler, D. Pirovich, R. Weissleder, *Nat. Commun.* **2017**, *8*, 15906-15919.
[202] M. Martínez-Calvo, J. R. Couceiro, P. Destito, J. Rodríguez, J. Mosquera, J. L. Mascareñas, *ACS Catal.* **2018**, *8*, 6055-6060.
[203] J. Miguel-Ávila, M. Tomás-Gamasa, A. Olmos, P. J. Pérez, J. L. Mascareñas, *Chem. Sci.* **2018**, *9*, 1947-1952.
[204] C. Vidal, M. Tomás-Gamasa, P. Destito, F. López, J. L. Mascareñas, *Nat. Commun.* **2018**, *9*, 1913-1923.

ketones.[205] Although the efficiency of these processes was low, these studies suggest that metal-hydride species can be generated inside cells. Considering these observations, our group reported the first metal-catalyzed isomerization reaction that can be carried out in the interior of living cells. The reaction is promoted by a ruthenium(IV) precatalyst that combines water solubility with some lipophilicity. The redox-neutral isomerization allows to transform non-functional, abiotic allyl alcohols into fluorescent or bioactive ketones, inside mammalian cells, with a remarkable intracellular turnover.[206]

Overall, a number of small-sized transition metal complexes can promote non-natural chemical reactions in biological and intracellular environments, albeit with very limited efficiency yet. In contrast, as indicated at the beginning of this introduction, intracellular reactions mediated by metalloproteins are essentially unknown. Considering our discovery that pallado-miniproteins like those described in chapter 1 can be readily translocated into the cell cytosol, we wondered whether they could promote designed palladium-dependent reactions inside cells. This could represent a significant first step towards the challenging endeavour of developing artificial catalytic metalloproteins that work in *in vivo* settings.

[205] a) S. Bose, A. H. Ngo, L. H. Do, *J. Am. Chem. Soc*. **2017**, *139*, 8792-8795; b) J. P. C. Coverdale, I. Romero-Canelón, C. Sanchez-Cano, G. J. Clarkson, A. Habtemariam, M. Wills, P. J. Sadler, *Nat. Chem.* **2018**, *10*, 347-354.

[206] C. Vidal, M. Tomás-Gamasa, A. Gutiérrez-González, J. L. Mascareñas, *J. Am. Chem. Soc.* **2019**, *141*, 5125–5129.

Objective

Develop metal-equipped synthetic miniproteins made of natural amino acids that are capable of promoting abiotic reactions inside living mammalian cells. We have demonstrated before that basic domains of bZIP transcription factors, mutated to include two histidine residues at *i, i+4* positions, can form metallocomplexes with palladium (II) sources. We wondered whether these stapled pallado-miniproteins with capacity to internalize into living mammalian cells, might perform palladium-promoted designed reactions.

Results and discussion

Internalization assays

Before testing catalytic reactivities, we considered fundamental to obtain more information on the cell internalization requirements of the miniprotein. Our previous studies had demonstrated that mixing the peptide **brHis$_2$** with $PdCl_2$(en) produces relatively stable peptide palladium complexes (detected by ESI-MS), and that these metallopeptides are able to internalize at low concentrations (5 μM) in HeLa cells. (Chapter I, section 3).[207]

We have now found that the metal-switch is perfectly operative in specifically modified peptides, such as the double mutant peptides [Leu^{247}→ Ala, Gln^{248}→ Ala] or [Thr^{236}→ Ala, Ala^{238}→ Thr], as well as the single mutant Ala^{238}→ Thr. Importantly, the cell entrance of longer peptides that include amino acid insertions, such as one containing a Leu-Ser fragment between Ala^{229} and His^{230}, can be also regulated by the Pd(II) reagent. None of the peptides internalized, at those concentrations, in the absence of palladium complex (Figure 70).

Not surprisingly, the cell uptake is sensitive to the presence of arginine residues, which is in consonance with the well-known role of these amino acids in cell penetrating peptides.[208] Thus, the double mutant [Arg^{245}→ Ser, Arg^{249}→ Ala] peptide failed to translocate across the plasmatic membrane, even in the presence of the palladium clip.[209]

[207] S. Learte-Aymamí, N. Curado, J. Rodríguez, M. E. Vázquez, J. L. Mascareñas, *J. Am. Soc. Chem.* **2017**, *139*, 16188-16193 (Chapter I, subchapter 3).

[208] a) J. B. Rothbard, E. Kreider, C. L. VanDeusen, L. Wright, B. L. Wylie, P. A. Wender, *J. Med. Chem.* **2002**, *45*, 3612-3618; b) K. Sakagami, T. Masuda, K. Kawano, S. Futaki, *Mol. Pharmaceut.* **2018**, *15*, 1332-1340.

[209] All the peptides have in their N-terminal an amino hexanoic acid as a linker and a TAMRA fluorophore connected by amide bond (see Experimental Data for further structure details).

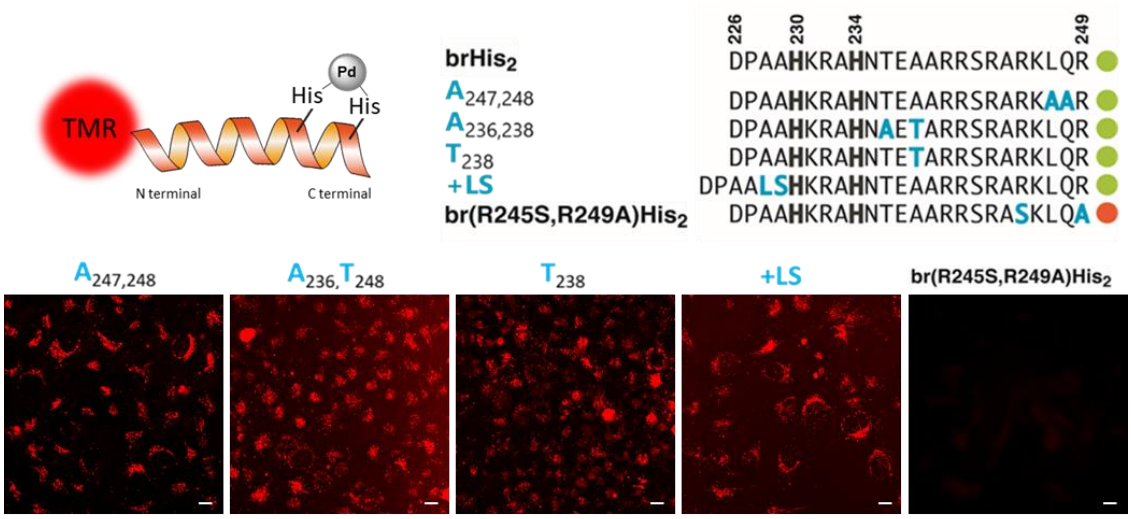


Figure 70. *Top Left:* Peptides structure. All peptides have in their N-terminal an amino hexanoic acid as a linker between the last amino acid (aspartic acid) and the TAMRA fluorophore, connected this by amide bond (see Experimental Data for further structure details). *Top right*: Sequences of modified peptides tested in the metal-switched cell uptake process, highlighting the mutated residues (blue) and whether the peptide is internalized (green dot), or does not show an observable translocation under the same conditions (red dot). *Bottom*: Fluorescence micrographies of Hela cells (confocal). Peptides were pre-mixed with $PdCl_2$(en) (1:1 ratio) in water for 10 min and incubated with cells for 30 min at 37 ºC. The cells were washed twice with PBS before observed under the microscope. All Incubations were made in Dulbecco's Modified Eagle Medium (DMEM) completed with 10% Fetal Bovine Serum (FBS). Red channel: λexc = 561 nm, λem = 620/60 nm. Scale bar: 10 μm.

In fact, keeping the six native arginines of the basic region fragment is key for the internalization, so that by replacing any of them by alanines leads to a significant decrease in the intracellular fluorescence of tetramethylrhodamine (TMR)-labelled derivatives. This can be easily observed in the next figure; only the original peptide containing the six native arginines goes inside living HeLa cells.

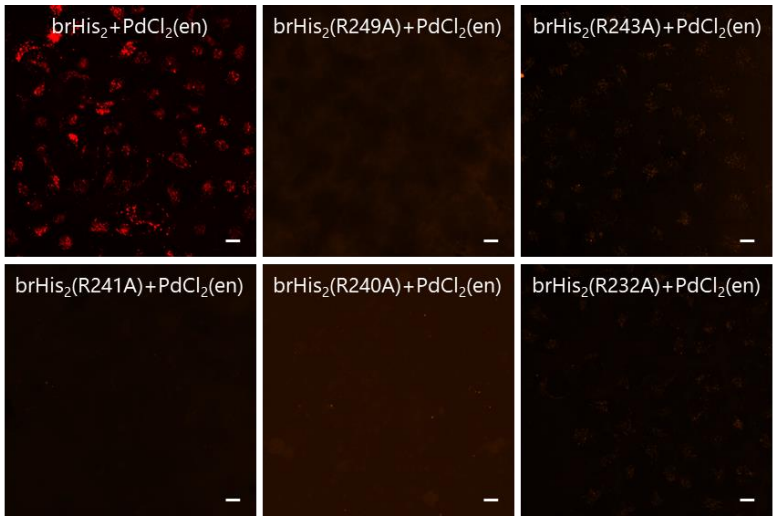


Figure 71. Fluorescence micrographies of HeLa cells (confocal). First, TMR-peptides were pre-incubated with $PdCl_2$(en) (1:1 ratio) in water for 10 min before the addition to cells. Then, we incubated HeLa cells with 5 µM of palladium-peptides for 30 min at 37 ºC. The cells were washed twice with PBS before being observed in a confocal microscope. All incubations were made in Dulbecco´s modified Eagle medium completed with 5% of fetal bovine serum (FBS-DMEM). λ_{exc} = 561 nm, λ_{em} = 620/60 nm. Scale bar: 10 µm.

Not surprisingly, when the internalization experiments were carried out at 4ºC instead of at 37ºC, the intracellular fluorescence, as measured by flow cytometry, was considerably lower. This result suggests that the cell penetration is energy dependent, and therefore very likely occurring by endocytosis, with the resulting peptide accumulated in vesicular structures.

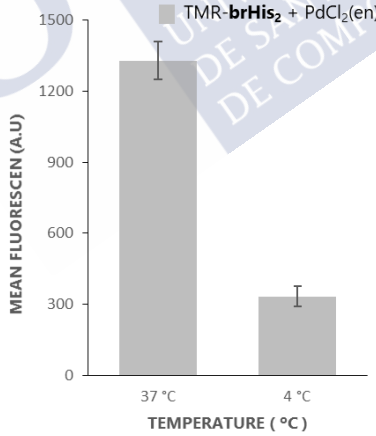


Figure 72. Study of internalization by cytometry. TMR-**brHis$_2$** was preincubated with $PdCl_2$(en) (1:1 ratio) in water for 10 min before the addition to cells. Then, HeLa cells were incubated with 5 µM TMR-**brHis$_2$**+$PdCl_2$(en) for 30 min at 37 ºC and 4 ºC. The cells were washed with PBS and trypsinized to measure. All incubations were made in FBS-DMEM. Data are mean ± SEM of three independent experiments performed in triplicate.

Monitoring palladium inside cells

The cell-penetrating staple peptides are not limited to the use of palladium dichloride with ethylenediamine (en) ligands, and can feature other bidentate ligands, for example, a bipyridine. Therefore, we envisioned the use of a bipyridine ligand equipped with a

different fluorophore from the TMR, already present in the peptide, to simultaneously track the localization of the peptide and the metal complex. Specifically, we synthesized the bipyridine-anthracene palladium derivative **Pd1** (Figure 73, top) and found that it also reacts efficiently with the bis-histidine peptide to give the expected complex, as corroborated by LC-MS (see Experimental Data). HeLa cells were then incubated for 1 h with the complex resulting from mixing equivalent amounts of TMR-**brHis$_2$** and **Pd1** (30 min). After washing (PBS, 2X), cells were observed by microscopy at different times. Figure 73a and b show the intracellular emissions corresponding to both fluorophores after 4 h, while Figure 73c is a colocalization image. Observation at shorter times of incubation, after washing, did not show differences in the localization of the fluorophores, which suggest that the complex remains stable. A control experiment using only the metal complex (**Pd1**), at the same concentration, revealed a very poor internalization (Figure 73d), confirming the role of the peptide as palladium transporter.

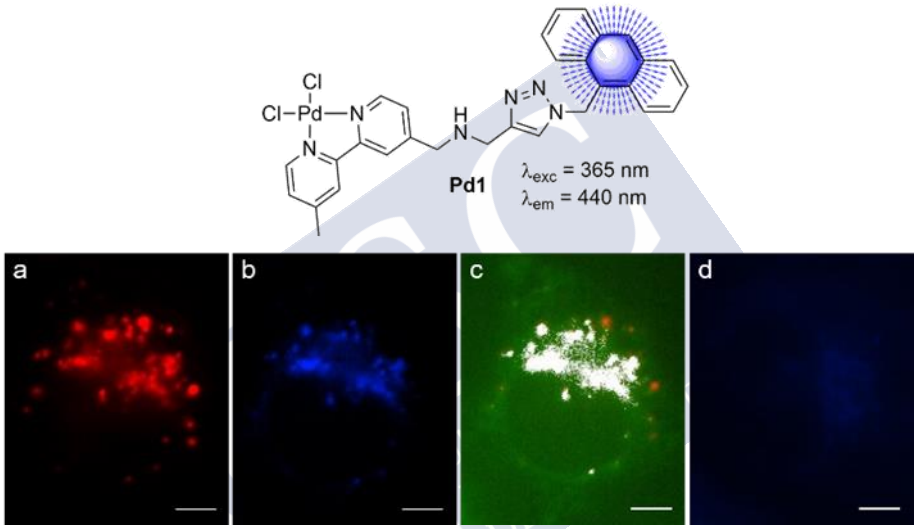


Figure 73. *Top:* Structure of **Pd1.** *Bottom:* Fluorescence micrographies of HeLa cells. The palladopeptide complex was made by mixing TMR-**brHis$_2$** (10 µM) with **Pd1** (1:1 ratio) in water for 30 min. Cells were incubated with the mixture for 1 h at 37 ºC, washed twice with PBS, replenished with fresh DMEM, and observed under the microscope at different times. a) Red emission channel corresponding to the TAMRA fluorophore; b) Blue emission channel corresponding to the anthracene; c) Colocalization image processed with Image J; d) Blue emission channel after incubation with **Pd1** alone. Incubations were made FBS-DMEM. Red channel: λ_{exc} = 550 nm, λ_{em} = 590-650 nm. Blue channel: λ_{exc} = 385 nm, λ_{em} = 450-510 nm. Scale bar: 5 µm, the image shows a single cell.

In vitro reactivities

With a method to internalize a relatively stable Pd(II) complex, we wondered whether this species could be engaged in catalytic reactions. As testing reaction we explored the depropargylation of the probe HBTPQ (**1,** Figure 74a),[202] which is not fluorescent, but upon uncaging generates a product **2** that emits light at 635 nm when excited in the far-ultraviolet (far-UV) region (λ_{exc} = 330 nm). Importantly, previous work from our and other groups had shown that standard palladium sources like PdCl(allyl)$_2$ or Pd(OAc)$_2$ fail to promote this type of reaction inside HeLa cells,[201,202] likely because of a poor cell

penetration of the palladium species, as well as a rapid deactivation in the presence of cellular components.

We first assessed the reactivities *in vitro*, using PBS as solvent. $Pd(OAc)_2$ and the palladium chloride sources used in the preparation of the hybrid bis-histidine metal complexes, $PdCl_2$(en) and $PdCl_2$(bpy), were able to promote the reaction (conditions: 200 μM solution of **1** and 10 mol % of the catalyst, Figure 74b). Noticeably, the palladopeptide resulting from mixing stoichiometric amounts of peptide and $PdCl_2$(en) in water was much less effective than the discrete palladium species.[210] Despite this result could be discouraging, it is not so surprising, as in this complex, the ethylendiamine ligand is not labile enough to allow an efficient coordination of the substrate. Gratifyingly, a similar metallopeptide made from $PdCl_2$(COD) (COD = 1,8-cyclooctadiene) instead of $PdCl_2$(en),[211] gave the product in a 65% yield after 24 h of reaction. This result confirms the need of using Pd(II) reagents with suitable labile ligands to obtain good reactivities. As a control reagent we also explored the reactivity of a purposely made bis-histidine dichloride complex, **$PdCl_2$(Boc-His)$_2$** which gave the product with lower yields (30%) than the metallopeptide, suggesting a positive effect of the peptide scaffold in the reactivity.

[210] S. E. Coelho, F. S. S. Schneider, D. C. de Oliveira, G. L. Tripodi, M. N. Eberlin, G. F. Caramori, B. de Souza, J. B. Domingos, *ACS Catal.* **2019**, *9*, 3792-3799.

[211] Not surprising, while in the MS of the palladopeptide resulting from mixing **brHis$_2$** and $PdCl_2$(en) the ethylenediamine ligand is conserved, when using $PdCl_2$(COD) the MS fits with a palladopeptide without ligands (see the Experimental Data for more details).

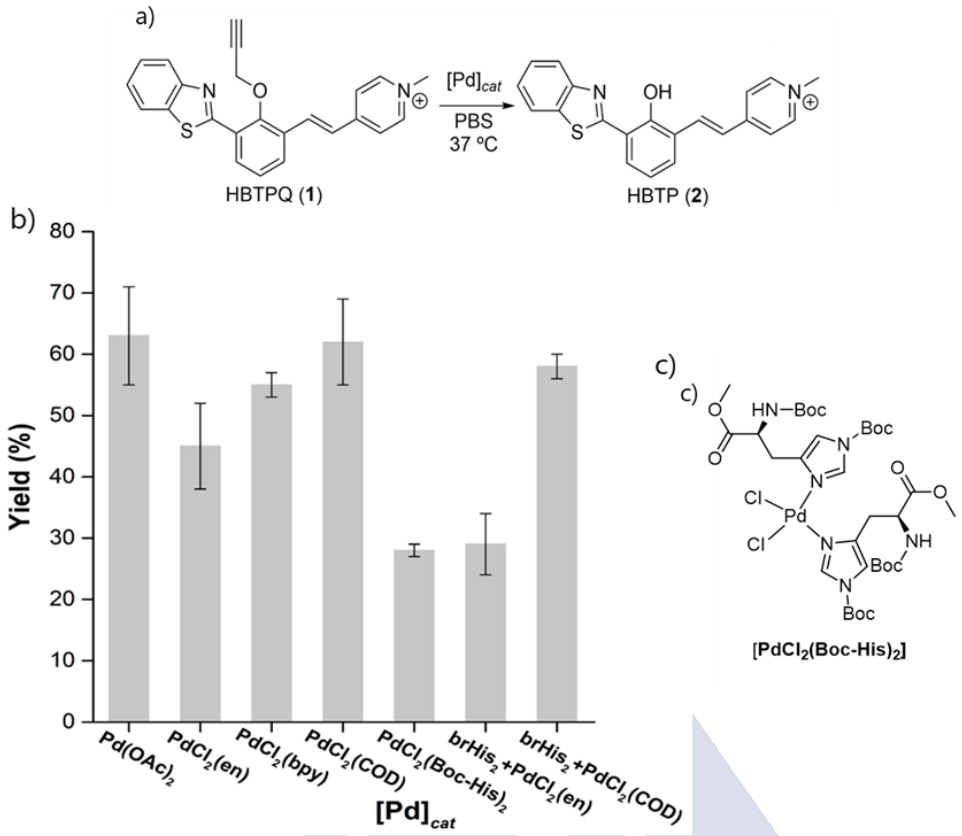


Figure 74. a) Uncaging of HBTPQ (**1**). b) Bar diagram representation of the yields obtained for each catalyst. Reaction conditions: HBPTQ (200 μM), Pd source (20 μM, 10 mol%) in 1 mL of PBS at 37 ºC, for 24 h. Yields were calculated by RP-HPLC-MS using internal standards. Data are mean ± SEM of three independent experiments performed in triplicate. c) Structure of **$PdCl_2$(Boc-His)$_2$**.

Reactions in live mammalian Cells

After these *in vitro* results, the stage was set for studying the intracellular reactivity. In consonance with the results observed with $PdCl_2$(en), the internalization of TMR-**brHis$_2$** can also be triggered by pre-treatment with the complex $PdCl_2$(COD) (Figure 75a). Interestingly, circular dichroism spectra of the peptide (TMR-**brHis$_2$**) after adding increasing amounts of the Pd(II) complex, at 4ºC, confirmed the propensity of the peptide to acquire an alpha-helix structure (Figure 75b).[212]

[212] J. T. Pelton, L. R. McLean, *Anal. Biochem*. **2000**, *277*, 167-176.

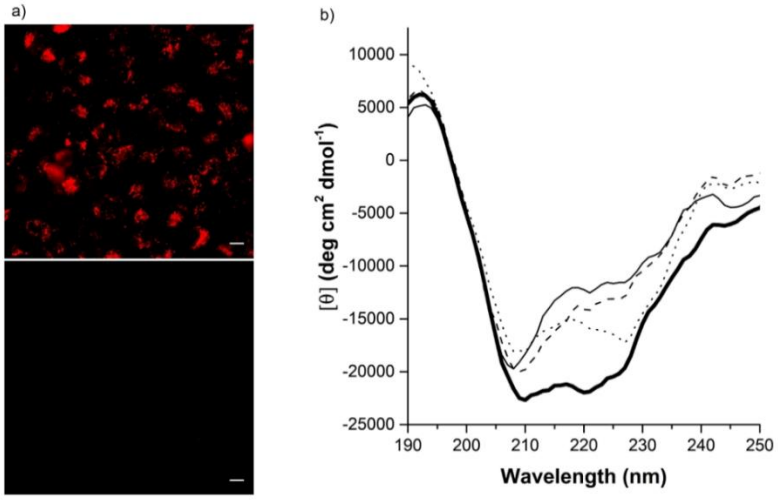


Figure 75. a) Fluorescence micrographies of HeLa cells (confocal). *Top:* TMR-**brHis$_2$** (5 μM) was pre-mixed with $PdCl_2$(COD) (1:1 ratio) in water for 10 min and incubated with cells for 30 min at 37 ºC. *Bottom:* Incubation with TMR-**brHis$_2$** alone. The cells were washed with PBS before being observed under the microscope. All the incubations were made in FBS-DMEM. Red channel: λ_{exc} = 561 nm, λ_{em} = 620/60 nm. Scale bar: 10 μm. b) Circular Dichroism titration of a 5 μM solution of **brHis$_2$** (narrow solid line) and increasing concentrations of $PdCl_2$(COD): 1 equiv. (dashed line), 3 equiv. (dotted line) or 10 equivalents (thick solid line). The experiments were carried out at 4 °C, in 10 mM phosphate buffer pH 7.5 and 100 mM of NaCl. Mean residue molar ellipticity (MRE) was calculated considering a 24-mer.

The cellular reactions were carried out by incubation of HeLa cells with 50 μM of probe **1** for 30 min at 37 °C. The medium was removed, and the cells were washed twice with FBS-DMEM to ensure the complete removal of the extracellular probe. The peptide-palladium reagents (50 μM) were then added in fresh media and, after 1 h and two washes with PBS, the cells were imaged by wide-field fluorescence microscopy. Gratifyingly, when cells were treated with the palladopeptide resulting from mixing TMR-**brHis$_2$** and $PdCl_2$(COD), we could observe a very intense intracellular fluorescence arising from the expected product **2** (Figure 76a). Not so surprisingly, with the metallopeptides made from $PdCl_2$(en) and $PdCl_2$(bpy) we did not observe fluorescence, which must be a consequence of the lack of reactivity, because of the presence of non-labile palladium ligands, since the palladopeptides were able to go inside the cells, as can be deduced from the TMR fluorescence (Figure 76b-c, bottom row).

Given that the palladopeptide was preformed just before the cellular incubations, it could not be discarded that the intracellular reaction was promoted by discrete palladium entities derived from the palladium source. However, in the absence of the peptide, the complex $PdCl_2$(COD) by itself is incapable of generating any intracellular fluorescence (Figure 76d). Other palladium sources like $Pd(OAc)_2$ or $PdCl_2$(allyl) also failed to generate intracellular fluorescence under the above conditions (Figure 77b-c). In addition, we observed a similar reactivity using lower amounts of the reactive metallopeptide (10 μM), which demonstrates the effectivity of the system (Figure 77a).

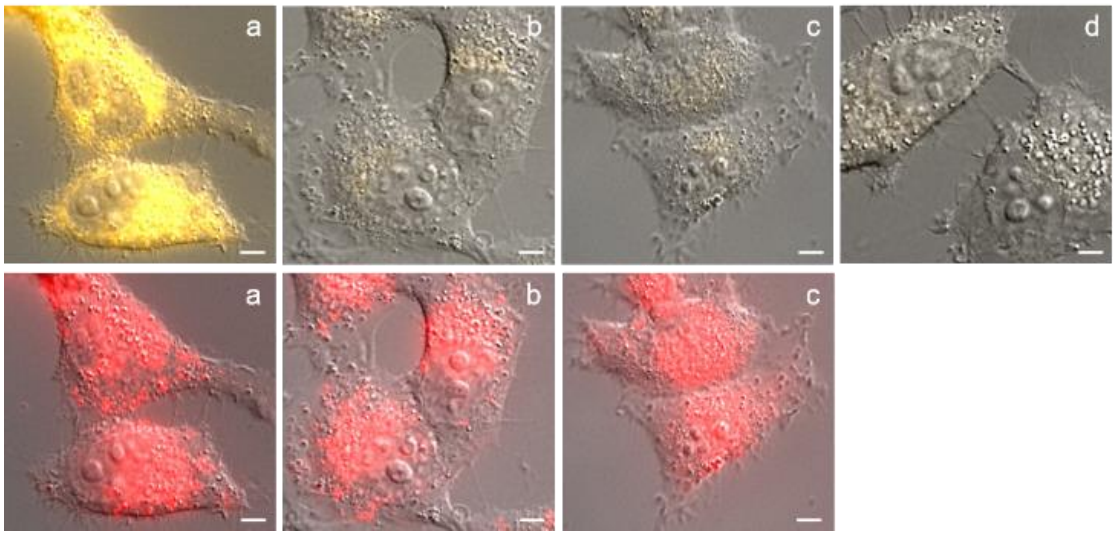


Figure 76. Intracellular depropargylation of **1** monitored by fluorescence microscopy in HeLa cells. Cells were treated with 50 μM of probe **1** for 30 min at 37 ºC and washed twice to remove excess of the probe. Peptides (50 μM) or solutions of peptide-palladium complexes obtained by mixing equivalent amounts of the peptide and the palladium source for 10 min, were added in fresh media, and the cells were incubated for 1 h. a) Incubation with the palladopeptide obtained from $PdCl_2$(COD); b) Incubation with the palladopeptide obtained from $PdCl_2$(en); c) Incubation with the palladopeptide obtained from $PdCl_2$(bpy); (d) Incubation with $PdCl_2$(COD), in absence of peptide. The cells were washed with PBS before being observed under the microscope. All the incubations were made in FBS-DMEM. Bright fields images are superimposed to the red or orange emission channel (the orange colour is arbitrarily chosen to differentiate the emission with the red channel). Top row: Orange channel: λexc = 385 nm, λem = 515-700 nm; Bottom row: Red channel: λexc = 550 nm, λem = 590-650 nm. Scale bar: 5 μm.

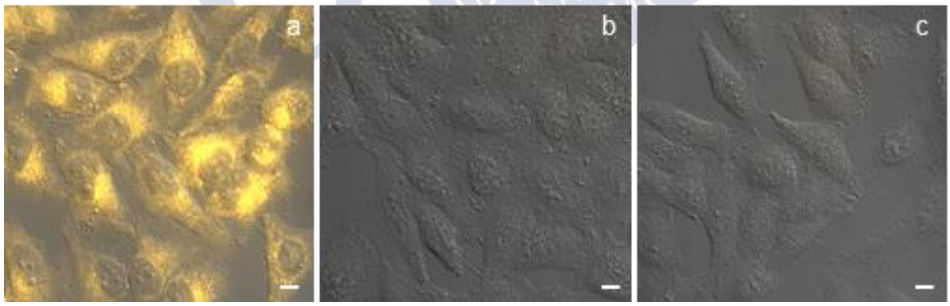


Figure 77. Fluorescence micrographies of HeLa cells. Standard conditions as in the previous case. a) Incubation with the palladopeptide resulting from TMR-**brHis$_2$** and $PdCl_2$(COD) (10 μM); b) Incubation with $Pd(OAc)_2$ (10 μM); c) Incubation with $PdCl_2$(allyl) (10 μM). λex= 385 nm, λem= 515-700 nm. Scale bar: 10 μm.

Control experiments with a peptide lacking the TMR label gave similar reactivity results, confirming that the fluorophore label does not affect the process. As expected, using the peptide TMR-**brHis$_2$** alone, we did not detect any fluorescence derived from the product. Moreover, the propargylated probe **1** did not present any fluorescence even with long times of incubations (Figure 78).

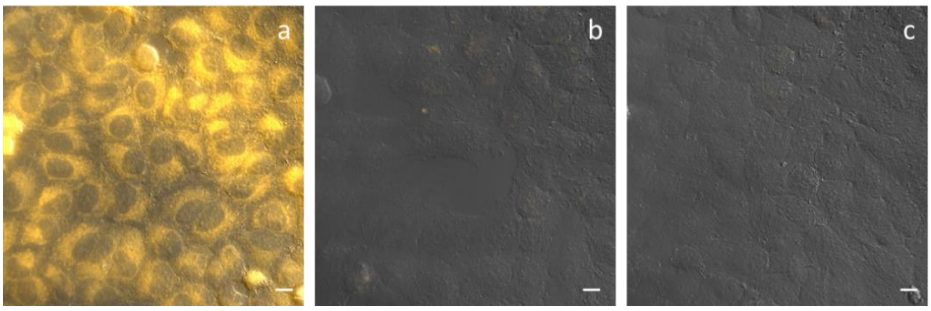


Figure 78. Fluorescence micrographies of HeLa cells. Standard conditions as in the previous cases. a) Incubation with the palladopeptide resulting from mixing **brHis$_2$** (without TAMRA label) and $PdCl_2$(COD); b) Incubation with TMR-**brHis$_2$** without metal; c) Incubation with HBTPQ (propargylated probe **1**, 50 μM) without peptide. λex= 385 nm, λem= 515-700 nm. Scale bar: 10 μm.

ICP-MS analysis of cellular extracts revealed that the palladium content is substantially higher in cells treated with the preformed palladoprotein hybrid than with equivalent amounts of $PdCl_2$(COD) (Figure 79). Nonetheless, this does not account for the lack of reactivity observed with these palladium sources, which suggests that the stapled peptide structure also plays a key role to favour the intracellular reaction, in part because it might stabilize the palladium.

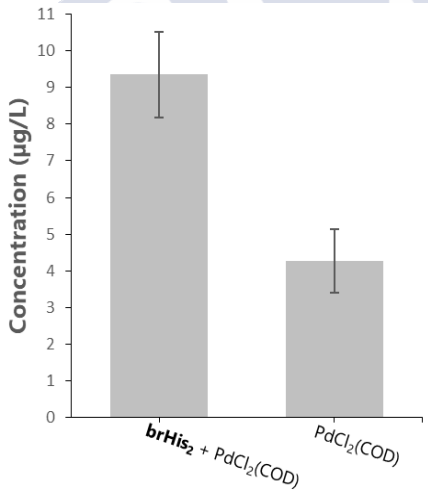
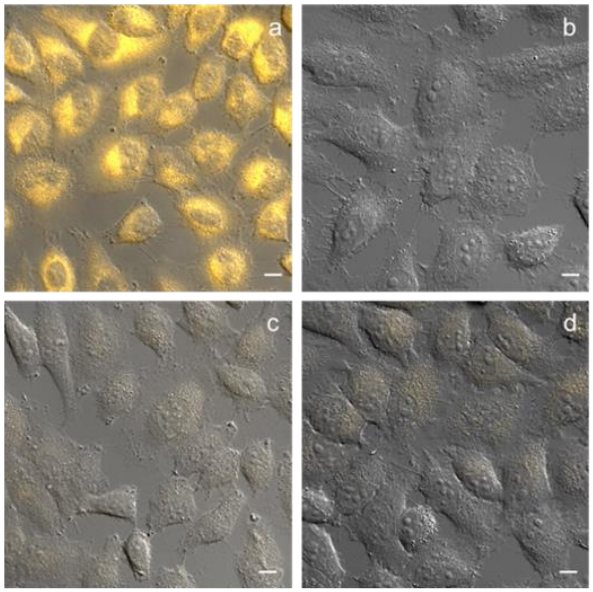


Figure 79. brHis$_2$ and $PdCl_2$(COD) (1:1 ratio) were pre-mixed in water for 10 min. Then, we incubated HeLa cells with 10 μM of resulting metallopeptide or just with $PdCl_2$(COD), for 2 h at 37 ºC. Finally, the cells were washed twice with PBS and digested with HNO_3. Data are mean ± SEM of three independent experiments.

Additional insights were obtained using as reagents the peptides **brHis** (His^{230} mutated to Leu), or **br** (native GCN4-basic region, without histidines), instead of **brHis$_2$**, which were not able to promote any intracellular reaction when premixed with $PdCl_2$(COD) (Figure 80b-c). Using a mutated peptide that lack key arginines for cellular uptake (R249A, R245A), also led to similar fails (Figure 80d). These results confirm the requirement of a bis-histidine stapling and an appropriate set of arginines for cellular uptake and intracellular reactivity.



TMR-**br**	TMR-Ahx-DPAALKRARNTEAARRSRARKLQR
TMR-**brHis**	TMR-Ahx-DPAALKRA**H**NTEAARRSRARKLQR
TMR-**brHis$_2$**	TMR-Ahx-DPAA**H**KRA**H**NTEAARRSRARKLQR
TMR-**brHis$_2$(R249A, R245S)**	TMR-Ahx-DPAA**H**KRA**H**NTEAARRSRA**S**KLQ**A**

Figure 80. *Top:* Fluorescence micrographies of HeLa cells. Standard conditions as in the previous cases. a) Incubation with the palladopeptide obtained from **brHis$_2$** and $PdCl_2$(COD) (10 μM); b) The same but using **brHis** instead of **brHis$_2$**; c) The same but using **br** instead **brHis$_2$**; d) The same but using a mutated TMR-**brHis$_2$**(R249A, R245S). λex= 385 nm, λem= 515-700 nm. Scale bar: 10 μm. *Bottom:* Sequences of peptides used in this assay.

Also importantly, the approach can be used in other human cells such as A549 and mammalian cells like Vero, which supports the generality of the approach.

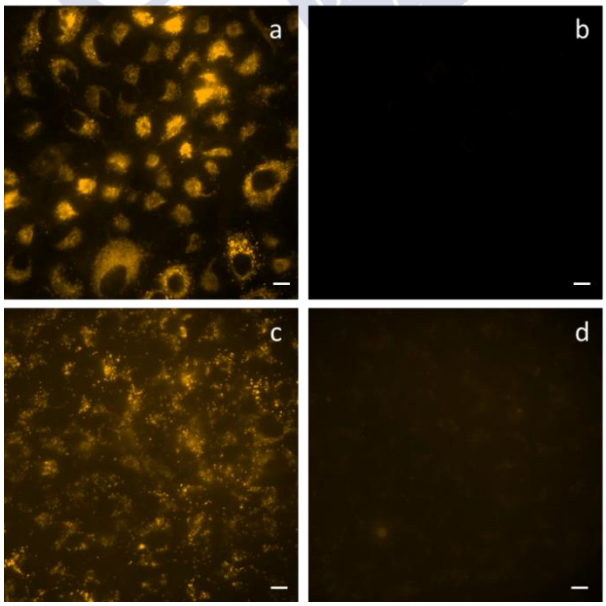


Figure 81. Fluorescence micrographies with A549 cells (top) and Vero cells (bottom). Standard conditions. a) Incubation with the palladopeptide resulting from **brHis$_2$** and $PdCl_2$(COD) (10 μM); b) Incubation with $PdCl_2$(COD) alone (10 μM); c) Incubation with the palladopeptide resulting from **brHis$_2$** and $PdCl_2$(COD) (10 μM); d) Incubation with $PdCl_2$(COD) (10 μM) alone. λex= 385 nm, λem= 515-700 nm. Scale bar: 10 μm.

The chemistry can be extended to other challenging substrates, like the bis-propargyl derivative **3**, which upon deprotection produces a far-red emitting product. As in the case of the phenol derivative **1**, only the palladominiprotein was able to perform the reaction inside mammalian cells.

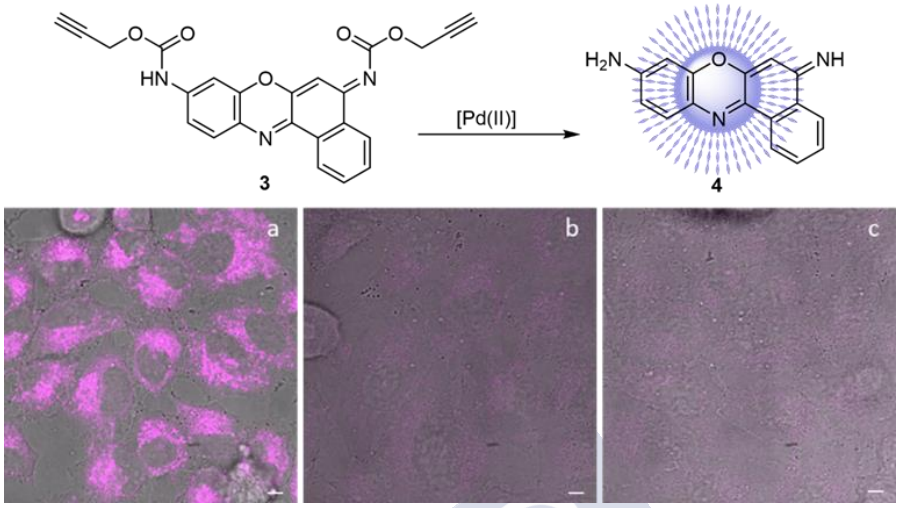


Figure 82. *TOP:* Uncaging of probe **3**. *Bottom*: Fluorescence micrographies of HeLa cells (confocal). Standard conditions as in the previous cases, with 75 μM of probe **3** and 10 μM of the reagents. Bright field images are superimposed to far-red emission channel. a) Incubation with the metallopeptide made from **brHis$_2$** and $PdCl_2$(COD); b) Incubation with $PdCl_2$(COD) alone; c) Incubation with substrate **3** alone. λ_{exc} = 561 nm, λ_{em} = 620/60 nm. Scale bar: 10 μm.

Reactivities using shorter peptides

Altogether, the above results confirm that engineering a bis-histidine handle in a basic miniprotein can provide for cellular internalization and cytosolic palladium catalysis owing to its ability to form palladium stapled derivatives. An ensuing question is whether the palladium stapling strategy is also effective for other shorter peptides.[213] Considering the well-known integrin targeting motif RGD (Arg-Gly-Asp), we wondered whether it could be also converted into a reactive metallopeptide by adding two histidines to both N- and C-terminal residues (Figure 83, panel 1). Thus, a TMR-labelled derivative (bonding directly the fluorophore with the N-terminal histidine) revealed that while the peptide **HRGDH** presents a relatively poor cellular uptake in HeLa cells, the internalization is significantly improved when pre-treated with $PdCl_2$(COD), though the uptake is not as efficient as with the above basic miniprotein (Figure 84).

Not surprisingly, the peptide-palladium complex is also capable of performing the intracellular depropargylation of substrate **1** (Figure 83, panel 2b), however the reactivity was lower compared with the pallado-miniprotein discussed before (Figure 83, panel 3). A similar peptide with only one histidine (**HRGDA**) presents some capacity to internalize

[213] A conjugate between a tetralysine moiety and a non-natural palladium ligand (carbene derivative) has been previously used for the depropargylation of rhodamine derivatives, using fixed cells: E. Indrigo, J. Clavadetscher, S. V. Chankeshwara, A. Megia-Fernandez, A. Lilienkampf, M. Bradley, *Chem. Commun.* **2017**, *53*, 6712-6715.

(the same with or without adding $PdCl_2$(COD), see Figure 84), but failed to act as a catalyst (Figure 83, panel 2c). Interestingly, the non-cyclic bis-histidine palladium complex **$PdCl_2$(Boc-His)$_2$** also failed to generate the intracellular product (Figure 83, panel 2d).

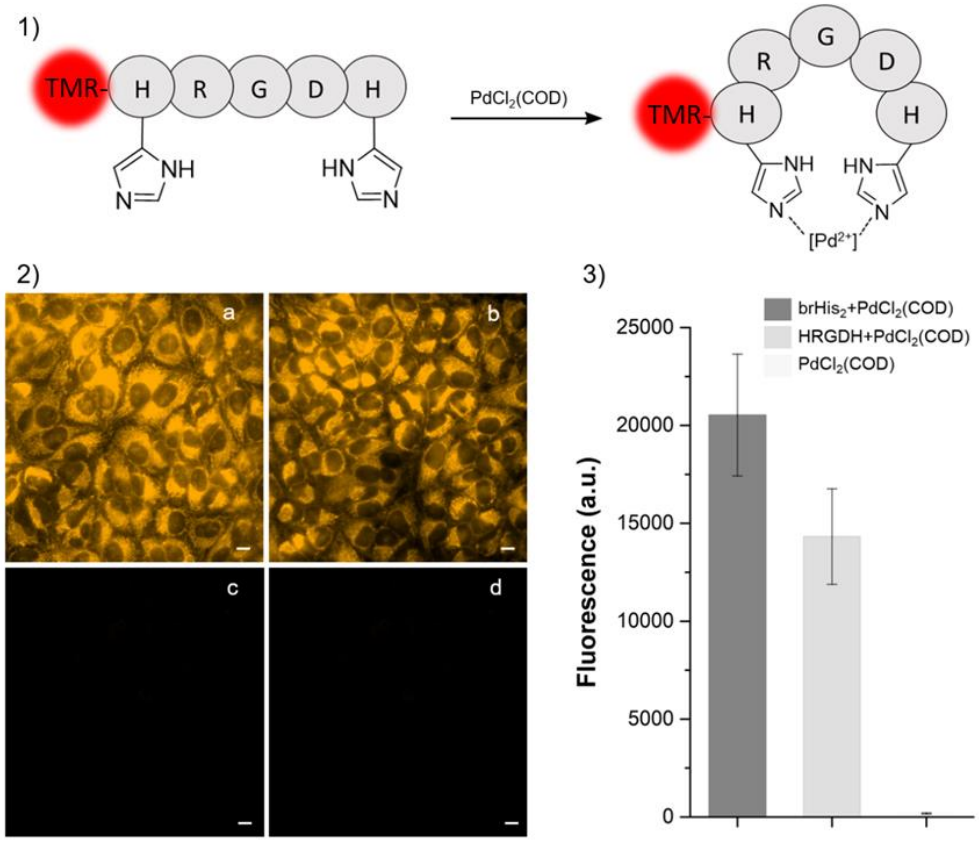


Figure 83. 1) Bis-Histidine RGD derivative and stapling produced by treatment with the palladium complex. 2) Intracellular depropargylation using the peptide TMR-**HRGDH** or control derivatives. HeLa cells were treated with 50 µM of probe 1 for 30 min at 37 ºC and washed twice. Peptides (10 µM) or solutions of peptide-palladium complexes obtained by mixing equivalent amounts for 10 min, were added in fresh media, and the cells were incubated for 1 h. a) Pallado-miniprotein resulting from $PdCl_2$(COD) and **brHis$_2$**; b) Mixture of TMR-**HRGDH** and $PdCl_2$(COD); c) Mixture of TMR-**HRGDA** and $PdCl_2$(COD); d) $PdCl_2$(Boc-His)$_2$. The cells were washed twice with PBS before being observed under the microscope. All incubations were in FBS-DMEM. λ_{exc} = 385 nm, λ_{em} = 515-700 nm. Scale bar: 10 µm. 3) CTFC measurements after the intracellular reaction. Data are mean ± SEM of three independent experiments performed in triplicate.

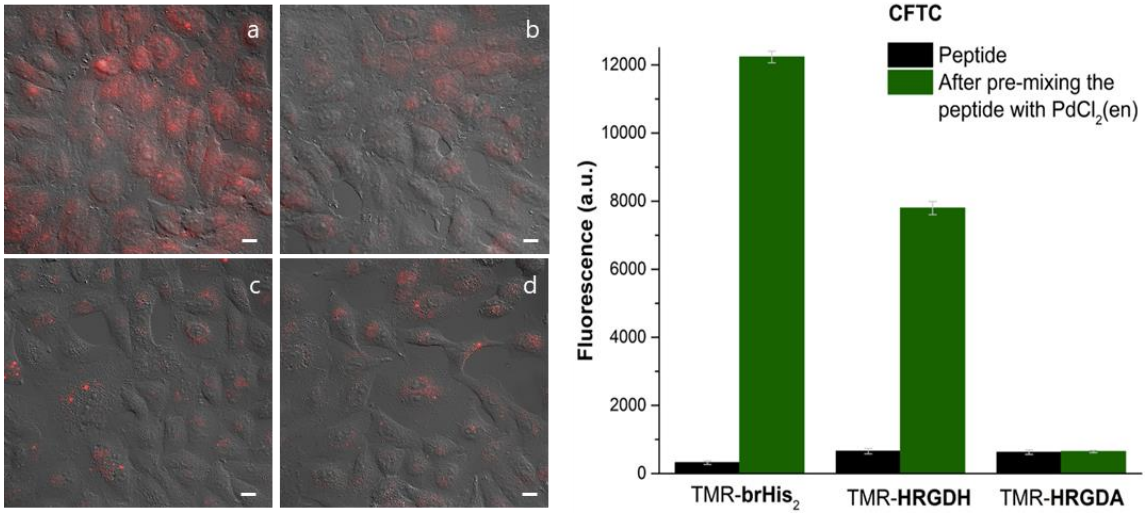


Figure 84. *Left*: Fluorescence micrographies of HeLa cells. Peptides (5 μM) was pre-mixed with $PdCl_2$(COD) (1:1 ratio) in water for 10 min and incubated with cells for 30 min at 37 ºC. The cells were washed with PBS before being observed under the microscope. All the incubations were made in FBS-DMEM. a) Incubation with TMR-**HRGDH** in presence of $PdCl_2$(COD); b) Incubation with TMR-**HRGDH** alone; c) Incubation with TMR-**HRGDA** in presence of $PdCl_2$(COD); d) Incubation with TMR-**HRGDA** alone. λex= 550 nm, λem= 590-650 nm. Scale bar: 10 μm. *Right*: CTFC measurements in HeLa Cells. The experiments of internalization were carried out following the standard protocol. Data are mean ± SEM of three independent experiments performed in triplicate.

Cytotoxicity assays

It is important to note that cell viability tests (MTT in HeLa Cells, 24 h) demonstrated that none of palladopeptide hybrids or palladium complexes were toxic below 50 μM (more than 90% viability after 24 h).

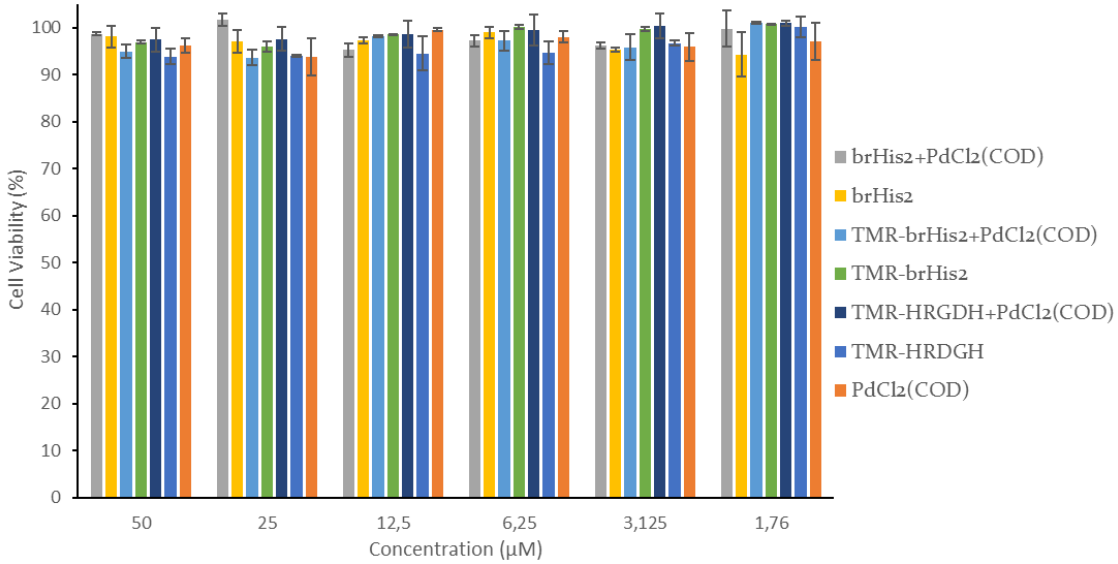


Figure 85. The palladopeptides were prepared by mixing the peptides and $PdCl_2$(COD) (1:1 ratio) in water for 10 min before the addition to cells. Then, HeLa cells were incubated with the palladopeptides or only with $PdCl_2$(COD) at different concentrations, for 24 h at 37 ºC. Finally, the cells were washed twice with PBS before carrying out the MTT assays. Data are mean ± SEM of three independent experiments performed in triplicate.

Selective intracellular reactivity

Finally, given that the cellular uptake of the palladium chelated complex of **HRGDH** in cells might be partially associated to the presence of integrin receptors, we anticipated that its cell penetration would depend on the type of cells and the presence of these receptors. Indeed, we were glad to observe that while in A549 or HeLa cells there is intracellular catalysis, in the MCF-7 breast cancer cell line, which expresses integrin at low levels, we barely detect any fluorescence. This preliminary result prompts well for the development of target selective intracellular metal catalysis.

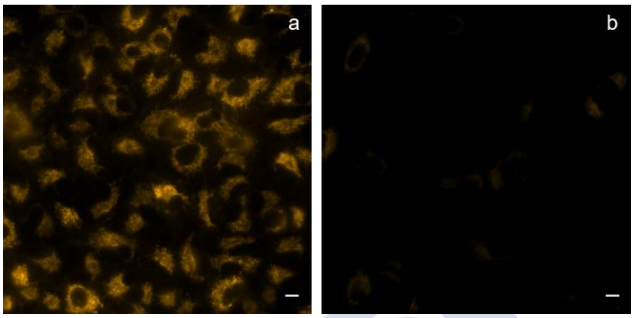


Figure 86. Fluorescence micrographies of A549 (a) and MCF-7 cells (b), using standard incubation conditions with the peptide-palladium complex resulting from TMR-**HRDGH** and $PdCl_2$(COD) (1:1 ratio). λexc = 385 nm, λem = 515-700 nm. Scale bar: 10 μm.

Conclusions

In summary, arginine-rich miniproteins made of natural amino acids, and equipped with *i, i+4* bis-histidine residues, can efficiently react with different Pd(II) salts to form stapled derivatives that show enhanced cell internalization properties. More importantly, the pallado-miniproteins obtained when using $PdCl_2$(COD) as palladium source, work as effective metalloreactors to promote depropargylation reactions inside living mammalian cells; transformations that cannot be performed using just the palladium sources. The efficiency of the approach is likely associated to a synergistic beneficial effect of the conformational constraint introduced by the metal bridge, and a protective role of the peptide scaffold, which avoids a rapid deactivation of the metal.

These results represent a first step towards the development of a "bottom-up" strategy for the generation of artificial catalytic metalloproteins capable of working in the native living environment of enzymes. Additionally, the well-known transformative potential of palladium catalysis prompts well for further applications of the strategy in other type of reactions.

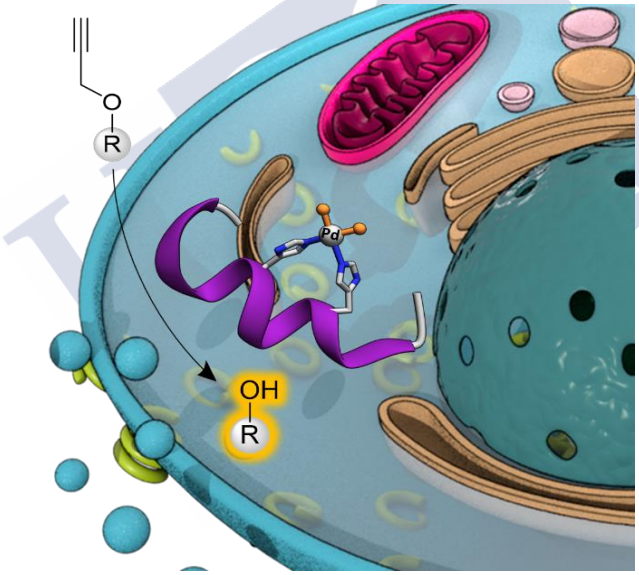


Figure 87. Schematic illustration of cell penetration of the pallado-miniprotein and its capacity to perform depropargylation reactions.

Chapter IV: Assembly of Artificial Metalloproteins In Live Mammalian Cells

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Introduction

Over one-third to one-half of all proteins contain a metal ion in their structure and thus they are called *metalloproteins*. The presence of the metal affects structure and properties of the protein, and in most of the cases, it is essential for their biological function.[214] Among all the metals, iron, zinc, magnesium, manganese and calcium are the most abundant; each one presenting different coordination and redox properties. For example, iron and copper are redox-active metals that can be switched between their oxidized and reduced forms, and normally participate in electron transfer processes.[215] They are also involved in dioxygen storage and carriage via metalloproteins such as hemoglobin (Figure 88). Other metals like zinc or magnesium are redox-inert and are used by the enzymes to stabilize negative charges, but they are also capable of working as Lewis acid in enzymatic processes.[216]

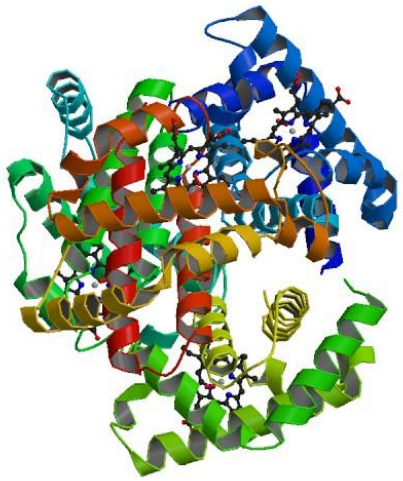


Figure 88. Crystal Structure of T State Hemoglobin with Oxygen Bound at Four iron-containing Heems units (PDB 1GZX).

As we have seen previously in TFs, metals can play structural and regulatory functions in many proteins, but they are especially relevant in catalysis.[217] Catalytic metalloproteins exhibiting a metal cofactor bound to their active site, are called *metalloenzymes*. Metalloenzymes catalyze many different types of biological reactions with high efficiency and remarkable selectivity. Some of these reactions have been even transferred to the bench of synthetic chemists and, indeed, the field of biocatalysis has led to many important applications, and continues to grow.[218]

---

[214] A. J. Thomson, H. B. Gray, *Curr. Opin. Struct. Biol*. **1992**, *2*, 155-158.

[215] D. E. Fenton, *Biocoordination Chemistry* **1995,** Oxford University Press, UK.

[216] H. B. Gray, *Proc. Natl. Acad. Sci. USA*. **2003**, *100*, 3563-3568.

[217] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton, *J. Biol. Inorg. Chem*. **2008**, *13*, 1205-1218.

[218] a) K. Drauz, H. Waldmann, *Enzyme Catalysis in Organic Synthesis* **2002,** Wiley-VCH Verlag GmbH; b) W. D. Fessner, T. Anthonsen, *Modern Biocatalysis* **2008,** Wiley-VCH Verlag GmbH & Co. KGaA.

Given that the number of transformations available for natural metalloenzymes is relatively small, there has been a long-standing interest in building artificial catalytic metalloproteins. The first example of a very primitive artificial metalloenzyme (ArM) was reported in 1956 by Fujii and co-workers. They used reduced palladium chloride adsorbed on silk fibers for the asymmetric reduction of dehydro-amino acid derivatives.[219]

At that time, isolating or producing proteins and testing their catalytic properties in the presence of metal ions was not straightforward, so the field developed very slowly. In the 1970s, two groups explored the use of proteins as scaffolds to accommodate non-native metal ions for catalytic purposes. In 1976, Kaiser and Yamamura reported that the hydrolytic enzyme carboxypeptidase A (CPA) could be repurposed into an oxidase upon substitution of the native Zn(II) by Cu(II).[220] In 1978, Whitesides and Wilson exploited the high affinity of biotinylated probes for avidin to anchor a Rh(diphosphine) within this protein.[221]

Despite the undeniable elegance of these pioneering studies, the true potential of ArMs had to await the advent of recombinant protein expression. In 1997, Distefano and Davies reported on a covalent scaffold modification of a recombinant adipocyte lipid binding protein (ALBP) with iodoacetamido-1,10-phenanthroline. Upon complexation with Cu(II), the resulting ArM catalyzed the stereoselective hydrolysis of racemic esters.[222] These papers set the stage for the resurgence of interest in artificial metalloenzymes at the turn of the millennium. In fact, in recent years there has been a boost in the development of transition metal protein hybrids capable of promoting otherwise difficult reactions (Figure 89).[223]

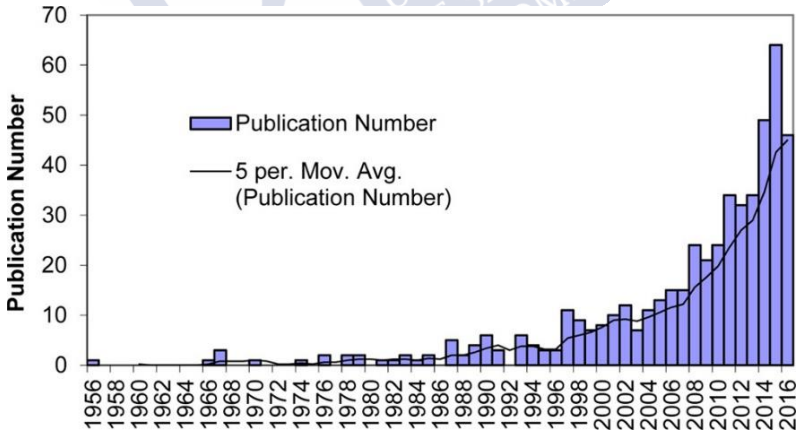


Figure 89. Number of publications concerning artificial metalloenzymes and the five-year moving average of this number plotted versus publication year.[223]

[219] S. Akabori, S. Sakurai, Y. Izumi, Y. Fujii, *Nature.* **1956**, *178*, 323-324.

[220] K. Yamamura, E. T. Kaiser, *J. Chem. Soc. Chem. Commun.* **1976**, 830-831.

[221] M. E. Wilson, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306-307.

[222] R. R. Davies, M. D. Distefano, *J. Am. Chem. Soc.* **1997**, *119*, 11643-11652.

[223] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142-231.

The assembly of the metalloenzymes has been generally achieved using one of the following strategies:

- **Covalent anchoring**: involves an irreversible reaction between metal-binding cofactors bearing a reactive functional group and an amino acid side-chain on the protein scaffold. An alternative consists of the *in vivo* incorporation of unnatural amino acids capable of binding a metal ion, using the amber stop codon suppression methodology.[224] It offers an exquisite degree of control over the position of the metal complex inside the protein.[225]

- **Dative**: relies on the coordination of a nucleophilic amino acid residue (His, Cys, Glu, Asp, Ser, etc.) to an unsaturated metal center. This type of anchoring and activation of the metal often complements either covalent or supramolecular strategies.[226]

- **Supramolecular interaction**: exploits the high affinity that proteins may display for a limited set of noncovalent inhibitors, natural cofactors, or substrates that could be eventually conjugated to a metal complex. This interaction could be supplemented also with a covalent modification.[227]

- **Metal substitution**: the native metal of a metalloenzyme may be substituted for another metal. The metal may be part of a prosthetic group or bound solely to amino acids.[228] Upon substituting the metal, new-to-nature reactivities can be introduced in the ArM's repertoire. In most of the cases, screening of the best options is based on *directed evolution* methods.[229] Introduced in the early nineties by Arnold, Chen and Stemmer,[230] this technique involves repeating cycles of gene mutagenesis, expression, and screening, thus mimicking the process used by nature to create new catalysts.

[224] a) L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science*. **2001**, *292*, 498-500; b) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413-444.
[225] H. Yang, P. Srivastava, C. Zhang, J. C. Lewis, *ChemBioChem.* **2014**, *15*, 223-227.
[226] D. Joseph Sommer, M. David Vaughn, G. Ghirlanda, *Chem. Commun.* **2014**, *50*, 15852-15855.
[227] a) J. Bos, W. R. Browne, A. J. M. Driessen, G. Roelfes, *J. Am. Chem. Soc.* **2015**, *137*, 9796-9799; b) T. Heinisch, T. R. Ward, *Acc. Chem. Res.* **2016**, *49*, 1711-1721; c) W. J. Song, F. A. Tezcan, *Science.* **2014**, *346*, 1525-1528.
[228] S. N. Natoli, J. F. Hartwig, *Acc. Chem. Res.* **2019**, 52, 326-335.
[229] H. Renata, Z. J. Wang, F. H. Arnold, *Angew. Chemie Int. Ed.* **2015**, *54*, 3351-3367.
[230] a) K. Chen, F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5618-5622; b) W. P. C. Stemmer, *Nature* **1994**, *370*, 389-391.

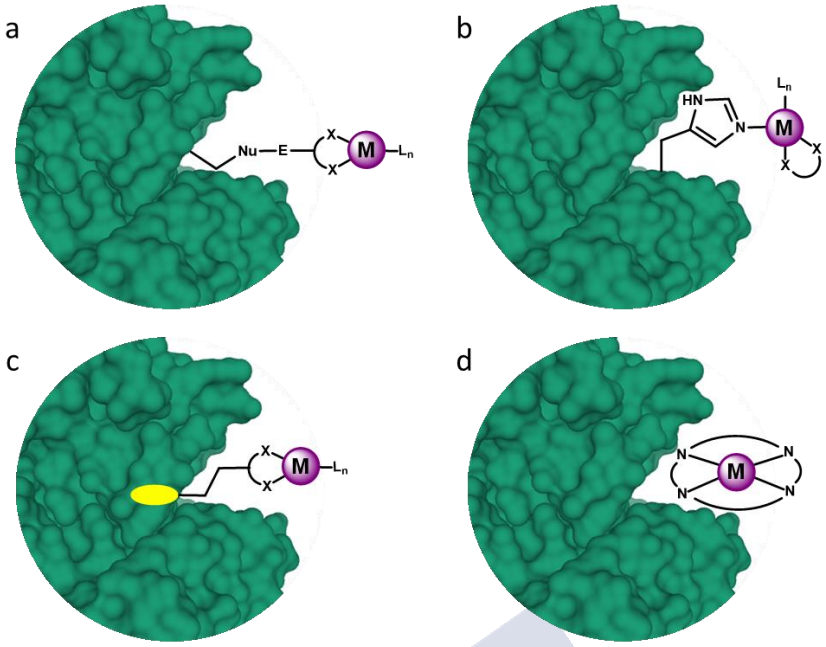


Figure 90. General approaches for the building of ArMs: a) Covalent immobilization. b) Dative coordination with an unsaturated metal complex. c) Supramolecular coordination using a high-affinity anchor. d) Metal substitution.

Using these strategies, a good number of active metalloenzymes that catalyze different type of transformations have been made.[231] Most of the reactions have been carried out in "*in vitro*" environments, and with synthetic purposes.

A major goal in the field consists in demonstrating that artificial metalloproteins can work in live settings, somewhat mimicking natural enzymes. Unfortunately, advances in this direction have been very slow. Recently, the Ward group demonstrated that conjugating a metalloprotein scaffold with a cell-penetrating module allows to achieve abiotic reactions inside cells.[232] In addition, the Mahy group reported an interesting approach to assemble an artificial metalloenzyme on cellular membranes by taking advantage of interactions of designed ligands with guanine nucleotide receptors expressed in the membrane.[233]

Therefore, creating reactive metalloproteins that work in the interior of living mammalian cells, and can truthfully mimic and complement natural enzymes, is an extraordinary challenge, that has not been yet realized. We envisioned that a potential approach to address these challenges could be based on a selective covalent linkage between externally added metal-containing ligands, that should easily transverse cellular membranes, and intracellularly expressed proteins containing reactive handles. In this context, we were especially attracted by the recently developed self-tagging

[231] a) H. J. Davis, T. R. Ward, *ACS Cent. Sci.* **2019**, *5*, 1120–1136; b) T. Matsuo, T. Miyake, S. Hirota, *Tetrahedron* **2019**, *60*, 151226–151233.
[232] Y. Okamoto, R. Kojima, F. Schwizer, E. Bartolami, T. Heinisch, S. Matile, M. Fussenegger, T. R. Ward, *Nat. Commun.* **2018**, *9*, 1943-1949.
[233] W. Ghattas, V. Dubosclard, A. Wick, A. Bendelac, R. Guillot, R. Ricoux, J. P. Mahy, *J. Am. Chem. Soc.* **2018**, *140*, 8756-8762.

technologies based on specific enzymes that are capable of selecting designed ligands and make a covalent interaction with them.

These self-labeling strategies include the SNAP-tag,[234] the CLIP-tag[235] and the HALO-Tag[236] methods. These enzymes, that can be genetically fused to a host protein, have been demonstrated to react quite efficiently, and in intracellular sites, with paired exogenously substrates equipped with fluorescent dyes. Because labeling by these strategies is irreversible and quantitative, they have opened up powerful options for live cell imaging approaches with a rich choice of different fluorophores. The SNAP-tag can be covalently labeled with O^6-benzylguanine derivatives, the CLIP-tag with O^2-benzylcytosine derivatives and HALO-Tag binds irreversibly chloroalkane ligands (Figure 91). These platforms are being employed for different applications like protein isolation and purification, analysis of protein function, studying protein–protein and protein–DNA interactions, performing biological assays, *in vitro* cellular imaging, and *in vivo* molecular imaging. One additional advantage of these genetic fusion techniques is that they can be designed to target selected organelles in live cells. Amongst others, they have been used to label proteins in the nucleus, in the ER, in mitochondria, as well as in components of the cytoskeleton.[237]

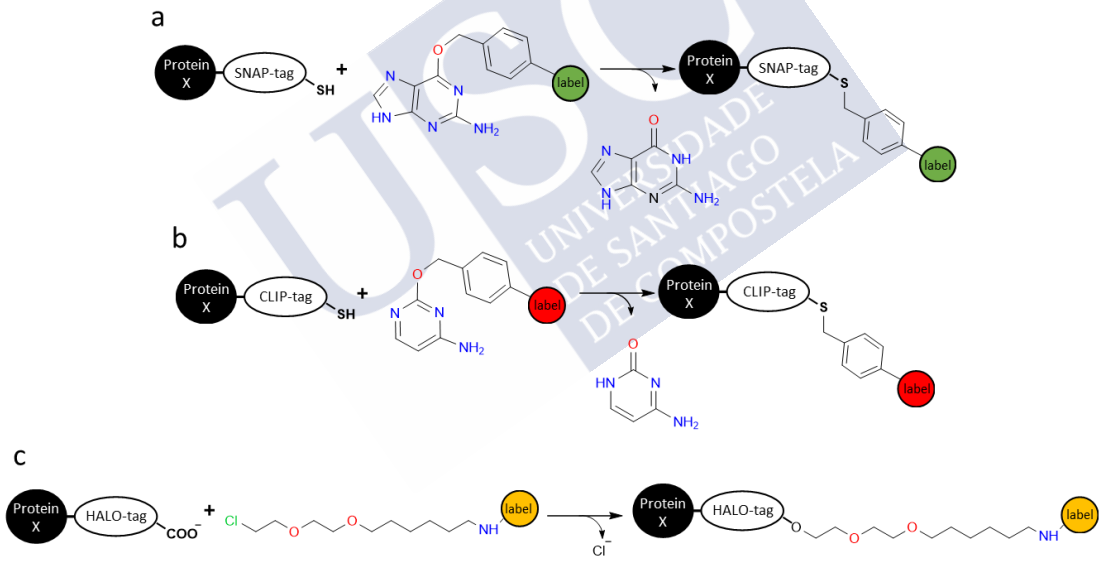


Figure 91. SNAP- CLIP- and HALO-tag Mechanism: these tags fused to the protein of interest labels themselves with "X" releasing a) guanine b) cytosine and c) Cl^-.

[234] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86-89.
[235] A. Gautier, A. Juillerat, C. Heinis, I. R. Jr. Corrêa, M. Kindermann, F. Beaufils, K. Johnsson, *Chem. Biol.* **2008**, *15*, 128-136.
[236] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, K. V. Wood, *ACS Chem. Biol.* **2008**, *3*, 373-382.
[237] a) D. Srikun, A. E. Albers, C. I. Nam, A. T. Iavarone, C. J. Chang, *J. Am. Chem. Soc.* **2010**, *132*, 4455-4465; b) B. G. Hughes, X. H. Fan, W. J. Cho, R. Schulz, *Am. J. Physiol.: Heart Circ. Physiol.* **2014**, **306**, H764-H770; c) J. C. Tseng, H. A. Benink, M. G. McDougall, I. Chico-Calero, A. L. Kung, *Curr. Chem. Genomics*. **2012**, *6*, 48-54.

In our work, we have focused on the HALO-tag technology because of its efficiency and simplicity. HALO-tag is a 297-residue peptide (33 kDa) derived from a bacterial enzyme designed to covalently bind to a synthetic ligand and work as a haloalkane dehalogenase. The mechanism consists in removing halides from halide-containing aliphatic hydrocarbons by a nucleophilic displacement mechanism, involving the carboxylic acid of an aspartate residue.[238] Base-catalyzed hydrolysis of the resulting covalent intermediate releases the hydrocarbon as an alcohol. The based-catalyzed cleavage is mediated by a conserved histidine located near the aspartate nucleophile (Figure 92). [239] Previous studies showed that mutating this histidine to phenylalanine, leads to the formation of a stable ester bond. [236] The reaction that forms the bond between the protein tag and the chloroalkane linker is fast and essentially irreversible under physiological conditions.

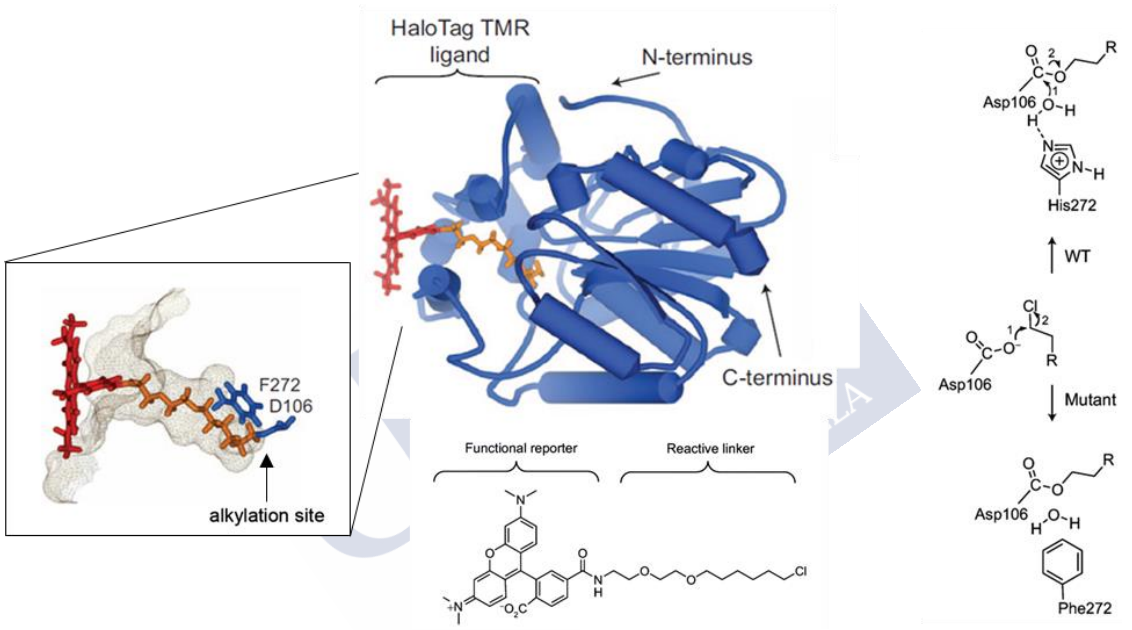


Figure 92. *Left:* The HALO-Tag protein (model) and close-up of the ligand tunnel (outlined by a mesh Connolly surface) with a covalently bound TMR ligand. The ligand has a fluorescent part (reporter) and a reactive linker with a terminal chloride. *Right*: Protein alkylation occurs when the chloride on the ligand gets displaced, upon nucleophilic attack by the carboxyl group of an aspartic acid (D106) side chain. As a result, the HALO-tag protein is covalently linked to the fluorescent reporter, leading to a stable adduct (the phenylalanine F272 in blue, instead of a histidine in the wild type protein, is key for this stability). (Source: Adapted from[236]).

[238] D. B. Janssen, *E. Curr. Opin. Chem. Biol.* **2004**, *8*, 150-159.

[239] F. Pries, G. H. Kingma, C. M. Jeronimus-Stratingh, A. P. Bruin, D. B. Janssen, *Biol. Chem.* **1995**, *270*, 10405-10411.

Objective

The long-term objective of our work in this topic is the generation of catalytically active artificial metalloproteins inside living cells using the HALO-Tag technology. This is a challenging goal that requires to be successful in several specific tasks: 1) The design and synthesis of well-designed chloroalkane derivatives incorporating the reactive metals in suitable positions, 2) the cellular penetration of these ligands; 3) the suitable assembly within the catalytic site of the HALO-tag enzyme, 4) the demonstration that the metal complex, once ligated to the enzyme, is capable of promoting the desired reaction. We first focused on the building of ruthenaproteins capable of performing deallylation or depropargylation reactions and uncage fluorescence probes.

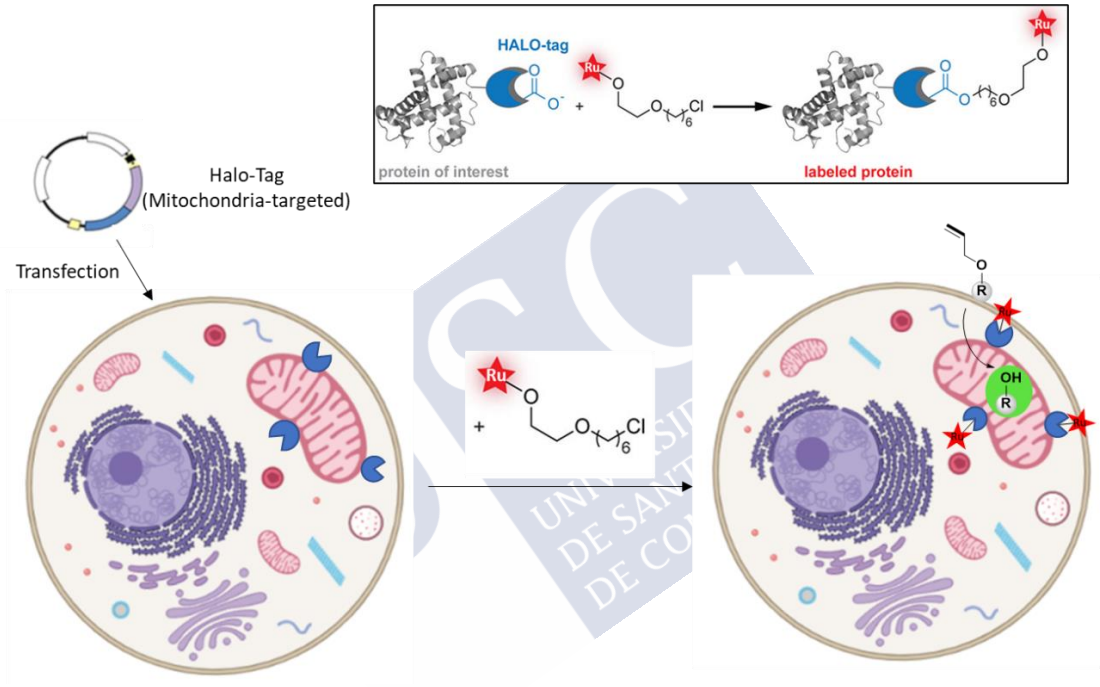


Figure 93. *Top:* Schematic illustration of the assembly of an artificial metalloprotein by the HALO-Tag system, using a ruthenium complex as a cofactor. *Bottom:* Schematic illustration of cells transfected with a HALO protein targeted to mitochondria. HALO is expressed in this organelle and after the addition of a ruthenium-containing HALO-tag ligand and subsequent binding, the ruthenaprotein is ready to perform the deallylation reaction.

Results and discussion

Development of the HALO-Tag protein system

Creating the metalloprotein required to develop an optimal chloroalkane ligand tethered to the metal complex, and capable of inserting into the HALO-tag binding pocket.

For creating the protein tag, we have employed a modified bacterial haloalkane dehalogenase designed to form a covalent bond with synthetic ligands. Using the phenyl alanine mutant of the natural dehalogenase previously commented, we designed and made different organelle-targeting HALO-Tag proteins.

The proteins contain two parts: An N-terminal HALO-Tag unit, required to react with bioorthogonal HALO-Tag ligands, and an organelle-targeting sequence. When expressed in mammalian cells, the organelle-targeting sequence directs expression of the HALO protein inside the lumen of the organelle of interest. The localized HALO-Tag protein might then react with the HALO-Tag ligands. We have constructed expression vectors to target proteins into the nucleus, endoplasmic reticulum (ER), cytosol and plasma membrane. In addition, we designed a negative control vector where the aspartate 106 (active site) is mutated to alanine (D106A),[240] in order to corroborate the specificity of the system (details in the Experimental Part).

In addition to cloning the mammalian constructs, we also designed bacterial plasmids with the HALO protein and a fusion tag, His-tag, to obtain purified protein through nickel-resin affinity in order to perform the *in vitro* assays. We also expressed in bacteria the construct HALO-mutated(D106A).

The chloroalkane-containing ligands

Other members of the group were able to synthesize a variety of ruthenium-containing structures that should work as HALO-tag ligands and were additionally equipped with different fluorophores. We selected those indicated in the figure, considering their homology to previously developed fluorophore-containing ligands.

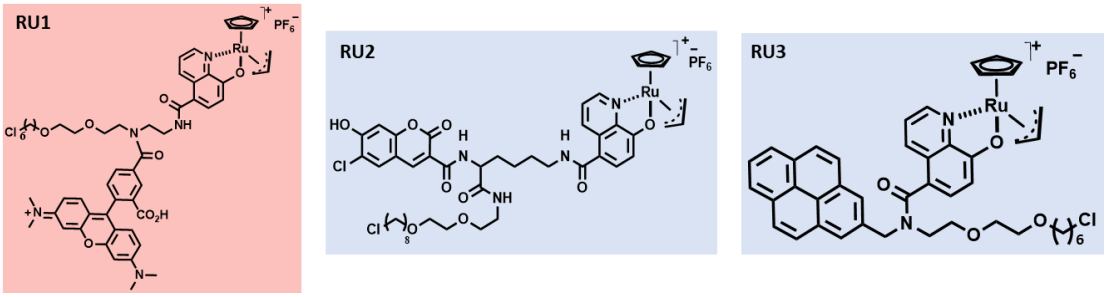


Figure 94. Structures of ruthenium-containing chloroalkane ligands.

[240] T. K. Neklesa, H. S. Tae, A. R. Schneekloth, M. J. Stulberg, T. W. Corson, T. B. Sundberg, K. Raina, S. A. Holley, C. M. Crews, *Nat. Chem. Biol.* **2011**, *7*, 538–543.

RU1 presents TAMRA as a fluorophore and it has a chloroalkane chain of 6 carbon units. **RU2** features a Coumarin, a carbon chain of 8 units and Lysine as a linker connecting the metal and the chloride. **RU3** presents a pyrene as fluorophore.

In vitro assays

We initially investigated the formation of the metalloprotein *in vitro,* using purified HALO proteins obtained by bacterial expression (both the HALO protein and the HALO-mutated(D106A) as negative control) (see the Experimental Part for details) and the above indicated chloroalkane ligands. To carry out the binding reaction, the proteins were quantified using a Nanodrop Spectrophotometer and mixed at a concentration of 10 μM with the ligands (20 μM) in PBS at 37ºC. After one hour, the proteins were concentrated using *Amicon® Ultra Centrifugal Filters*, carrying out two wash steps with PBS to ensure the removal of excess ligand. We measured the concentration of the proteins using a Nanodrop Spectrophotometer and performed a polyacrylamide gel electrophoresis (SDS-PAGE) assay to identify the possible incorporation of the ligands into the protein by fluorescence.

As we can see in the Figure 95, we were glad to detect fluorescence in the bands corresponding to reaction crudes resulting from the mixing of the ligands with the HALO protein that contains the reactive Asp residue (Figure 95, left, lanes 1-3). However, the same experiments with the protein mutated in that residue did not provide any fluorescent bands (Figure 95, left, lanes 4-6). Coomassie Brilliant Blue protein staining detected equal amounts of protein at the expected MW of 33 KDa (Figure 95, right panel) demonstrating that the variations of fluorescence levels are the result of differences in reactivity and not owed to different amounts of protein being used for each experimental condition.

Figure 95. *Left:* Polyacrylamide gel: Lanes 1-3: reactions made with HALO protein (20 μg each line); Line 1: reaction performed with **RU1**, Line 2: reaction performed with **RU2**, Line 3: reaction performed with **RU3**. Lanes 4-6: reactions made with HALO-mutated(D106A) protein (20 μg each line); Line 4: reaction performed with **RU1**, Line 5: reaction performed with **RU2**, Line 6: reaction performed with **RU3.** *Right:* Same polyacrylamide gel dyed with Coomassie Brilliant Blue to detect the presence of proteins: Lanes 1-3: HALO protein. Lanes 4-6: HALO-mutated(D106A) protein. SDS-PAGE experiments were resolved on a 12,5% denaturing polyacrylamide gel (with SDS) and Tris-glycine-SDS buffer over 40 min at r.t. Analysis was performed directly using a ChemiDoc MP Imaging System (*Bio-rad*) for fluorescence visualization (Left) or by staining with Coomassie Brilliant Blue for 5 min (Right).

These results suggest that we can build the desired conjugates, and therefore we decided to test their ability to promote the designed ruthenium-promoted reactions *in vitro*.

We first prepared the protein-ruthenium complexes following the protocols described above (including washes to removal of excess ligand). The reactivity assays were carried out using Rho(alloc)$_2$ as probe (Figure 96, top). We performed these reactions in PBS during 4h at 37ºC, using 10 μM of protein-complex conjugate and 50 μM of probe. We assumed that all the protein is bound to the ruthenium complex, although the proportion of the metal is probably lower. In addition to the reaction with the protein-ruthenium complexes we also carried out control reactions with the Ru-containing ligands without the protein, including one without the chloroalkane chain (Figure 96, bottom right).

As we can see in Figure 96, using either **RU1** or the product resulting from its conjugation to the protein we were able to observe a clear increase in fluorescence that must be associated to the formation of the desired product. However, using the crude product mixture resulting from mixing the HALO protein with **RU3** we did not observe fluorescence. Moreover, if the HALO ligand used for the coupling lacks the chloride-chain, the protein-metal hybrid is not formed, and this is the reason why we did not observe reactivity (HALO-**RUcontrol**). Despite these hopeful results, the conversion using the protein conjugates was very low (less than 10%, see the Experimental Part for the use of calibration curves).

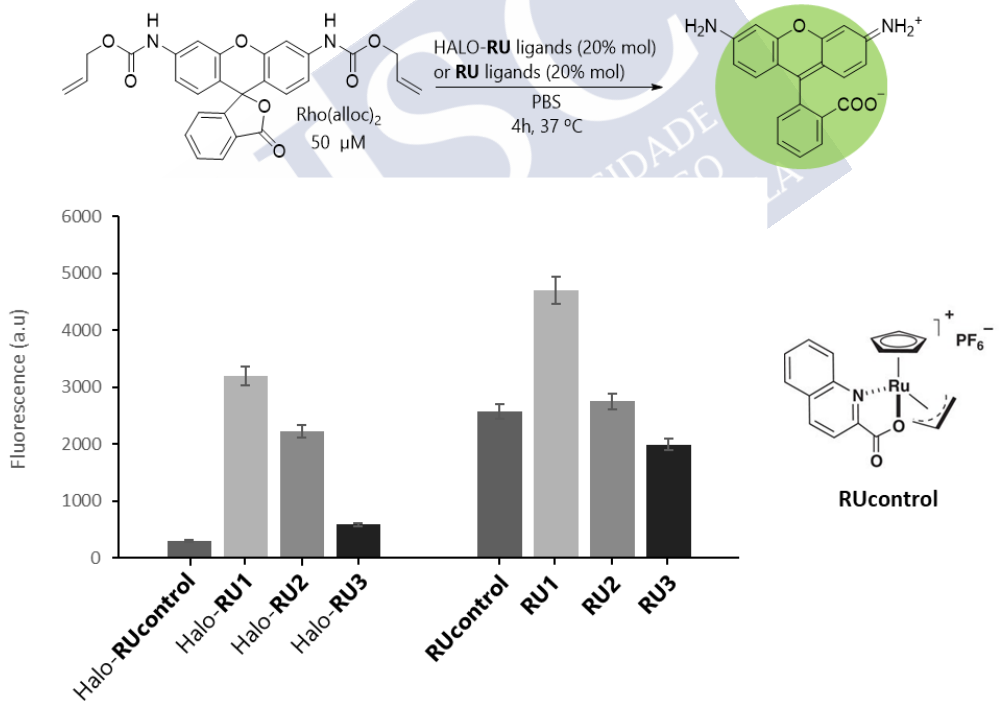


Figure 96. *Top:* Uncaging of Rho(alloc)$_2$. *Bottom Left:* Bar diagram representation of the fluorescence obtained after the reaction with different ruthenium containing species. The HALO-**RU** species were obtained mixing the protein (10 μM) with the ligands (20μM) during 1h in PBS at 37 ºC. Then, the proteins were concentrated using *Amicon® Ultra Centrifugal Filters*, carrying out two wash steps with PBS to ensure the removal of excess ligand. Reaction conditions: Rho(alloc)$_2$ (50 μM), HALO-**RU** species or **RU**-ligands (approx. 10 μM, 20 mol%) in 100 μL of PBS at 37 ºC for 4 h. The reactions were measured in a microtiter plate reading spectrophotometer by recording the fluorescence (Excitation Wavelength: 485nm, Emission Wavelength: 535nm). Data are mean ± SEM of three independent experiments performed in triplicate. *Bottom Right:* Structure of **RUcontrol**.

To further demonstrate that there is reactivity with the hybrids, we made a **RU**-ligand with the same structure of **RU1** but without the terminal chlorine (**RU4**), as well as the chloroalkane-ligand without ruthenium (**L1**). We performed the same protocols for the protein binding reaction and resolved the proteins by SDS-PAGE. The HALO-mutated(D106A) protein was again used as a negative binding control experiment.

The results confirmed that **L1** binds very well to the protein (Figure 97, Lane 1), but as expected, using the chloride-free ligand **RU4** (Figure 97, Lane 3), the interaction is much worse than with **RU1** (Figure 97, Lane 2). Using the HALO-mutated(D106A) protein and **RU1** we observed no ligation (Figure 97, Lane 4).

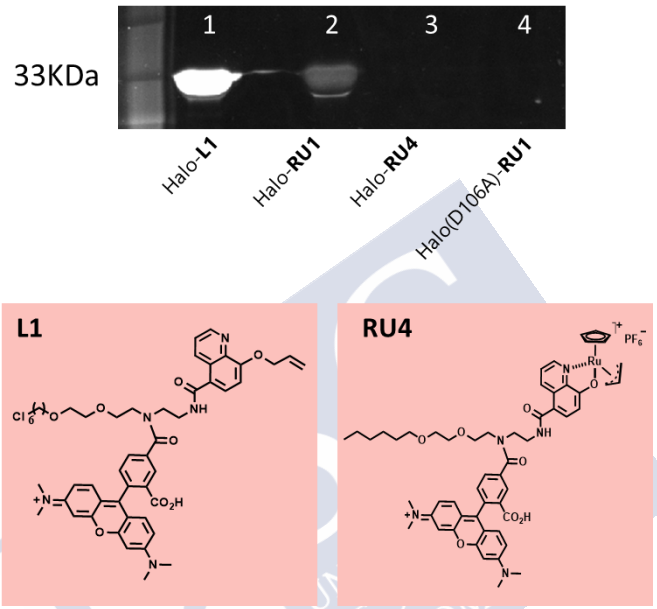


Figure 97. *Top:* SDS-PAGE: Lanes 1-3: conjugation reactions to the protein HALO (20 µg each line); Line 1: with **L1**, Line 2: positive control with **RU1**, Line 3: with the dechlorinated ligand **RU4**. Lanes 4: conjugation between the negative control HALO-Mutated(D106A) protein and **RU1**. *Bottom:* Structures of new chloroalkane-ligands. SDS-PAGE experiments were resolved on a 12,5% denaturing polyacrylamide gel (with SDS) and Tris-glycine-SDS buffer over 40 min at r.t and analyzed directly using ChemiDoc MP Imaging System from *Bio-rad* by fluorescence visualization.

The detection of fluorescence in the gel confirms the presence of the fluorophore of the ligand, but not the presence of ruthenium. Therefore, we carried out ICP/MS analysis of these gel bands. For this purpose, we cut the bands of this gel and digested them with Nitric acid (70%). The data obtained in the ICP/MS analysis allowed us to deduce that the amount of ruthenium detected in the reaction with **RU1** corresponded to a 10% molar ratio with respect to the total amount of protein, confirming that the desired conjugation takes place, but the yield is modest. As expected, the protein resulting from reaction with **RU4** lacked ruthenium (less 0.5% with respect to the total protein). In the case of the reaction of the HALO-mutated(D106A) with the ligand **RU1**, the ICP revealed about a 1% of ruthenium.

All these results together demonstrated that the covalent complex is formed, albeit the binding reaction presents a low efficiency, and the ability of the resulting complex to

catalyze the deallylation reaction is also very modest. Despite the low efficiency, we decided to carry out a test of the behaviour of the complex in living cellular environments.

Intracellular assembly

We first checked the assembly of the protein-ligand complexes inside mammalian cells using organelle-targeted HALO proteins. We transfected the different constructs in HeLa Cells for 48h, and after one wash with PBS, we added the ruthenium-containing ligands for 1 hour (10 μM). As it can be seen in the Figure 98, we observed fluorescence in the desired intracellular compartment when **RU1** was used as ligand for the corresponding HALO protein. On the contrary, we could not detect any fluorescence when HALO-mutated(D106A) was used as binding protein for the ligand.

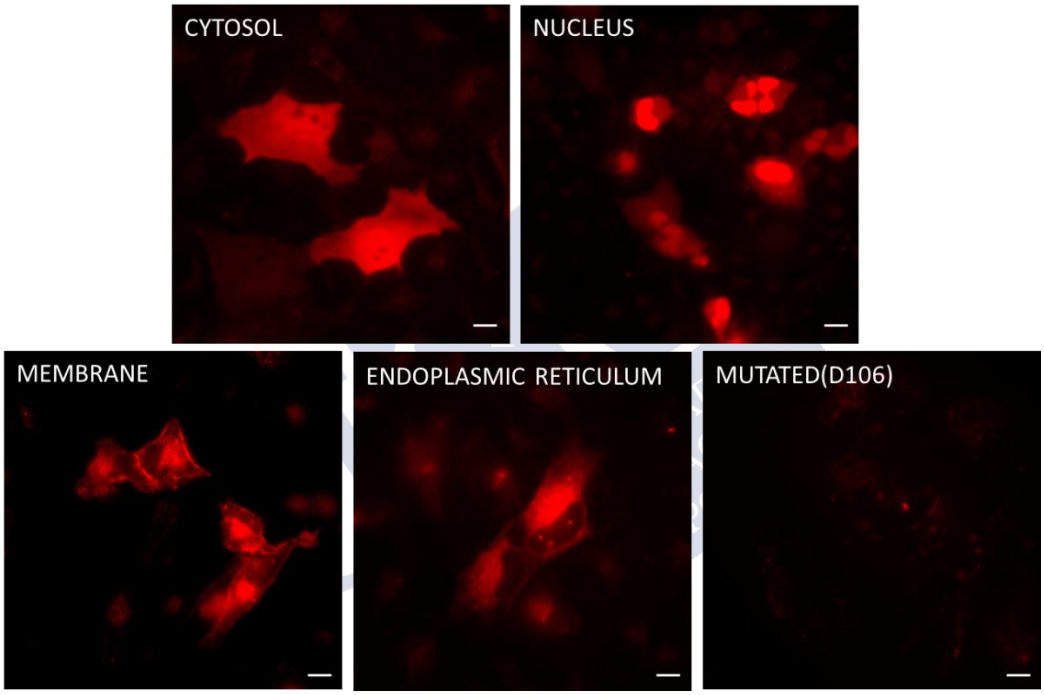


Figure 98. Fluorescence micrographies of HeLa cells (confocal). First, plasmids were transfected using *FuGENE*® HD Transfection Reagent. After 48h, we washed the cells with PBS and incubated them with 10 μM of **RU1** for 1h at 37 ºC. The cells were washed twice with PBS before being observed in a confocal microscope. All incubations were made in Dulbecco´s modified Eagle medium completed with 5% of fetal bovine serum (FBS-DMEM). λ_{exc} = 561 nm, λ_{em} = 620/60 nm. Scale bar: 10 μm. For **RU2** and **RU3** we performed the same experiment, the pictures are not shown because no fluorescence was detected.

Reactivity

We assayed the same reaction than in the *in vitro* assay, using as substrate Rho(alloc)$_2$. After the incubation of HALO transfected cells with the ruthenium-haloalkane **RU1** as indicated above, we made two washings with PBS and added the caged probe (50 μM). As a control of the intracellular activity of the unbound ligand, we also added **RU1** to cells that did not express the HALO proteins.

Unfortunately, after 4 h, we did not detect any fluorescence corresponding to the uncaged product in any organelle of the cells that were expected to contain the protein-

ruthenium conjugates (Figure 12, top). In cells expressing the HALO-mutated(D106) we did not observe any fluorescence either, indicating that the unbound **RU1** is not active inside the cells in these conditions. However, increasing the concentration of the complex to 50 µM produced a clear increase in the intracellular fluorescence, indicating that the complex was capable of promoting the deallylation reaction inside living cells if given adequate conditions (Figure 12, down, left panel).

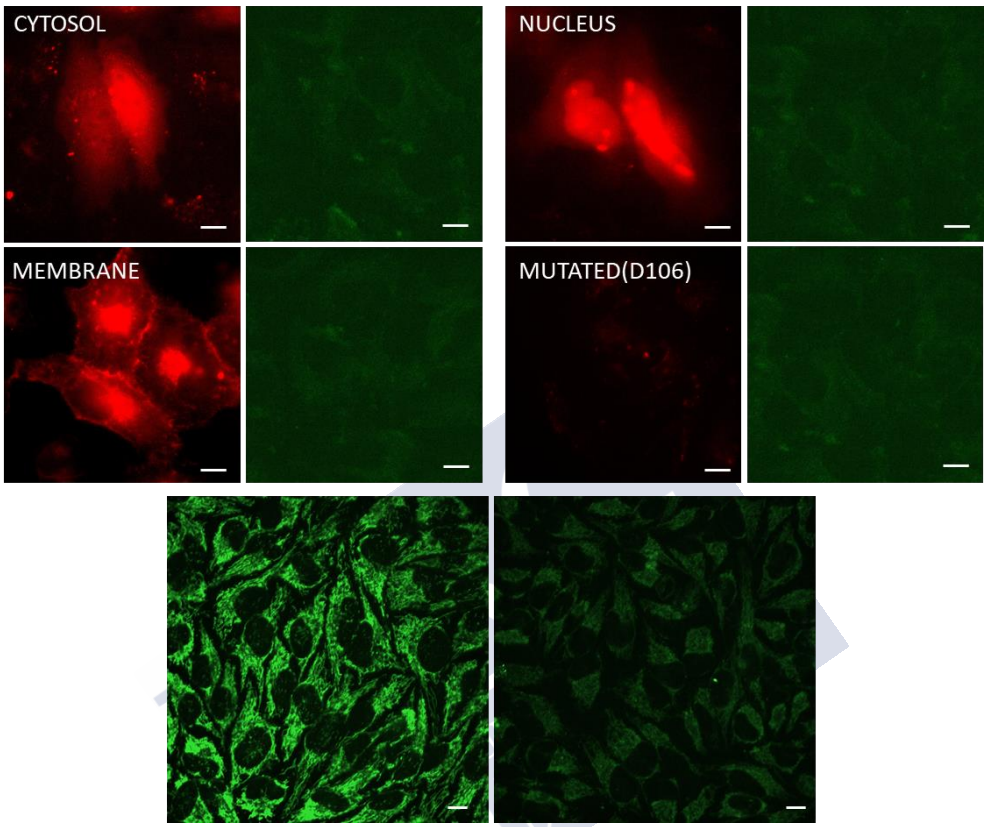


Figure 99. Intracellular deallylation of Rho(alloc)$_2$ monitored by fluorescence microscopy in HeLa cells (confocal). *Top:* Cells expressing HALO protein. First, plasmids were transfected using *FuGENE*® HD Transfection Reagent. After 48h, we washed the cells with PBS and incubated them with 10 µM of **RU1** for 1h at 37 ºC. The cells were washed twice with PBS and incubated in fresh media with 50 µM of Rho(alloc)$_2$ for 4h. Scale bar: 20 µm. *Bottom:* Control cells without HALO protein. Cells were treated with 50 µM of **RU1** for 1h at 37 ºC and washed twice with PBS to remove the excess. Then the caged probe (50 µM) was added in fresh media, and the cells were incubated for 4 h. *Left*: with **RU1**; *Right*: Rho(alloc)$_2$ alone as a control. Scale bar: 10 µm. All incubations were made in Dulbecco´s modified Eagle medium completed with 5% of fetal bovine serum (FBS-DMEM). Red channel: λ_{exc} = 561 nm, λ_{em} = 620/60 nm. Green channel: λ_{exc} = 488 nm, λ_{em} = 525/50 nm.

All these results are consistent with the formation of a small proportion of the desired ruthenaproteins inside cells; however, these are not able to work as catalysts for the desired deallylation reaction. Likely, both, the formation of the hybrid, and its catalytic activity are not efficient enough. Nevertheless, in qualitative terms, the results are promising.

Conclusions

In summary, we demonstrated that the HALO-tag system can be used for the incorporation of ligands that contain both a fluorophore and a metal complex, both *in vivo* and *in vitro*. The resulting metalloproteins were able to show some reactivity *in vitro*, but not *in vivo*. Further research to optimize the systems is needed.

General Conclusions

1) We have devised a new type of **fully-peptidic** DNA binder with a new artificial DNA binding motif. Bivalent and trivalent constructs were prepared in a straightforward manner owing to their peptidic constitution, and displayed excellent DNA recognition properties in terms of affinity and selectivity. Remarkably, we have demonstrated that these artificial DNA binders could be incorporated into expressed miniproteins as sequence specific **artificial gene activators**. We have also used **nanopore force spectroscopy** to obtain further data on the DNA interaction.

2) We have demonstrated the viability challenging **tripartite DNA binding**, involving the simultaneous formation of **two metal coordination** complexes. The recognition systems involved three individual peptide modules, to bind specific, **long sequences of DNA**. Key for the success of the approach was the dual role of the metal as an a-helix-nucleating factor and heterodimerization clip. The multicomponent nature of the system and the kinetic lability of the metal coordination facilitated the disassembly of the supramolecular structure upon addition of external agents that sequester the nickel cation.

3) We have presented a **minimal, synthetic peptide** that was entirely made of natural residues, capable of binding to specific sites of the DNA with high affinity and selectivity. The DNA binding depends on the presence of a palladium complex (or gold) that favors the α-helix nucleation required for the interaction. Importantly, the DNA binding and disassembly processes can be repeated several times, therefore providing for a fully **reversible switchable system**. In addition to these important novelties, we have demonstrated for the first time the possibility of triggering the **cell internalization** of designed basic peptides by adding a metal clipping agent.

4) We have shown that **metal-based stapled-peptides**, by favouring the alpha-helical folding ability, influenced drastically the interaction of these peptides with RAS proteins. This interaction could be **switched** in a reversible manner by using palladium chelating agents. Importantly, the resulting metallopeptides were functionally active, and behave as potent **RAS inhibitors**.

5) We have described **miniproteins** made of natural amino acids and stapled with Pd(II) salts, capable of work as effective **metalloreactors** to promote depropargylation reactions inside **living mammalian cells**; transformations that cannot be performed using just the palladium sources. The efficiency of the approach was likely associated to a synergistic beneficial effect of the conformational constraint introduced by the metal bridge, and a protective role of the peptide scaffold, which avoids a rapid deactivation of the metal. These results represent a first step towards the development of a "bottom-up" strategy for the generation of artificial catalytic metalloproteins capable of working in the native living environment of enzymes.

6) We have demonstrated that the **HALO-tag** system can be used for the incorporation of ligands that contain both a fluorophore and a metal complex, both *in vivo* and *in vitro*. The resulting **metalloproteins** were able to show some reactivity *in vitro*, but not *in vivo*. Further research to optimize the systems is needed.

Summary

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter I. Section 1:
DNA-binding miniproteins based on zinc fingers.

Obtaining artificial proteins that mimic the DNA binding properties of natural transcription factors could open new ways to manipulating gene expression. Inspired by the modularity of natural transcription factors, we have designed synthetic miniproteins that combine the zinc finger module of the transcription factor GAGA and AT-hook peptide domains.

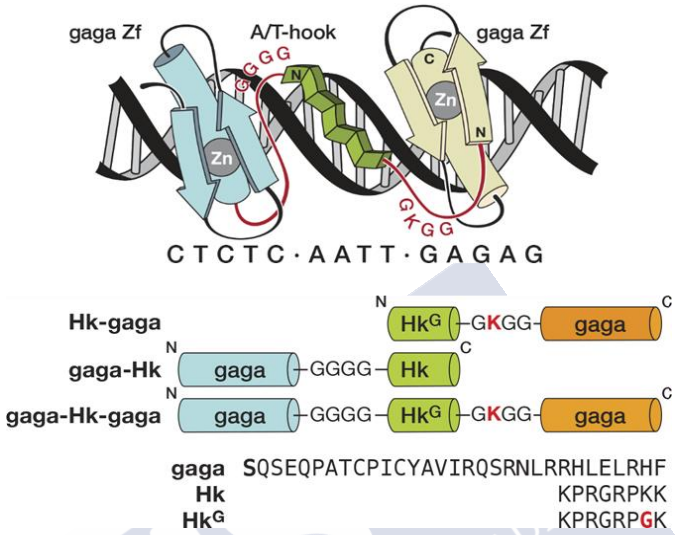


Figure 1. Representation of the synthesized hybrids and sequences of the modules.

Having at hand the desired bivalent conjugates, we studied their DNA binding properties using EMSA. Thus, oligonucleotide containing the AT-hook and GAGA binding sites, was mixed with increasing concentrations of the conjugates **gaga-Hk** or **Hk-gaga**. The gels showed concentration-dependent slow-migrating bands in both cases, which were consistent with the formation of the desired complexes peptides/DNA. Fluorescence anisotropy titrations confirmed that both peptides bind with high affinity to their target sites.

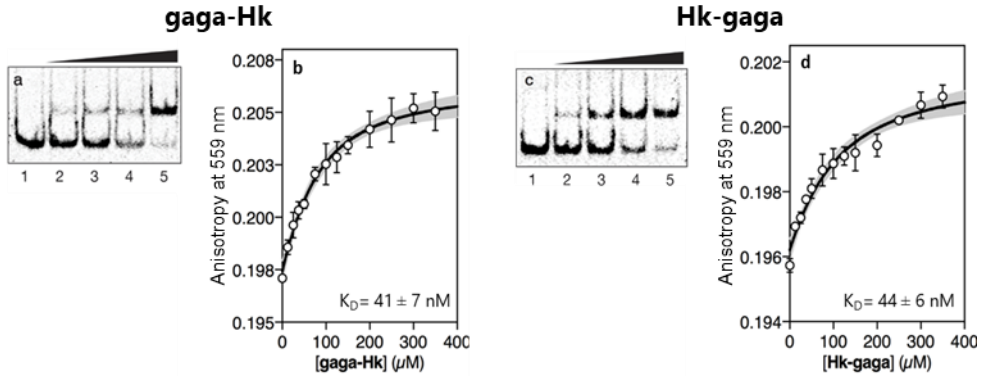


Figure 2. *Panel a and Panel c*: Polyacrylamide gel electrophoresis with target dsDNA and increasing concentrations of **gaga-Hk** (a)**/ Hk-gaga** (c). *Panel b and Panel d*: Anisotropy titration of TMR-target dsDNA with increasing concentrations of **gaga-Hk** (b) and **Hk-gaga** (d).

We then moved to the more challenging ternary "major-minor-major" groove interaction. Gratifyingly, addition of increasing concentrations of the ternary chimera **gaga-Hk-gaga** to its target dsDNA, produced a slower migrating band (Figure 3a), consistent with the formation of ternary miniprotein/DNA complex. Importantly, the synthetic miniprotein did not elicit retarded bands when incubated with a sequence lacking both GAGA binding sites (Figure 3b, lanes 1-5), or missing the A/T-rich site (Figure 3b, lanes 6-10). Fluorescence anisotropy titrations using a TMR-target dsDNA confirmed the high affinity binding of the trivalent peptide chimera (Figure 3c). These results supported the formation of a trivalent DNA complex at the specific DNA site of 14 base pairs.

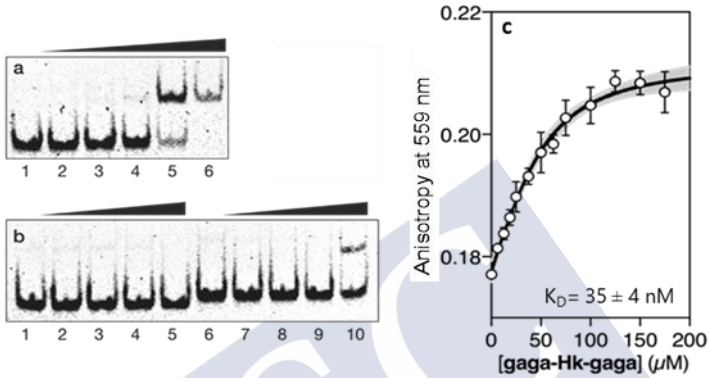


Figure 3. a) Target dsDNA with increasing concentrations of **gaga-Hk-gaga**. b) Lanes 1-5, gaga-mutated dsDNA with **gaga-Hk-gaga.** Lanes 6-10, AT-mutated dsDNA with **gaga-Hk-gaga**. c) Anisotropy titration of TMR-target dsDNA with increasing concentrations of **gaga-Hk-gaga.**

We next analyzed the interaction of the trivalent chimera with DNA using nanopores which works by stochastically examining DNA states in presence of a given amount of the peptide binder, deducing a K_D. The results agreed with the obtained by anisotropy.

The linear and fully peptidic structure of our peptides has allowed their expression in mammalian cells. After engineering and cloning our sequences in the appropriate vectors, and performed the cell reporter´s assays, the results demonstrated that our non-natural peptides could increase gene expression processes in mammalian cells that contained designed sites in their genome. These results open the door to generation of new genetic tools based in this typed of miniaturized, artificial versions of transcription factors, shorter than the zinc fingers described until now.

Chapter I. Section 2:

Assembly of a ternary metallopeptide complex at specific DNA sites mediated by an AT-Hook adaptor.

We have explored the application of metal coordination as a powerful approach to obtaining multimeric DNA binders. For our design, we equipped an AT-Hook with two bipyridine ligands (**HkBpy$_2$**) to recruit GCN4 peptide modified with two His residues (**brHis$_2$**) to adjacent DNA major grooves, in response to a metal clip (Nickel(II), Figure 4). The Ni(II) mediates the self-assembly of the peptide binders and also acts as α-helix nucleating agent.

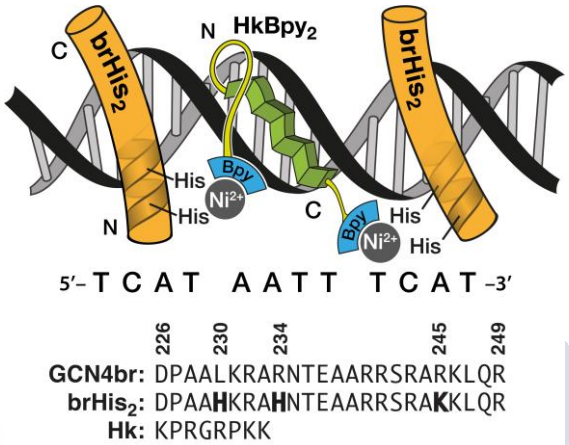


Figure 4. *Top:* Schematic structure of the supramolecular assembly. *Bottom*: Sequence of the natural GCN4 basic region, as well as the **brHis$_2$**, and AT-hook peptide.

We studied the DNA binding properties of the system by EMSA. As expected, incubation of the peptide **brHis$_2$** with the target dsDNA, did not induce the formation of retarded bands in the gel (Figure 5a, lane 2). Addition of $Ni(ClO_4)_2$ did not generate new retarded bands (Figure 5a, lane 3), but subsequent addition of **HkBpy$_2$** gave rise to a new retarted band that is consistent with the formation of the desired complex (Figure 5a, lane 4). Gratifyingly, addition of EDTA to the mixture of the complex, led to the disappearance of the slowed band in the gel, which agrees with the expected disassembly of the complex (Figure 5a, lane 5). Importantly, control experiments with AT-Hook derivatives featuring a single N-terminal (**Bpy^{N}Hk**) or C-terminal bipyridines (**HkBpy^C**) did not give rise to slower-migrating bands under the same conditions in the presence of the target dsDNA (Figure 5b).

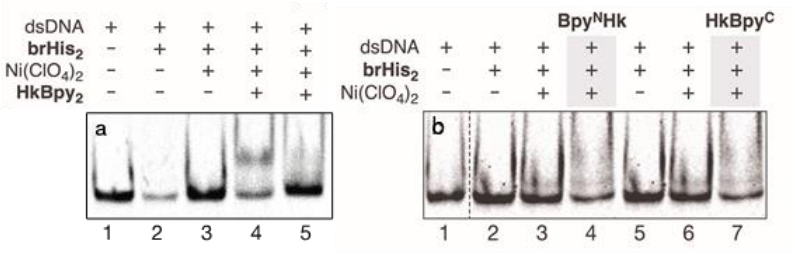


Figure 5. a) DNA binding studies for [(**brHis₂**)₂(**HkBpy₂**)Ni(II)₂] with its target dsDNA. Lane 5: Same mixture as in 4 after addition of EDTA. b) EMSA of the N-terminal (**Bpy^{N}Hk**) and C-terminal (**HkBpy^{C}**) mono-bipyridine AT-Hook derivatives with the target dsDNA.

We performed fluorescence anisotropy titrations by adding **HkBpy₂** to a solution containing a TMR-target dsDNA, the **brHis₂** peptide and $Ni(ClO_4)_2$. As shown in Figure 6, an increase in the anisotropy was detected in consonance with the formation of a larger complex (K_D~22.3 ± 0.4 nM).

As expected for a poorly structured peptide, the circular dichroism spectrum of **brHis₂** presented a relatively weak negative signal at 222 nm, even in the presence of **HkBpy₂** and consensus DNA; the addition of $Ni(ClO_4)_2$ to the mixture promoted a considerable increase in the negative ellipticity intensity at 222 nm, which was consistent with the folding of the peptide chain into an α-helix. In accordance with the results obtained by EMSA, the addition of EDTA to the supramolecular complex promoted a drastic decrease in the helicity of the peptide (Figure 6, right), which correlated with the disruption of the DNA complex.

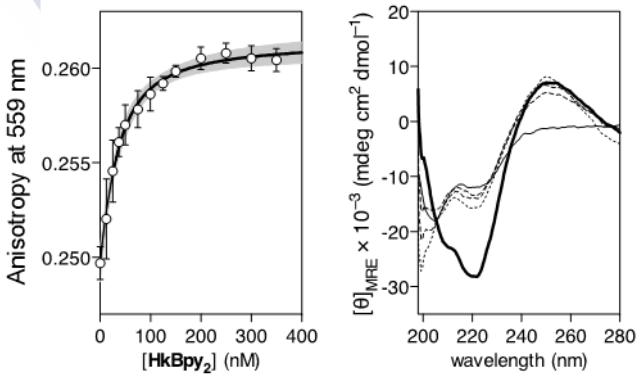


Figure 6. *Left*: Anisotropy titration of the TMR-target dsDNA in the presence of **brHis₂** and $Ni(ClO_4)_2$ with increasing concentrations of **HkBpy₂**. *Right*: Circular dichroism of **brHis₂** (solid line), after addition of target dsDNA and **HkBpy₂** (dashed lines), and after addition of $Ni(ClO_4)_2$ (thick solid line); addition of EDTA (dotted line).

Chapter I. Section 3:

Metal-dependent DNA recognition and cell internalization of designed, basic peptides.

Previously, it has been demonstrated that the complete leucine zipper of the GCN4 factor could be replaced by artificial dimerizing elements without compromising the DNA recognition properties of the system. In contrast, monomeric GCN4 peptides display low DNA-binding affinity. We have designed a monomeric bis-histidine peptide (**brHis$_2$**) based on the basic region of GCN4, which could bind to its DNA sites when treated with $PdCl_2$(en).

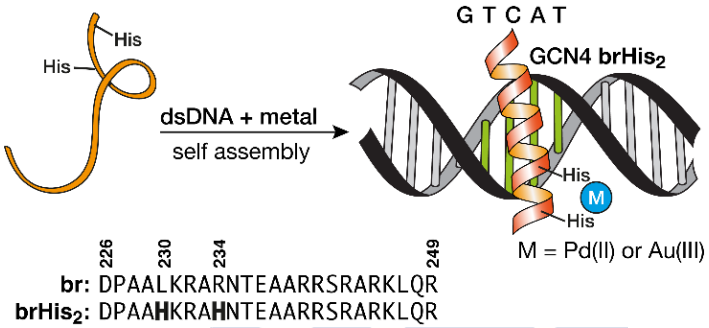


Figure 7. *Top*: Schematic illustration of the system. *Bottom*: Sequences of the natural GCN4 basic region and the mutated peptide **brHis$_2$**.

EMSA studies showed that the incubation of the peptide **brHis$_2$** with its consensus dsDNA, did not induce the formation of retarded bands (Figure 8a, lane 2). However, addition of $PdCl_2$(en) to this mixture gave rise to a new retarded band corresponding to the formation of the desired peptide/DNA complex (Figure 8a, lane 3). Furthermore, the DNA complex could be reversibly dismounted and assembled by controlling the relative amounts of the Pd complex and metal chelator, DEDTC (Figure 8a, lane 4-7). Remarkably, the metal-promoted interaction was highly selective; therefore, mutation in a single position of the target site was enough to completely abolish the DNA binding (Figure 8b).

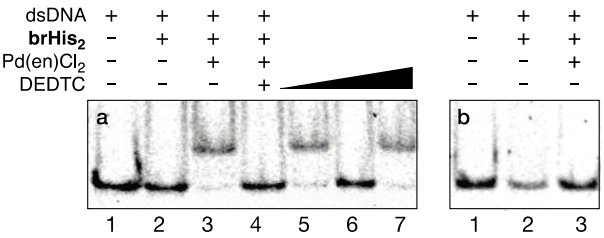


Figure 8. a) Incubation with the target dsDNA and switching. Lane 5: addition of $PdCl_2$(en); Lane 6: addition of DEDTC; Lane 7: addition of $PdCl_2$(en). b) Incubation with mutated dsDNA.

After that, we performed fluorescence anisotropy titrations by adding **brHis$_2$** to a solution containing $PdCl_2$(en) and a (TMR)-labelled dsDNA containing the target sequence. The addition of **brHis$_2$** led to a progressive increase in the fluorescence

anisotropy, calculating an apparent K_D~24 ± 17 nM (Figure 9, left). Circular dichroism of **brHis$_2$** presented a relatively weak negative signal at 222 nm, even in the presence of the consensus DNA. Addition of $PdCl_2$(en) to the mixture promoted the folding of the peptide into an α-helix. Addition of DEDTC to the mixture resulted in a decrease in the helicity of the peptide, correlating with the disruption of the DNA complex (Figure 9, right).

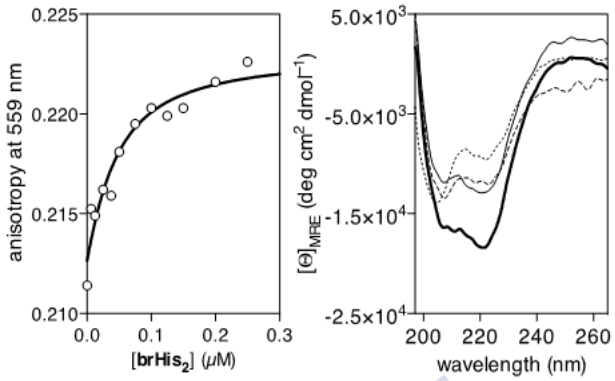


Figure 9. *Left*: Anisotropy titration of TMR-target dsDNA in the presence of $PdCl_2$(en) and increasing concentrations of **brHis$_2$**. *Right*: CD of **brHis$_2$** (dotted line), after addition of target dsDNA (dashed line) and $PdCl_2$(en) (thick solid line), and subsequent addition of DEDTC (solid line).

After corroborating that the metal complex was able to coordinate to histidines of the peptide, even in absence of DNA, we were curious to know if this palladium clipping could influence the cell penetration ability of **brHis$_2$**. Incubation of mammalian Hela cells with TMR-**brHis$_2$**, revealed that despite the basic character of the peptide, it showed a poor internalization. Gratifyingly, addition of 1 eq. of $PdCl_2$(en) led to the appearance of bright intracellular emission, which was mainly localized in endosomes (Figure 10). These results represented the first examples of a metal-triggered cell internalization of designed basic peptides.

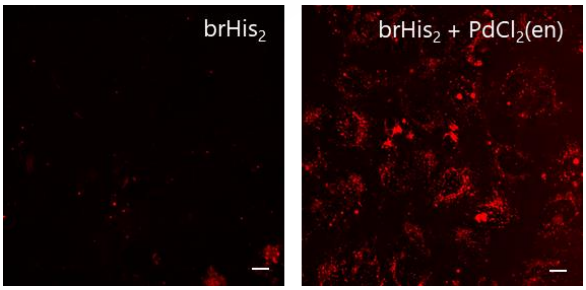


Figure 10. Fluorescence micrographies of Hela cells after incubation with TMR-**brHis$_2$** for 30 min at 37 ºC (*left*), and the same experiment in the presence of 1 eq. of $PdCl_2$(en) (*right*).

Chapter II:

Controlling oncogenic KRAS signaling pathways in living cells with a coordination staple.

RAS oncoproteins are molecular switches associated with critical signaling pathways. Mutations in the *RAS* gene family, mainly the *KRAS* isoform, are responsible of some of the deadliest cancers. It has been shown that stabilized α-helical peptides based on the SOS1 αH helix bind KRAS with high affinity and inhibit RAS activation. Following our interest in the application of metal chelation for modulating the properties of peptides, we have chosen for our designed the fragment comprising SOS1 residues Phe^{929} to Asn^{944} replacing two residues solvent-exposed face by His, to chelate a Pd(II) ion (Figure 11).

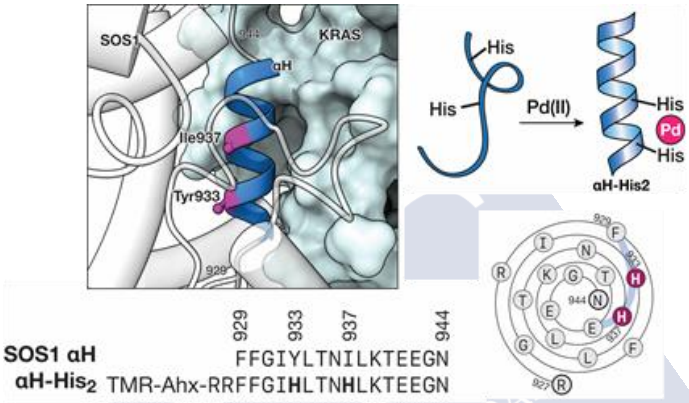


Figure 11. *Top left*: αH helix of SOS1 with RAS (PDB code 1BKD). *Bottom left*: Sequence of the natural SOS1 and the mutant **αH-His$_2$**. *Top right*: Helical stabilization by Pd(II) chelation. *Bottom right*: Wenxiang diagram of the **αH-His$_2$**.

Circular dichroism studies showed that metal chelation had a modest effect in the α-helical content of the **αH-His$_2$** (Figure 12a). Coordination of the Pd(II) ion was confirmed by MS, as well as by the reduction in the helicity upon addition of the Pd(II) chelator DEDTC (Figure 12b).

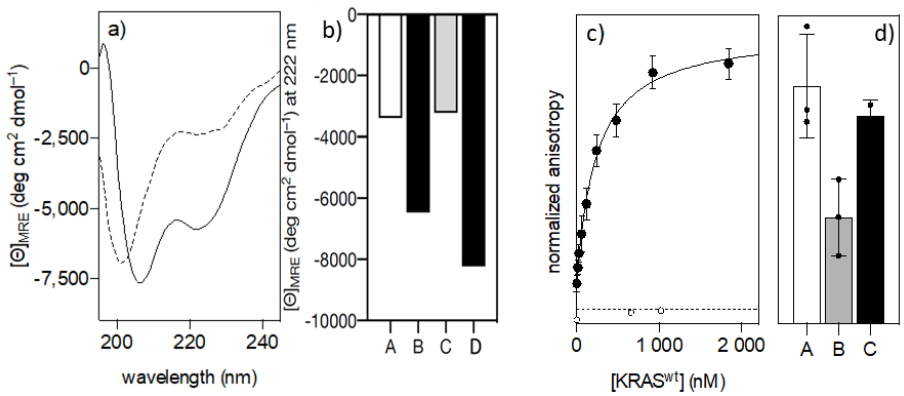


Figure 12. a) Circular dichroism of **αH-His$_2$** (dashed line) and the spectrum 30 min after addition of 1 eq. of $PdCl_2$(en). b) Circular dichroism values at 222nm for: A, **αH-His$_2$**; B, **αH-His$_2$**[Pd]; C, + DEDTC; D, + $PdCl_2$(en). c) Anisotropy of **αH-His$_2$**[Pd] with increasing concentrations of KRAS^{wt}. d) Anisotropy values for: A, saturated **αH-His$_2$**[Pd]/KRAS; B, + DEDTC; C, + $PdCl_2$(en).

Anisotropy assays with **αH-His$_2$** did not show any increase in the florescence, while with the metallopeptide we observed a marked increase in the anisotropy (K_D~240nm, Figure 12c). Importantly, addition of DEDTC to the solution with the complex resulted in a decrease of the anisotropy, consistent with the disassembly of the system (Figure 12d).

After this, we studied the effect in the nucleotide release process required for KRAS^{wt} activation. Using a fluorescent mantGTP, we found that the binding of the metallopeptide **αH-His$_2$**[Pd] to KRAS^{wt} induced the displacement of the bound mantGTP nucleotide from the preformed complex, affecting the GTPas activity of the protein.

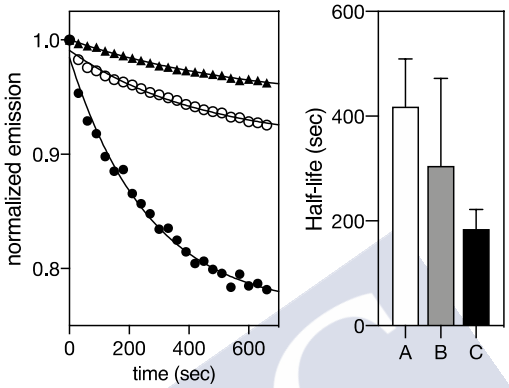


Figure 13. *Left:* Black triangles, intrinsic nucleotide displacement of the preformed KRAS^{wt}/mantGTP; White circles, same solution in the presence of **αH-His$_2$**; Black circles, same solution in the presence of **αH-His$_2$**[Pd]. *Right*: half-life of kinetics shown on the left. A, intrinsic KRAS^{wt}/mantGTP complex; B, with **αH-His$_2$**; C, with **αH-His$_2$**[Pd].

Incubation of A549 cells with **αH-His$_2$** revealed a rather poor internalization (Figure 14, left a). Gratifyingly, **αH-His$_2$**[Pd] showed an intense intracellular fluorescence signal (Figure 14, left b), validating the suitability of metallopeptide for cell-based assays. The activation of KRAS leads to stimulation of various signal transduction pathways that could be affected for the inhibitory effect of **αH-His$_2$**[Pd] on KRAS. Thus, we studied the activation of the MAPK RAF-MEK-ERK cascade, and the results (Figure 14, right) demonstrated that only **αH-His$_2$**[Pd] inhibited ERK phosphorylation dose-responsively.

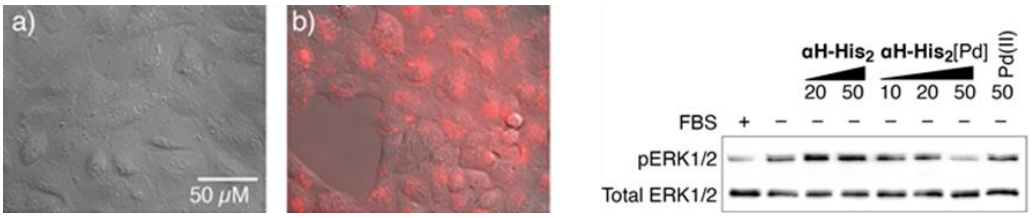


Figure 14. *Left*: Internalization in A549 cells. a) Incubation with **αH-His$_2$**; b) incubation with the **αH-His$_2$**[Pd] complex (30 min at 37 ºC). *Right*: Serum-starved A549 cells were incubated with **αH-His$_2$**, the metallopeptide **αH-His$_2$**[Pd], or with $PdCl_2$(en) for 4 h. Cellular lysates were subjected to Western blot analysis by using antibodies to phospho- and total ERK1/2.

Chapter III:

Intracellular reactions promoted by bis-histidine miniproteins stapled with Pd(II) complexes.

The generation of catalytically active metalloproteins inside living mammalian cells is a major research challenge at the interface between catalysis and cell biology. The peptide **brHis$_2$** reacts with palladium(II) sources to generate stapled pallado-miniproteins, capable of cross the cell membrane and that could promote designed reactions inside cells.

Knowing this, we tested the reactivity of these metallopeptides *in vitro,* exploring the depropargylation of the probe HBTPQ (**1**) (Figure 15, left). The results demonstrated that metallopeptides made from Pd(II) reagents with suitable labile ligands as $PdCl_2$(COD), were be able to perform the reaction with good yields. In addition, the control reagents suggested a positive effect of the peptide scaffold in the reactivity (Figure 15, right).

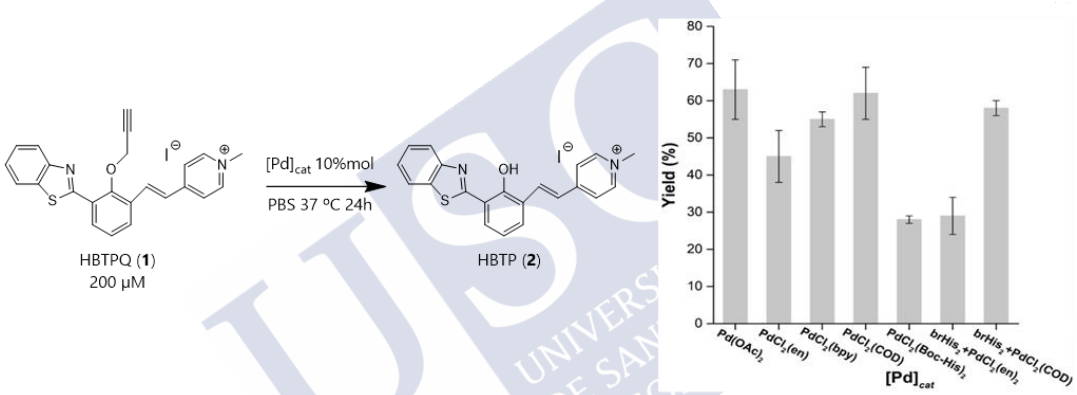


Figure 15. *Left*: Uncaging of HBTPQ (**1**). *Right*: Bar diagram representation of the yields obtained for each catalyst, calculated by RP-HPLC-MS.

The cellular reactions were performed incubating HeLa cells with probe **1** and with the peptide-palladium reagents. Gratifyingly, when cells were treated with palladopeptide resulting from TMR-**brHis$_2$** and $PdCl_2$(COD), we could observe an intense intracellular fluorescence arising from the expected product **2** (Figure 16a). However, with the metallopeptides made from $PdCl_2$(en) and $PdCl_2$(bpy), we did not observe fluorescence (Figure 16b-c). Moreover, in the absence of the peptide, the complex $PdCl_2$(COD) by itself was incapable of generating any intracellular fluorescence (Figure 16d).

ICP-MS analysis of cellular extracts revealed that the palladium content was substantially higher in cells treated with the preformed palladoprotein hybrid than with equivalent amounts of $PdCl_2$(COD). Remarkably, cell viability tests demonstrated that the none of palladopeptide hybrids or palladium complexes were toxic.

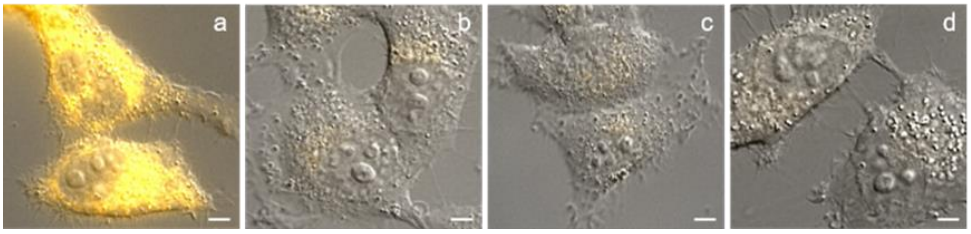


Figure 16. Fluorescence micrographies of HeLa cells. Cells were treated with probe **1** for 30 min at 37 ºC, washed and incubated with palladoproteins for 1h. a) palladopeptide resulting from **brHis$_2$** and $PdCl_2$(COD); b) palladopeptide obtained from $PdCl_2$(en); c) palladopeptide obtained from $PdCl_2$(bpy); d) $PdCl_2$(COD), in absence of peptide.

Considering the well-known integrin targeting motif RGD (Arg-Gly-Asp), we wondered whether it could be also converted into a reactive metallopeptide by adding two histidines at the edges (Figure 17, top). The experiments showed that this new complex was also capable of intracellular depropargylation of substrate **1**. Given that the cellular uptake of the palladium chelated complex of peptide **HRGDH**, might be partially associated to the presence of integrin receptors, we anticipated that its cell penetration would depend on the type of cells and the presence of these receptors. Indeed, we were glad to observe that while in A549 or HeLa cells there was intracellular catalysis (Figure 17, bottom a), in the MCF-7 breast cancer cell line, which expressed integrin at low levels, we barely detected any fluorescence (Figure 17, bottom b). This preliminary result could impulse the development of selective intracellular metal catalysis.

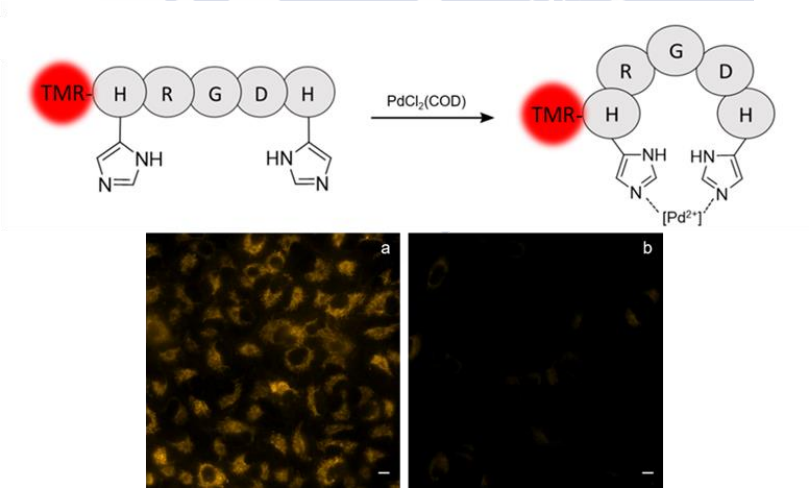


Figure 17. *Top***:** Bis-Histidine RGD derivative and stapling produced by treatment with the palladium complex. *Bottom*: Fluorescence micrographies of A549 (a) and MCF-7 cells (b), incubating probe HBPTQ **1** and the peptide-palladium complex resulting from TMR-**HRDGH** and $PdCl_2$(COD).

Chapter IV:
Assembly of artificial metalloproteins in live mammalian cells.

Based on the HALO-Tag technology, we have proposed the generation of catalytically active artificial metalloproteins inside living cells. With this aim, a variety of ruthenium-containing structures were synthesized, that should work as HALO-tag ligands and were additionally equipped with different fluorophores (Figure 19, top). In the other hand, we designed expression vectors to mammalian cells and also for bacteria, containing the HALO protein unit and an organelle-targeting sequence.

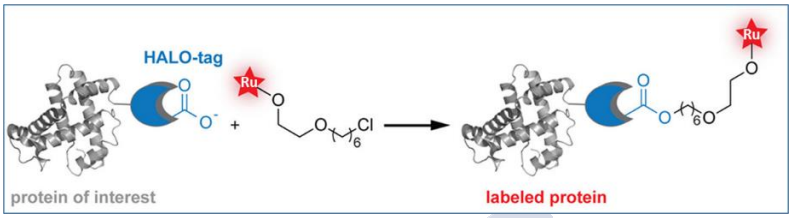


Figure 18. Schematic illustration of the assembly of the artificial metalloprotein.

We initially investigated the formation of the metalloprotein *in vitro,* using purified HALO proteins (HALO protein and the HALO-mutated(D106A), as negative control) and the chloroalkane ligands. We were glad to detect fluorescence only in the bands corresponding to reaction crudes resulting from the mixing of the ligands with the HALO protein, but the same experiments with the protein mutated did not provide any fluorescent band (Figure 19, bottom). For detected the presence of ruthenium in these band, we carried out ICP/MS analysis, detecting the ruthenium in the reaction with HALO-**RU1** (10% molar ratio with respect to total protein), while HALO-mutated(D106A) protein with same ligand, lacked the ruthenium.

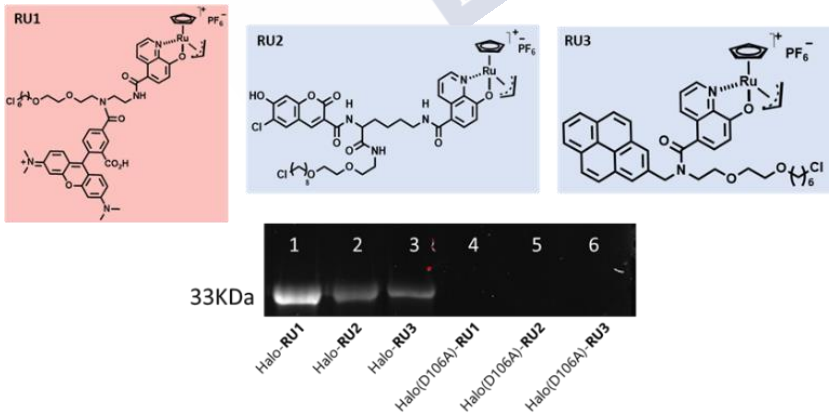


Figure 19. *Top*: Structures of ruthenium-containing chloroalkane ligands. *Bottom*: Polyacrylamide gel: Lanes 1-3: reactions made with HALO protein. Lanes 4-6: reactions made with HALO-mutated(D106A) protein. The HALO-RU species were obtained mixing the proteins with the ligands (ratio 1:2) during 1h in PBS at 37 ºC.

The designed reaction using Rho(alloc)$_2$ as probe (Figure 20, top), showed either **RU1** or the product resulting from its conjugation to the protein we were able to form the desired product. However, using the crude product mixture resulting from mixing the

HALO protein with **RU3** we did not observe fluorescence (Figure 20, bottom). Moreover, if the HALO ligand used for the coupling lacks the chloride-chain we did not observe reactivity (HALO-**RUcontrol**).

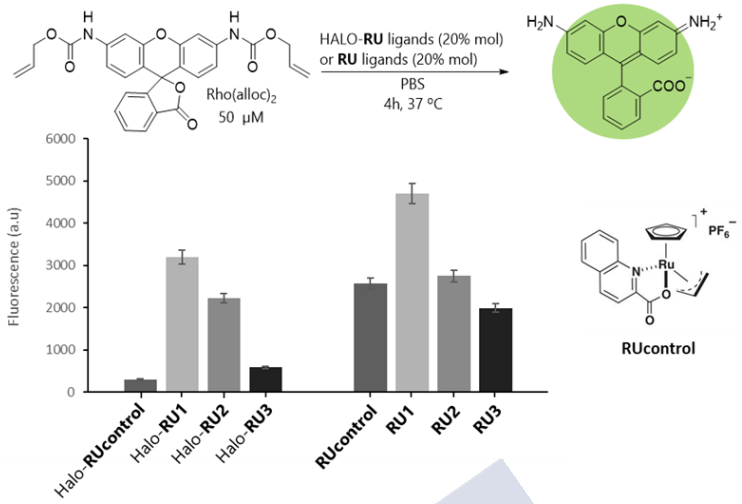


Figure 20. *Top:* Uncaging of Rho(alloc)$_2$. *Bottom left:* Bar diagram representation of the fluorescence obtained after the reaction. *Bottom right*: Structure of **RUcontrol**.

The experiments made inside HeLa mammalian cells, to corroborate the formation of the metalloprotein *in vivo*, demonstrated the assembly in the desired intracellular compartment when **RU1** was used as ligand. We could not detect any fluorescence when HALO-mutated(D106A) was used as binding protein.

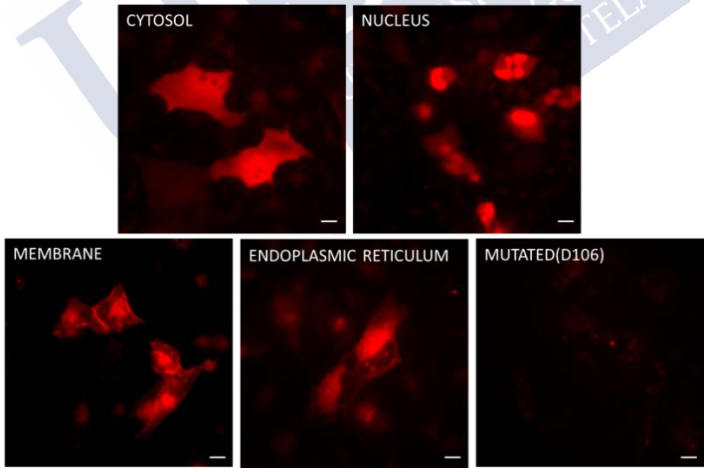


Figure 21. Fluorescence micrographies of HeLa cells. First, plasmids were transfected (48h) and then **RU1** was incubated for 1h at 37 ºC.

We assayed the same reaction, using as substrate Rho(alloc)$_2$, inside mammalian cells. Unfortunately, after 4 h, we did not detect any fluorescence corresponding to the uncaged product in any organelle of the cells that were expected to contain the protein-ruthenium conjugates.

Anyhow, all these results are consistent with the formation of the desired ruthenaproteins inside cells; however, these are not able to work as catalysts for the desired deallylation reaction. Nevertheless, in qualitative terms, the results are promising.

Resumo

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Capítulo I. Sección 1:
Miniproteínas de unión a ADN baseadas en dedos de zinc.

A obtención de proteínas artificiais que imiten as propiedades de unión ao ADN dos factores de transcrición naturais abre novas formas de manipular a expresión xénica. Inspirándonos nesta modularidade, deseñamos miniproteínas sintéticas que combinan módulos de dedos de zinc do factor de transcrición GAGA e dominios peptídicos do AT-Hook

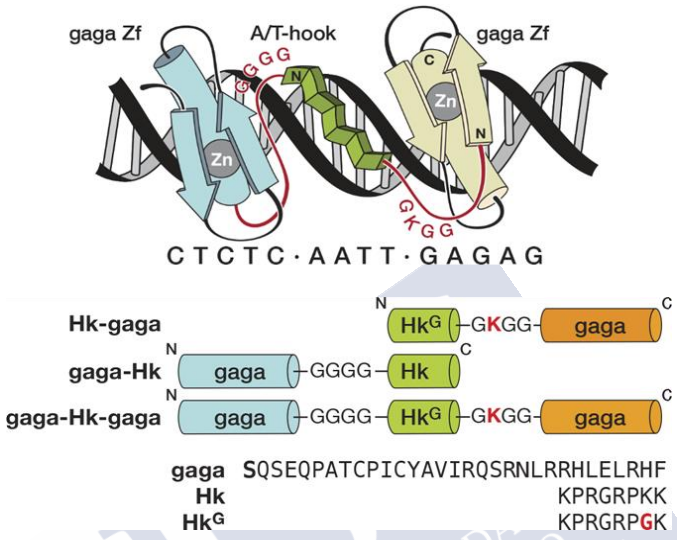


Figura 22. Representación dos híbridos sintetizados e as súas secuencias.

Tendo a man os conxugados bivalentes desexados, estudiamos as súas propiedades de unión ao ADN utilizando EMSA. Para isto, o oligonucleótido que contiña sitios de unión para o AT-Hook e GAGA mesturouse con concentracións crecentes de **gaga-Hk** ou **Hk-gaga**. Os xeles mostraron bandas de migración lenta dependentes da concentración en ambos casos, que foron consistentes coa formación dos complexos péptido/ADN desexados. As valoración de anisotropía de fluorescencia confirmaron que ámbolos péptidos unían con alta afinidade aos seus sitios diana.

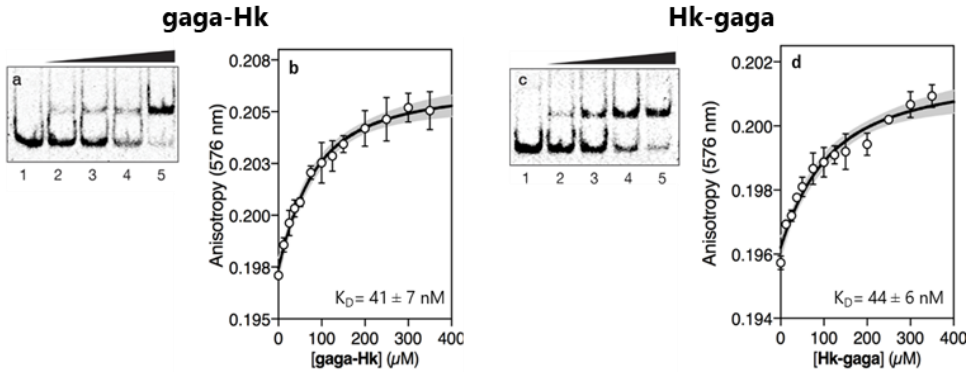


Figura 2. *a)-c)* Electroforesis en xel de poliacrilamida, utilizando o ADN consenso e concentracións crecentes de **gaga-Hk** (a) o **Hk-gaga** (c). *b)- d)* Anisotropía do TMR-ADN diana con concentracións crecentes de **gaga-Hk** (b) o **Hk-gaga** (d).

Tras isto, pasamos a estudiar a interacción ternaria (sucos "maior-menor-maior"). Afortunadamente, a adición de concentracións crecentes da quimera **gaga-Hk-gaga** ao seu oligonucleótido obxectivo, produciu unha banda de migración máis lenta (Figura 3a), consistente coa formación do complexo ternario miniproteína/ADN. É importante destacar que a miniproteína sintética non provocou bandas retardadas cando se incubou cunha secuencia que carecía de ámbolos sitios de unión para GAGA (Figura 3b, carrís 1-5), o do sitio rico en A/T (Figura 3b, carrís 6-10). As valoracións de anisotropía de fluorescencia utilizando un ADN diana marcado con TMR, confirmaron a unión de alta afinidade da quimera peptídica trivalente (Figura 3c).

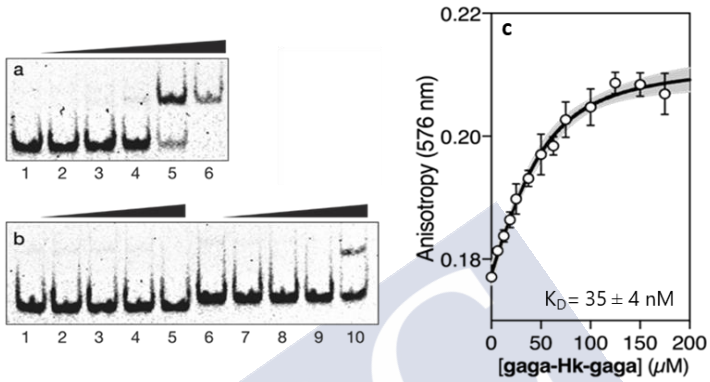


Figura 3. a) ADN obxectivo con concentracións crecentes de **gaga-Hk-gaga**. b) Carrís 1-5: ADN mutado nos sitios GAGA; Carrís 6-10: ADN mutado para AT-Hook, ámbolos dous con concentracións crecentes de **gaga-Hk-gaga**. c) Anisotropía do TMR-ADN diana con concentracións crecentes de **gaga-Hk-gaga.**

Ademais, analizamos a interacción da quimera trivalente co seu ADN diana e mutado utilizando nanoporos. Esta técnica funciona examinando estatisticamente os estados do ADN en presenza da cantidade determinada do aglutinante peptídico, deducindo así unha K_D. Os resultados correspondéronse cos obtidos mediante anisotropía.

A estrutura lineal e completamente peptídica dos nosos péptidos permitiu a súa expresión en células de mamífero. Despois de deseñar e clonar as nosas secuencia nos vectores apropiados, e realizar os ensaios con xenes reporteiros, os resultados demostraron que os nosos péptidos non naturais aumentaron os procesos de expresión xénica en células de mamíferos, que contiñan sitios deseñados no seu xenoma. Estes resultados abren a porta á xeración de novas ferramentas xenéticas baseadas neste tipo de versións artificiais e miniaturizadas de factores de transcrición, máis curtos que os dedos de zinc descritos ata o de agora.

Capítulo I. Sección 2:

Ensamblaxe de un complexo metalopéptido ternario en sitios específicos de ADN, mediado por un adaptador AT-Hook.

Exploramos a aplicación da coordinación de metais como un método poderoso para obter aglutinantes de ADN multiméricos. No noso deseño, equipamos un AT-Hook con dous ligandos de bipiridina (**HkBpy$_2$**) para recrutar ao péptido GCN4 modificado con dous residuos de His (**brHis$_2$**) nos sucos maiores de ADN adxacentes, en resposta a un clip metálico (níquel (II), Figura 4). O Ni (II) media a autoensamblaxe dos motivos peptídicos e tamén actúa como axente nucleante da hélice α.

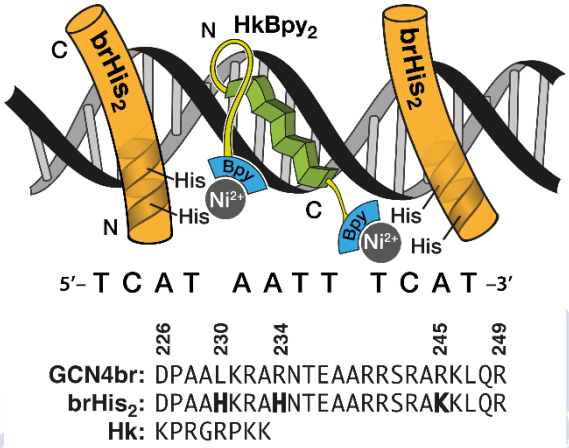


Figura 4. *Arriba*: Estrutura esquemática do conxunto supramolecular. *Abaixo*: Secuencia da rexión básica natural do GCN4 e da mutada **brHis$_2$**, e do péptido AT-Hook.

Estudamos as propiedades de unión ao ADN do sistema mediante EMSA. Como se esperaba, a incubación do péptido **brHis$_2$** co oligonucleótido diana non induciu a formación de bandas retardadas no xel (Figura 5a, carril 2). A adición de i$(ClO_4)_2$ non xerou novas bandas retardadas (Figura 5a, carril 3), pero a posterior adición de **HkBpy$_2$** deu lugar a unha nova banda que é consistente coa formación do complexo desexado (Figura 5a, carril 4). Afortunadamente, a adición de AEDT á mestura do complexo levou á desaparición da banda retardada no xel, o que concorda co desensamblaxe esperado do sistema (Figura 5a, carril 5). É importante destacar que os experimentos control con derivados do AT-Hook cunha soa bipiridina N-terminal (**Bpy^{N}Hk**) ou C-terminal (**HKBpy^C**) non deron lugar a bandas de migración máis lentas coas súas dianas de ADN (Figura 5b).

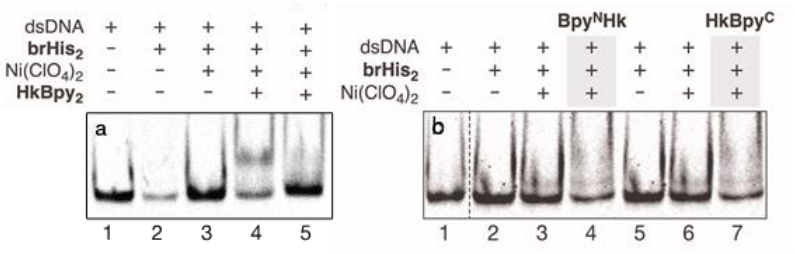


Figura 5. a) Estudos de unión do [(**brHis₂**)₂(**HkBpy₂**)Ni(II)₂] co seu ADN diana. Carril 5: Mesma mestura que en 4 despois da adición do AEDT. b) EMSA dos derivados N-terminal (**Bpy^{N}Hk**) e C-terminal (**HkBpy^{C}**) mono-bipiridina AT-Hook, coas súas dianas de ADN.

Tras isto, realizamos valoracións de anisotropía de fluorescencia engadindo **HkBpy₂** a unha solución que contiña o oligonucleótido diana marcado con TMR, o péptido **brHis₂** e $Ni(ClO_4)_2$. Como se mostra na Figura 6, detectouse un aumento na anisotropía en consonancia coa formación dun complexo máis grande (K_D~22,3 ± 0,4 nM).

Como era de esperar para un péptido mal estruturado, o espectro de dicroísmo circular de **brHis₂** presentou un sinal negativo relativamente débil a 222 nm, incluso en presenza de **HkBpy₂** e o seu ADN consenso; a adición de $Ni(ClO_4)_2$ á mestura promoveu un aumento considerable na intensidade de elipticidade negativa, o que é consistente co pregamento da cadea peptídica en hélice α. De acordo cos resultados obtidos por EMSA, a adición de AEDT ao complexo supramolecular promoveu unha diminución drástica da helicidade do péptido, que se pode correlacionar coa separación do complexo do ADN (Figura 6, dereita).

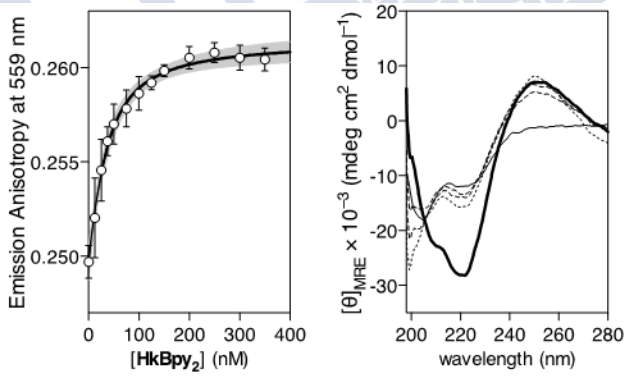


Figura 6. *Esquerda*: Anisotropía do TMR-ADN diana en presenza de **brHis₂** e $Ni(ClO_4)_2$ con concentracións crecentes de **HkBpy₂**. *Dereita*: Dicroísmo circular de **brHis₂** (liña continua) despois da adición do ADN diana e **HkBpy₂** (liñas descontinuas), despois da adición de $Ni(ClO_4)_2$ (liña continua grosa); e adición de EDTA (liña de puntos).

Capítulo I. Sección 3:

Recoñecemento de ADN dependente de metais e internalización celular de péptidos básicos deseñados.

É un feito demostrado que a cremalleira de leucina do factor GCN4 pódese substituír completamente por elementos dimerizantes artificiais, sen comprometer as propiedades de recoñecemento ao ADN do sistema. Polo contrario, os péptidos GCN4 monoméricos mostran unha baixa afinidade ao ADN. Móstrase neste capítulo o deseño dun péptido monomérico bis-histidina (**brHis$_2$**) baseado na rexión básica de GCN4, que podería unirse ao seu sitio do ADN cando é tratado con $PdCl_2$(en).

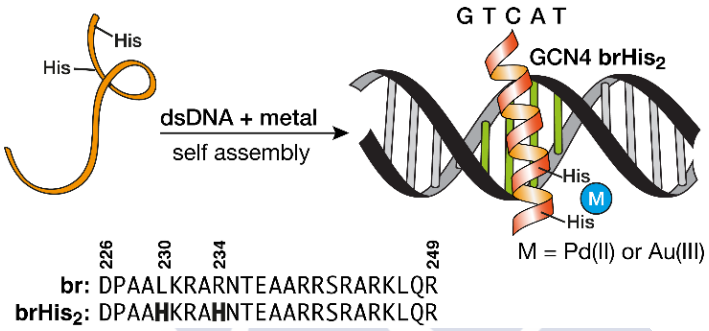


Figura 7. *Arriba*: Ilustración esquemática do sistema. *Abaixo*: Secuencias da rexión básica natural do GCN4 e o péptido mutado **brHis$_2$**.

Estudos EMSA mostraron que a incubación de **brHis$_2$** co seu ADN consenso non inducía a formación de bandas retardadas (Figura 8a, carril 2). Sen embargo, a adición de $PdCl_2$(en) a esta mestura deu lugar a unha nova banda retardada correspondente á formación do complexo péptido/ADN desexado (Figura 8a, carril 3). Ademais, este complexo púidose desmontar e ensamblar de forma reversible controlando as cantidades de Pd e quelante metálico, DEDTC (Figura 8a, carril 4-7). Sorprendentemente, a interacción for moi selectiva e a mutación dunha soa posición do sitio obxectivo aboliu completamente a unión ao ADN (Figura 8b).

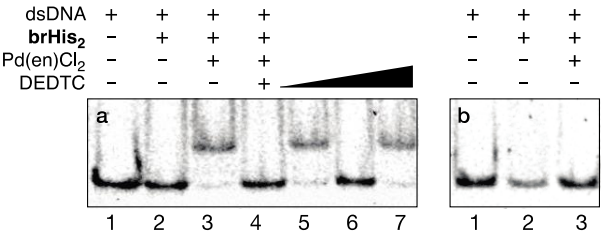


Figura 8. a) Incubación co ADN diana e reversibilidade. Carril 5: Adición de $PdCl_2$(en); Carril 6: adición de DEDTC; Carril 7: adición de $PdCl_2$(en); b) Incubación co ADN mutado.

Despois disto, realizamos valoracións de anisotropía de fluorescencia engadindo **brHis$_2$** a unha solución que contiña $PdCl_2$(en) e o ADN diana marcado (TMR). A adición de **brHis$_2$** conduciu a un aumento progresivo da anisotropía de fluorescencia, calculando unha Kd~24 ± 17 nM (Figura 9, esquerda). O dicroísmo circular de **brHis$_2$** presentou un sinal

negativo relativamente débil a 222 nm, incluso en presenza do ADN consenso. A adición de $PdCl_2$(en) á mestura promoveu o pregamento do péptico nunha hélice α. A adición de DEDTC á mestura deu coma resultado unha diminución na helicidade do péptido, que se correlacionou coa ruptura do complexo (Figura 9, dereita).

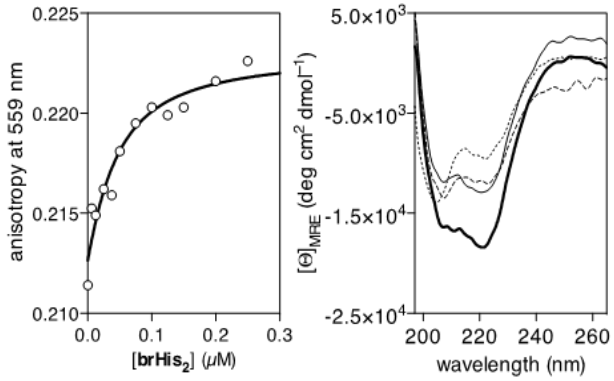


Figura 9. *Esquerda*: Anisotropía do TMR-ADN diana en presenza de $PdCl_2$(en) e concentracións crecentes de **brHis₂**. *Dereita*: CD de **brHis₂** (liña punteada), despois da adición do ADN consenso (liña descontinua) e do $PdCl_2$(en) (liña continua grosa); adición de DEDTC (liña continua).

Despois de corroborar que o complexo metálica era capaz de coordinar as histidinas do péptido, incluso en ausencia de ADN, tivemos curiosidade por saber se este paladio-clip podía influír na capacidade de penetración celular do **brHis₂**. A incubación de células de mamífero HeLa con TMR-**brHis₂**, revelou que, a pesar do carácter básico do péptido, mostraba unha mala internalización. Afortunadamente, a adición de $PdCl_2$(en) deu lugar á aparición dunha emisión intracelular brillante, que se localizou principalmente en endosomas (Figura 10). Estes resultados representaron os primeiros exemplos dunha internalización celular activada por metais de péptidos básicos deseñados.

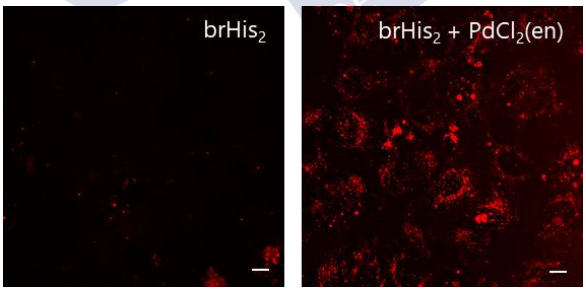


Figura 10. Micrografías de fluorescencia de células HeLa despois da incubación con TMR-**brHis₂** durante 30 min a 37 ºC (*esquerda*), e o mesmo experimento en presenza de 1 eq. de $PdCl_2$(en) (*dereita*).

Capítulo II:

Control da vías de sinalización oncoxénicas de KRAS en células vivas mediante grampa de coordinación.

As oncoproteínas *RAS* son interruptores moleculares asociados con vías de sinalización críticas. As mutacións nesta familia de xenes *RAS*, principalmente a isoforma KRAS, son responsables dos cancros máis mortais. Atópase demostrado que péptidos α-helicoidais estabilizados baseados na hélice SOS1 unen KRAS con alta afinidade e inhiben a súa activación. Para o noso deseño eliximos o fragmento comprendido entre os residuos de SOS1 Phe^{929} e Asn^{944}, substituíndo os residuos da cara exposta ao disolvente por His para quelar un ión Pd(II) (Figura 11).

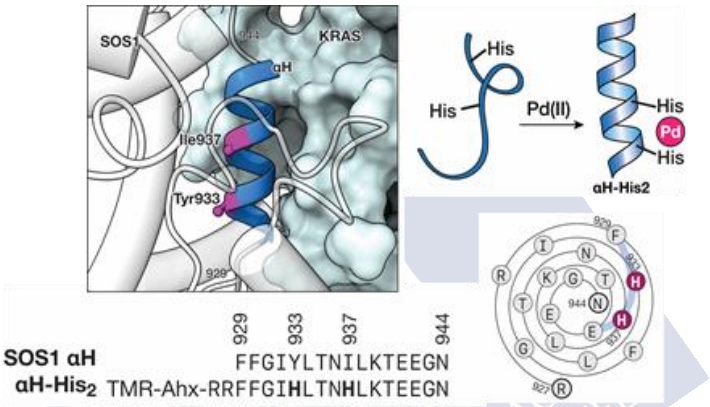


Figura 11. *Arriba Esquerda*: Hélice αH de SOS1 con RAS (código PDB 1BKD). *Abaixo esquerda:* Secuencia de SOS1 natural y del mutante **αH-His$_2$**. *Arriba dereita*: Estabilización helicoidal por quelación de Pd (II). *Abaixo dereita:* Diagrama de Wenxiang del **αH-His$_2$**.

Os estudos de dicroísmo circular mostraron que a quelación do metal tivo un efecto modesto no contido α helicoidal de **αH-His$_2$** (Figura 12a). A coordinación de Pd(II) foi confirmada por MS, así coma pola redución da helicidade trala adición do quelante de Pd(II) DEDTC (Figura12b).

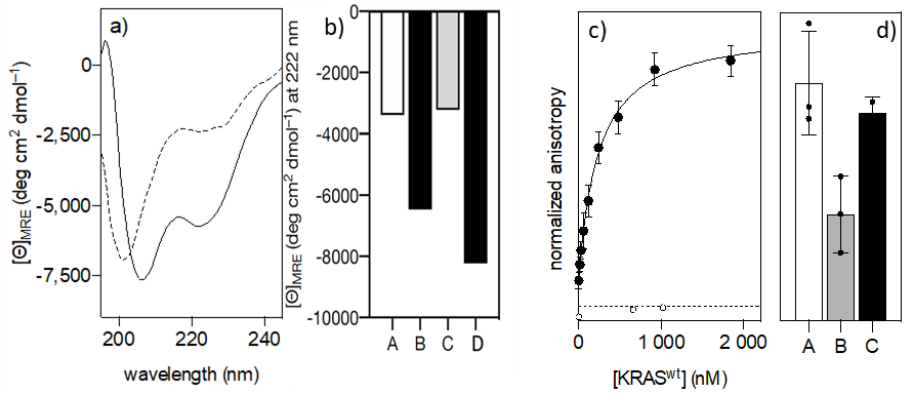


Figura 12. a) DC do **αH-His$_2$** (liña descontinua) e tras 30 min despois da adición de 1 eq. de $PdCl_2$(en). b) DC a 222 nm: A, **αH-His$_2$**; B, **αH-His$_2$**[Pd]; C, + DEDTC; D, + $PdCl_2$(en). c) Anisotropía de **αH-His$_2$**[Pd] con concentracións crecentes de KRAS^{wt}. d) Valores de anisotropía para: A, **αH-His$_2$**[Pd]/KRAS saturado; B, + DEDTC; C, + $PdCl_2$(en).

A anisotropía de **αH-His$_2$** non mostrou ningún aumento na fluorescencia, mentres que co metalopéptido observamos un marcado aumento na anisotropía (K_D~240 nm, Figura 12c). É importante destacar que a adición de DEDTC ao complexo resultou nunha diminución da fluorescencia consistente coa desensamblaxe do sistema (Figura 12d).

Tras isto, estudamos o efecto no proceso de liberación de nucleótidos requirido para a activación de KRAS^{wt}. Usando un mantGTP fluorescente, atopamos que a unión de **αH-His$_2$**[Pd] a KRAS^{wt} inducía o desprazamento do nucleótido mantGTP do complexo preformado, afectando á actividade GTPasa da proteína.

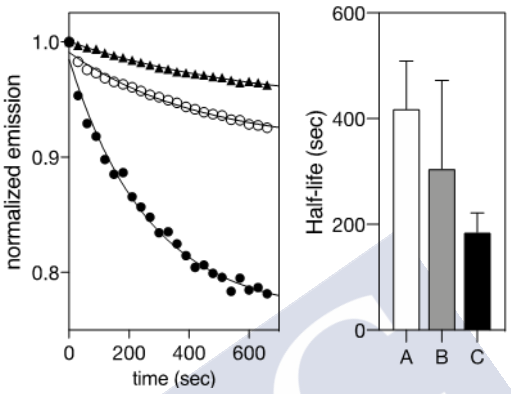


Figura 13. *Esquerda*: triángulos negros, desprazamento de nucleótidos intrínseco de KRAS^{wt}/mantGTP preformado; Círculos brancos, mesma solución en presenza de **αH-His$_2$**; Círculos negros, mesma solución en presenza de **αH-His$_2$**[Pd]. *Dereita*: Vida media das cinéticas da esquerda. A, complexo KRAS^{wt}/mantGTP; B: con **αH-His$_2$**; C: con **αH-His$_2$**[Pd].

A incubación de células A 549 con **αH-His$_2$** revelou unha internalización pobre. Afortunadamente, **αH-His$_2$**[Pd] mostrou un sinal intenso de fluorescencia intracelular (Figura 14, esquerda), validando a idoneidade do metalopéptido para ensaios celulares. A activación de KRAS conduce á estimulación de varias vías de transdución de sinais que poderían verse afectadas polo efecto inhibidor de **αH-His$_2$**[Pd] sobre KRAS. Polo tanto, estudamos a activación da cascada MAPK RAF-MEK-ERK e os resultados (Figura 14, dereita) demostraron que só **αH-His$_2$**[Pd] inhibía a fosforilación de ERK en resposta á dose.

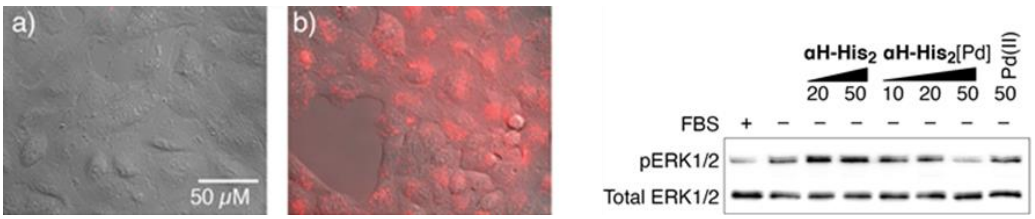


Figura 14. *Esquerda*: Internalización en A549. a) Incubación con **αH-His$_2$**; b) incubación co complexo **αH-His$_2$**[Pd] (30 min, 37 ºC). *Dereita*: Células A549 privadas de soro incubáronse con **αH-His$_2$**, con **αH-His$_2$**[Pd] ou con $PdCl_2$(en) durante 4 h. Os lisados celulares sometéronse a análise Western-Blot usando anticorpos para ERK1/2 fosfo- e total.

Capítulo III:

Reaccións intracelulares promovidas por miniproteínas bis-histidina grampadas mediante complexos de Pd (II).

A xeración de metaloproteínas cataliticamente activas dentro de células vivas é un importante desafío na interface entre a catálise e a bioloxía celular. O péptido **brHis$_2$** reacciona con paladio(II) para xerar paladio-miniproteínas grampadas, capaces de atravesar a membrana celular e que poderían promover reaccións deseñadas no interior celular.

Sabendo isto, probamos a reactividade destes metalopéptido *in vitro*, explorando a desproparxilación da sonda HBTPQ (**1**) (Figura 15, esquerda). Os resultados demostraron que os metalopéptidos preparados a partir de reactivos de Pd(II) con ligandos lábiles adecuados coma $PdCl_2$(COD) eran reactivos con bos rendementos. Ademais, os controis suxeriron un efecto positivo do andamio peptídico nesta reactividade (Figura 15, dereita).

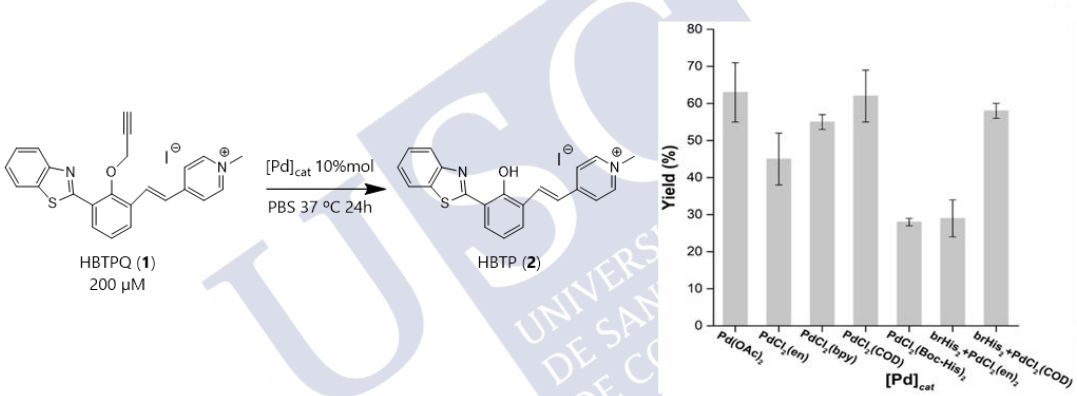


Figura 15. *Esquerda*: Desprotección de HBTPQ (**1**). *Dereita*: Diagrama de barras dos rendementos obtidos para cada catalizador, calculados mediante RP-HPLC-MS.

As reaccións celulares realizáronse incubando células HeLa coa sonda **1** e cos péptidos-paladio. Afortunadamente, cando as células se trataron co péptido-paladio resultante de mesturar TMR-**brHis$_2$** e $PdCl_2$(COD), puidemos observar unha intensa fluorescencia intracelular correspondente ao produto esperado **2** (Figura 16a). Sen embargo, cos metalopéptidos feitos co $PdCl_2$(en) y $PdCl_2$(bpy), non observamos fluorescencia (Figura 16b-c). Ademais, en ausencia do péptido, o complexo $PdCl_2$(COD) por si so foi incapaz de xerar o produto (Figura 16d).

Análises ICP-MS de extractos celulares revelaron que o contido de paladio era substancialmente maior nas células tratadas co híbrido péptido-paladio preformado que con cantidades equivalente de $PdCl_2$(COD). Ademais, as probas de viabilidade celular demostraron que ningún dos péptidos-paladio ou os complexos de paladio eran tóxicos.

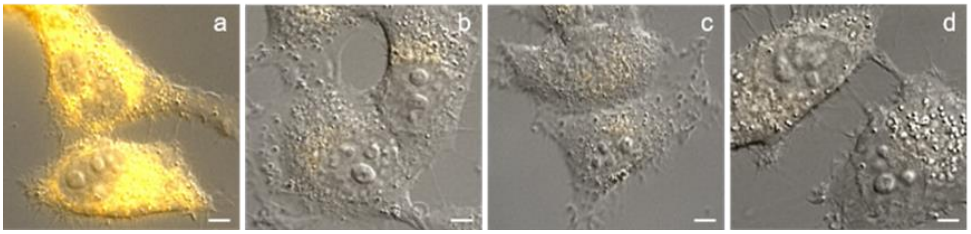


Figura 16. Micrografías de fluorescencia de HeLa. As células tratáronse coa sonda **1** durante 30 min a 37 ºC, laváronse e incubáronse coas proteínas-paladio durante 1h. a) Péptido-paladio **brHis$_2$** e $PdCl_2$(COD); b) Péptido-paladio resultante de $PdCl_2$(en); c) Péptido-paladio resultante de $PdCl_2$(bpy); d) $PdCl_2$(COD), sen péptido.

Tendo en conta o coñecido motivo de recoñecemento de integrinas RGD (Arg-Gly-Asp), preguntámonos se tamén podería converterse nun metalopéptido reactivo engadindo dúas histidinas nos extremos (Figura 17, arriba). Os ensaios mostraron que este novo complexo tamén era capaz de realizar a desproparxilación intracelular do substrato **1**. Dado que a internalización celular do metalopéptido **HRGDH**-Pd podería estar parcialmente asociada á presenza de receptores de integrina, anticipamos que a súa penetración dependería do tipo de células e da presenza destes receptores. De feito, alegrounos observar que mentres que en células A549 ou HeLa había catálise intracelular (Figura 17, inferior a), na liña celular MCF-7, que expresaba integrinas en niveis baixos apenas detectamos fluorescencia (Figura 17, inferior b). Estes resultados preliminares poderían impulsar o desenvolvemento de catálises metálicas intracelulares e selectivas.

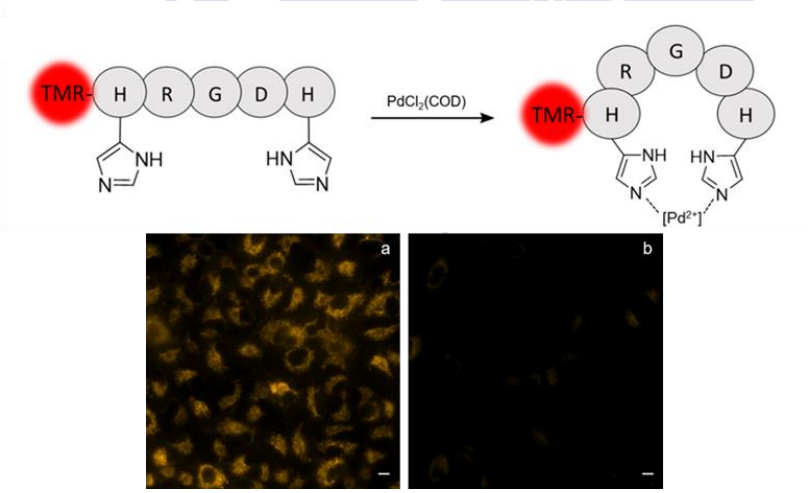


Figura 17. *Arriba*: Derivado RGD bis-histidina y grampado producido polo tratamento co complexo de paladio. *Abaixo*: Micrografías de fluorescencia de células A549 (a) e MCF-7 (b), incubando con **1** e o complexo TMR-**HRDGH** e $PdCl_2$(COD).

Capítulo IV:

Ensamblaxe de metaloproteínas artificiais en células vivas de mamíferos.

Baseándonos na tecnoloxía HALO-Tag, propuxemos a xeración de metaloproteínas artificiais cataliticamente activas dentro de células vivas. Para isto, sintetizáronse varios ligandos de HALO que contiñan rutenio, equipados con diferentes fluoróforos (Figura 19, arriba). Por outro lado, deseñáronse vectores de expresión para células de mamíferos que contiñan a proteína HALO e unha secuencia de dirección a orgánulos.

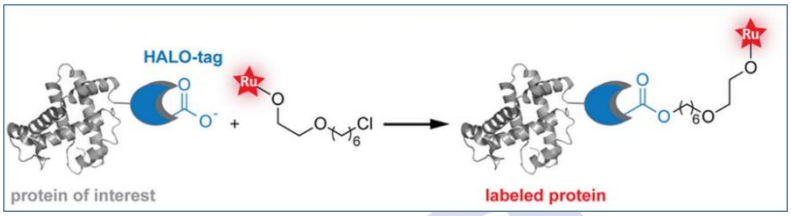


Figura 18. Ilustración esquemática da ensamblaxe da metaloproteína artificial.

Inicialmente investigamos a formación da metaloproteína *in vitro*, utilizando proteínas HALO purificadas (HALO y HALO-mutada(D106A), como control negativo) e os ligandos de cloroalcano. Detectouse fluorescencia só nas bandas correspondentes a reaccións cos ligandos e proteína HALO, pero os mesmos experimentos coa proteína mutada non proporcionaron ningunha banda fluorescente (Figura 19, abaixo). Para detectar a presenza de rutenio nestas bandas realizouse ICP/MS, detectando o rutenio na reacción con HALO-**RU1** (relación molar 10% con respecto ao total de proteína), mentres que coa proteína HALO-mutada(D106A) e o mesmo ligando, non foi detectado.

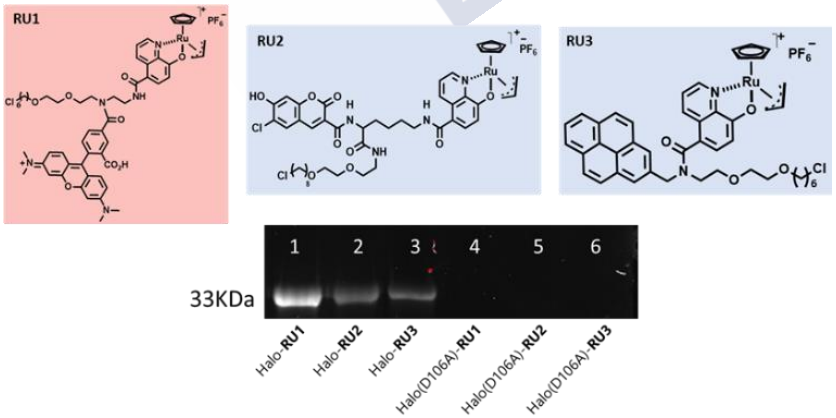


Figura 19. *Arriba*: Ligandos de cloroalcano con rutenio. *Abaixo*: Xel de poliacrilamida: Carrís 1-3, reaccións realizadas coa proteína HALO. Carrís 4-6, reaccións realizadas coa proteína HALO-mutada(D106A). As especies HALO-**RU** obtivéronse mesturando as proteínas cos ligandos (relación 1: 2) durante 1h en PBS a 37 ºC.

A reacción usando **Rho(alloc)$_2$** como sonda (Figura 20, arriba) mostrou que **RU1** ou o produto resultante da súa conxugación con HALO, podían formar o produto desexado. Sen embargo, con HALO-**RU3** non observamos fluorescencia (Figura 20, abaixo).

Ademais, se o ligando HALO usado para o acoplamento carece da cadea de cloruro, non observamos reactividade (HALO-**RUcontrol**).

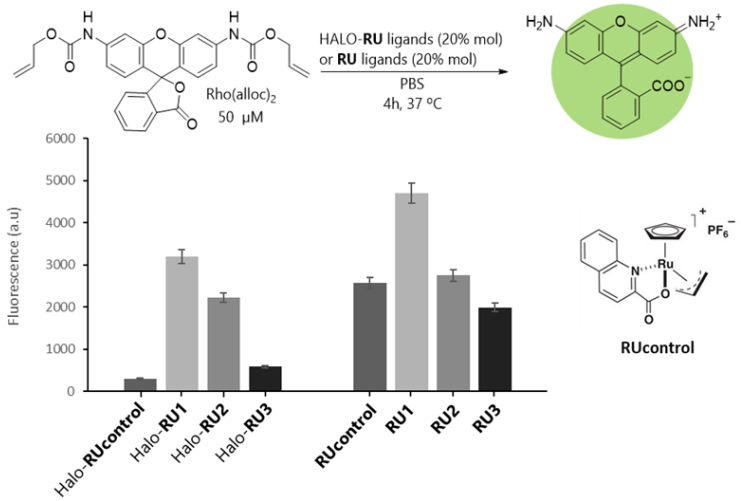


Figura 20. *Arriba*: Desprotección de Rho(alloc)$_2$. *Abaixo esquerda:* Diagrama de barras de la fluorescencia obtida despois de la reacción. *Abaixo dereita*: Estrutura de **RUcontrol**.

Os experimentos realizados no interior de células HeLa para corroborar a formación da metaloproteína demostraron a ensamblaxe no orgánulo intracelular específico cando se utilizou **RU1** como ligando. Non puidemos detectar ningunha fluorescencia cando se utilizou HALO-mutada(D106A) como proteína de unión.

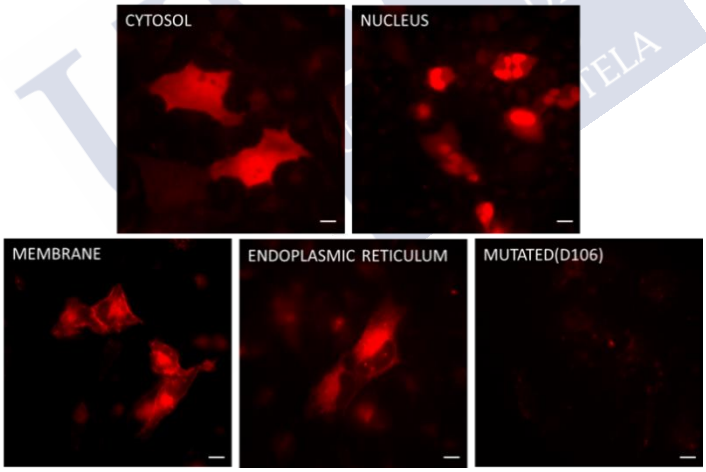


Figura 21. Micrografías de fluorescencia en HeLa. Primeiro se transfectaron os plásmidos (48 h) e logo incubouse **RU1** durante 1 h a 37 ºC.

A reacción de desprotección da sonda Rho(alloc)$_2$ probouse en células HeLa. Desafortunadamente, despois de 4 h, non se mostrou ningunha fluorescencia correspondente ao produto en ningún orgánulo onde detectamos os conxugados proteína-ligando.

Así, os resultados son consistentes coa formación das rutenioproteínas desexadas dentro das células, aínda que non funcionaron como catalizadores para a reacción de desalilación deseñada. Non obstante, en termos cualitativos, os resultados son prometedores.

Experimental Data

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

General information

All peptide synthesis reagents and amino acid derivatives were purchased from Sigma Aldrich and Iris Biotech; amino acids were purchased as protected Fmoc amino acids with the standard side chain protecting scheme: Fmoc-Ala-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Ser(*t*-Bu)-OH, Fmoc-Glu(O*t*-Bu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Ile-OH, Fmoc-Thr(*t*-Bu)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-His(Trt)-OH, and Fmoc-Asp(O*t*-Bu)-OH, except for the orthogonally protected Fmoc-Lys(Alloc)-OH, which was purchased from *Bachem*. The resins employed for the SPPS were H-Rink-Amide ChemMatrix (0.57 mmol/g loaging) from *Biotage AB,* or PAL-PEG-PS resin (0.19 mmol/g loading) from *Applied Biosystems*, depending on the case.

Chemical were purchased from *Sigma Aldrich*, *Alfa Aesar*, and *Strem* and used without further purification. Dry solvents were directly purchased from *Sigma Aldrich* and used without further purification.

The abbreviation "r.t." refers to reactions carried out a temperature between at 20-25°C.

Procedures and experimental techniques

Solid phase peptide synthesis

Peptides were synthesized following standard Fmoc-peptide synthesis protocols on a 0.1 mmol scale using a 0.5 mmol/g loading *H-Rink* amide *ChemMatrix* resin with a *Liberty Lite* automatic microwave assisted using peptide synthesizer from CEM Corporation. The amino acids were coupled in 5-fold excess DIC as activator, Oxime as base and DMF as solvent. Couplings were conducted for 4 min at 90 °C. Deprotection of the temporal Fmoc protecting group was performed by treating the resin with 20% piperidine in DMF for 1 min at 75 °C.

Deprotection of the alloc group was carried out following the methodology proposed by Sainlos et al.[241] Thus, a suspension of the resin in DCM was degassed by gently bubbling of N_2 for 5 minutes. After this time $Pd(PPh_3)$ (0.8 eq.) and PhSiH (25 eq.) were added and the flux of N_2 was maintained for 15 min. After this time, the resin was washed with DCM three times and the previous process repeated twice. Finally, the resin was filtered and washed with DMF, DEDTC and DMF again.

Coupling of bipyridine ligand: The resin containing the peptide (50 mg, aprox. 0.01 mmol) was suspended in DMF and shaken for 1 h to ensure a good swelling. A solution of the bipyridine ligand (7 mg, 0.02 mmol, 2 eq.), HATU (9 mg, 0.02 mmol, 2 eq.) and DIEA (250 µL, 0.195 M in DMF, 0.04 mmol, 4 eq.) was added. The reaction mixture was shaken for 1 h. The resin was washed with DMF (3 ×, 5 min) and DCM (2 ×, 5 min).

[241] M. Sainlos, B. Imperiali, *Nat. Protoc.* **2007**, *2*, 3201-3209.

Coupling of TMR and fluorescein: Fmoc-6-aminohexanoic acid (Fmoc-Ahx-OH) was added as a spacer between the peptide amino terminal side, and the fluorophore, and, after standard removal of the Fmoc protecting group, the corresponding fluorophore was added: TMR (5(6)-carboxytetramethylrhodamine) purchased from *Carbosynth,* was coupled using 3 eq. e (0.15 mmol, 64.5 mg), 3 eq. of HATU and 5 eq. of DIEA 0.2 M in DMF for 60 min; Fluorescein-5-Isothiocyanate was coupled using 4 eq., and 5 eq. of DIEA 0.2 M in DMF for 120 min.

Cleavage/deprotection step was performed by treatment of the resin-bound peptide for 2h with the following cleavage cocktail: 900 µL TFA, 50 µL CH_2Cl_2, 25 µL H_2O and 25 µL TIS (1 mL of cocktail / 40 mg resin). The resin was filtered, and the cocktail was added onto ice-cold diethyl ether. After 10 - 30 min, the precipitate was centrifuged and washed again with of ice-cold ether. The solid residue was dried under argon and re-dissolved in water.

br, **brHis, br(S)** and **brHis$_2$**, included the p-acetamidobenzoic acid (p-ABA) at their N-terminal to facilitate quantification.

HPLC and UHPLC

UHPLC-MS: Peptide analysis was performed by analytical UHPLC-MS with an *Agilent 1200* series LC/MS using a *SB C18* (1.8 µm, 2.1 × 50mm) analytical column from *Phenomenex*. Standard conditions for analytical UHPLC consisted on a linear gradient from 5% to 95% of solvent B in 12 min at a flow rate of 0.350 mL/min (A: H_2O 0.1% TFA, B: CH_3CN 0.1% TFA). Compounds were detected by UV absorption at 222, 270 and 330 nm. Electrospray Ionization Mass Spectrometry (ESI/MS) was performed with an *Agilent 6120 Quadrupole* LC/MS model in positive scan mode using direct injection of the purified peptide solution into the MS detector.

Semipreparative HPLC: Purification of the peptides was performed on a semipreparative RP-HPLC with an *Agilent* 1100 serie LC equipped with a UV-visible detector using a *Phenomenex Luna–C18* (250 × 10 mm) reverse-phase column. Standard conditions for purification by RP−HPLC consisted on a linear gradient 5 to 75% B over 40 min at a flow rate of 4 mL/min. (A: H_2O 0.1% TFA, B: CH_3CN 0.1% TFA). Collected fractions with pure products were lyophilized with a *ThermoSavant Modulyo D* lyophilizator equipped with an *Edwars RV* hihgh vacuum pump.

DNA hybridization

Oligonucleotides were obtained from *Thermo Fischer*. Concentration of the oligonucleotides was measured by UV-VIS by using the extinction coefficient given by the Comercial House. For the hybridization process, a stochiometric mixture of the DNA strands was prepared in PBS buffer (10 mM, 100 mM NaCl, pH 7.0) and heated at 90 °C for 10 minutes. After this time, the mixture was slowly cooled down until room temperature, obtaining the doble strand of DNA.

EMSA experiments

EMSA experiments were 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 over 30 min in 18 mM Tris-HCl (pH 7.5), 90 mM KCl, 1.8 mM MgCl2, 0.2 mM TCEP, 9% glycerol, 0.11 mg/mL BSA and 2.2% NP40. For the experiments we used 75 nM of the dsDNAs, and a total incubation volume of 20 µL. After incubation for 30 min, products were resolved by PAGE using a 10% non-denaturing polyacrylamide gel and 0.5× TBE buffer (0.445M Tris, 0.445 M Boric acid) for 40 min and analyzed by staining with SyBrGold (Molecular Probes: 5 µL in 50 mL of 1× TBE) for 10 min. and visualized by fluorescence. The temperatures used in each step in the different experiments, is indicated in the caption of each experiment.

Fluorescence measurements

Fluorescence assays: Measurements of fluorescence were performed using a *Varian Cary Eclipse* Fluorescence Spectrophotometer coupled to a *Cary* Single Cell Peltier accessory *(Agilent Technologies)* temperature controller. All measurements were made with a *Hellma* semi-micro cuvette (108F-QS) at 20 °C. The measurements were made with the following settings: increment 1.0 nm, averaging time 0.1 s, excitation slit width 5.0 nm, emission slit width 10.0 nm, PMT voltage 700 V.

Anisotropy assays: Measurements were made with a *Jobin-Yvon Fluoromax-3*, (DataMax 2.20) coupled to a *Wavelength Electronics* LFI−3751 temperature controller, using a *Hellma* micro cuvette of 1 ml. Settings: integration time: 2.0 s; excitation slit width: 5.0 nm; emission slit width: 20.0 nm; excitation wavelength 559 nm; emission wavelength 585 nm.

Generally, *TMR-**oligo*** (5 µL, 5 µM) was added to 995 µL of Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl, 0.02 mM of corresponding metal complex and calf thymus DNA (when present, 50 µM in base pairs), and the anisotropy was measured. Aliquots of a stock solution in water of the corresponding peptide (12.5 µM) were successively added to this solution, and the anisotropic value was recorded after each addition.

Fluorescence anisotropy curve fitting: Experimental data were fitted with the *DynaFit 4.0* software, which performs a numerical treatment of the system.[242] The program is available free of charge for academia at *http://www.biokin.com/dynafit/*.
Dynafit requires plain text files called scripts that contains information about the chemical model underlying the experimental data, the values of model parameters, such as starting concentrations of reactants, as well as information about location of the files. A typical script used in the analysis titrations is included below. The file has been commented to indicate the purpose of the keywords and sections, but the reader is

[242] a) P. Kuzmič, *Methods in Enzymology*. **2009**, *467*, 247–280; b) P. Kuzmic, *Anal. Biochem.* **1996**, *237*, 260–273.

recommended to review the *DynaFit* scripting manual distributed along the program or available at the *DynaFit* website.

```
[task]                          ;semicolons indicate comments from actual instructions
 task = fit                     ;nature of the calculation to be performed
 data = equilibria

[mechanism]                     ;Free-form 1:1 binding model with Kd
 R + L <==> RL : Kd dissoc      ;to be calculated as dissociation constant

[constants]                     ;Initial Kd value for iteration
 Kd = 1.0 ?                     ;the "?" indicates that this will be optimized

[concentrations]
 R = 2.0                        ;Fixed conc. of the DNA during the peptide titration

[responses]                     ;contribution to the spectroscopic signal of each
 R = 0.1 ?                      ;of the different components of the equilibrium
 RL = 1.5 ?                     ;these will be optimized ("?" after the values).

[data]                          ;location of files and information about the data
 variable  L                    ;the species that changes conc. during the titration
 offset auto ?
 directory ./exp/brHis          ;file path (relative to DynaFit program location)
 extension txt
 file f1                        ;name of the experimental data file

[output]
 directory ./exp/brHis/out      ;path indicating location of DynaFit output files

[settings]                      ;cosmetic settings that control DynaFit graphics
 {Output}                       ;fits were exported & finally plotted with
 XAxisUnit = uM                 ;GraphPad Prism 7.0c. GraphPad Software,
 BlackBackground = n            ;La Jolla California, www.graphpad.com
 XAxisLabel = [peptide]
 YAxisLabel = anisotropy
 WriteTXT = y
```

Circular Dichroism

Circular Dichroism experiments were performed on a *Jasco-715* coupled with a thermostat *Nestlab* RTE-111. The settings used were: Acquisition range: 300-195 nm; band width: 2.0 nm; resolution: 0.2 nm; accumulation: 5 scans; sensitivity 10 mdeg; response time: 0.25 s; speed: 100 nm/min. Measurements were carried out in a 2 mm *Hellma* cuvette at 25 ºC or 4 ºC, depending of experiment required. The mixtures were incubated for 5 min before registering. The spectra are the average of 5 scans and were processed using the "smooth" macro implemented in the program *KaleidaGraph* (v 3.5 by Synergy Software).

Generally, the samples contained 10 mM phosphate buffer pH 7.5 and 100 mM of NaCl, 5 μM peptide, 5 μM of corresponding dsDNA (when present) and different equivalents of metal complexes (when present). 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.

UV-VIS spectroscopy

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 (108.002-QS) with a light path of 10 mm. Measurements were made at 20 °C.

Acquisition parameters were: 220-700 nm range, scan speed of 200 nm/min, resolution of 0.2 nm.

Mass Spectrometry

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-adsorbed peptide. Matrix was selected in function of the molecular weight of the peptides, being 4-HCCA (α-cyano-4-hydroxycinnamic acid) the chosen one for smaller peptides and sinapinic acid for those larger than 3 KDa.

NMR

^{1}H NMR (300 MHz) and ^{13}C NMR (75 MHz) spectra were recorded at room temperature on a *Varian Mercury* 300 MHz spectrometer. ^{13}C NMR (126 MHz) were recorded on a *Bruker DRX-500* spectrometer. Data are represented as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, m = multiplet, bs = broad singlet, dd = doublet doublets, dt = doublet triplets, dq = doublet quartets, coupling constants in Hertz (Hz)). The chemical shifts for protons (δ) are reported in parts per million downfield from tetramethylsilane and are referenced to residual protium in the NMR solvent ($CHCl_3$ δ = 7.26). Chemical shifts for carbon are reported in parts per million downfield from tetramethylsilane and are referenced to the carbon resonances of the solvent ($CDCl_3$ δ = 77.0). NMR spectra were analyzed using *MestreNova© NMR* data processing software (www.mestrelab.com).

Dynamic Light Scattering

The experiments were carried out with a *Malvern Nano ZS* (Malvern Instruments, U.K.) operating at 633 nm with dispersion angle of 173 deg. Measurements were performed in a 2 mm cuvette at 25 ºC and attenuator 11.

Cell assays

All steps were performed on a sterile clean bench *Tesltar AV-100* at room temperature. Solutions stored in a fridge were warmed beforehand in a water bath (37 °C). Unless otherwise specified, all incubations were performed in DMEM containing 5% of fetal bovine serum (FBS-DMEM).

Cell Culture: All cell lines were cultured in DMEM (Dulbecco´s modified Eagle´s medium), 5 mM glutamine, penicillin (100 units/mL) and streptomycin (100 units/mL) (all from *Invitrogen*). Proliferating cultures were maintained in a 5% CO_2 humidified incubator at 37 °C. For all the experiments, cells were seeded in the corresponding well at the indicated concentration two days before treatment.

Fluorescence microscopy: All images were obtained with an *Andor Zyla* mounted on a *Nikon TiE*. Confocal images were acquired in an *Andor Dragonfly* High Speed Confocal Platform. Images were further processed with *Image J* or *NIS* software (Nikon).

Microscopy settings: The filter sets for the observation of the fluorescence of the products were as follows:

- Blue channel

Widefield: LED λ excitation: 385 nm. Filter cube DAPI-1160B-000 (Semrock): BP 387/11 nm, LP447/60 nm and DM 409 nm. HBTP filter: LED λ excitation: 385 nm. Filter cube: BP 375/28x nm, LP 515lp nm and DM 415 nm.
Confocal: Laser excitation: 405 nm. LP 450/50 and DM 418 nm.

- Green channel

Widefield: LED λ excitation: 470 nm. Filter cube FITC-3540C-000: BP 482/35 nm, LP 536/40 nm and DM 506 nm.
Confocal: Laser excitation: 488 nm. LP 525/50 and DM 501 nm.

- Red channel

Widefield: LED λ excitation: 550 nm. Filter cube TRITC-B-000 (Semrock): BP 543/22 nm, LP 593/40 nm and DM 562 nm.
Confocal: Laser excitation: 561 nm. LP 620/60 and DM 567 nm.

Cell internalization studies: Cells were seeded on glass-bottom plates 48 h before treatment. Culture medium was removed and DMEM containing 5% fetal bovine serum (FBS-DMEM) and peptides (5 μM) or crude mixtures of palladopeptides (5 μM) were added. Indeed, these palladopeptides were made just before the addition to cells by pre-incubating with metal complexes (1:1) in water for 10 min. After 30 min of incubation with cells at 37 ºC (or 4ºC on ice), these were washed twice with PBS and replaced with fresh FBS-DMEM to observe under the microscope (or measure by cytometer) with adequate filters. Digital pictures of the different samples were taken under identical conditions of gain and exposure.

Viability test (MTT assay): The toxicity of the palladopeptide, peptides and metal complexes were tested by MTT assays in HeLa cells as follows: 100,000 cells per well were seeded in 96 well plates two days before treatment with different concentrations of the catalysts/peptides. After 24 h of incubation, HEPES containing 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) was added to the cell culture medium to a final concentration of 0.5 mg/ml. Cells were then incubated for 4 h to allow the formation of formazan precipitates by metabolically active cells. A detergent solution of 10% SDS (sodium dodecyl sulphate) and 0.01 M HCl was then added and the plate was incubated overnight at room temperature to allow the solubilization of the precipitates. The quantity of formazan in each well (directly proportional to the number of viable cells) was measured by recording changes in absorbance at 570 nm in a microtiter plate reading spectrophotometer (*Tecan Infinite F200 PRO*).

Cytometry: Flow cytometry was performed on a *Guava easyCyteTM* cytometer. Data analysis was performed with *InCyte* software included in *GuavaSoft 3.2* (*Millipore*). For cytometry assays cells were seeded in 96 well plates two days before treatment. Culture medium was removed and FBS-DMEM with corresponding compounds was added in triplicates. The incubation was performed for 30 min at 37 ºC or 4 ºC on ice. The cells were washed with PBS and trypsinized to measure in the cytometer.

ICP-MS: The metal content of the samples was determined by ICP-MS (*Agilent 7700x*) with a sample introduction system consisting of a *Micromist* glass low-flow nebulizer, a double-pass glass spray chamber with a Peltier system (3 ºC) and a quartz torch.

Buffers and gels

PBS: 137 mM NaCl; 2,7 mM KCl; 8 mM Na_2HPO_4 and 1,5 mM KH_2PO_4.

Lysis buffer bacteria: 20 mM Tris-HCl (pH 7,5); 200 mM NaCl; 1mM DTT; 2 mM $MgCl_2$; 0,25 % Tween 20; 1 mg/ml lysozyme and 10 mM protease inhibitor.

Agarose Gel: 1 g of agarose was mixed in 100mL of TAE (1%) in a microwavable flask. Microwaved for 1-3 min until the agarose was completely dissolved and led cool down the agarose solution to about 50 °C for 5 mins. Ethidium bromide (EtBr) was added to a final concentration of approximately 0.2-0.5 µg/mL (10 µL per 100 mL). The agarose was poured into a gel tray with the well comb in place, until it had completely solidified. Carefully loaded a molecular weight ladder into the first lane of the gel and the samples into the additional wells. The gel was run at 150 V about 30min. The visualization was performed with ChemiDoc™ MP imaging system by *Biorad*.

Electrophoresis buffer for Agarose gels (TAE): 40 mM Tris; 1,1%, acetic acid and 1 mM EDTA, (pH 8,3).

SDS-PAGE gel (12%):

10 ml of separating gel	4 ml of stacking gel
3.4 ml H_2O	2.7 ml H_2O
4 ml acrylamide (30 %)	0.8 ml acrylamide (30 %)
2.5 ml 1.5 M tris (pH 8.8)	0.5 ml 1 M tris (pH 6.8)
50 µl SDS (20%)	20 µl SDS (20%)
100 µl APS (10%)	40 µl APS (10%)
10 µl TEMED	4 µl TEMED

After the gel was polymerized (30 min), the glass plates were took out of the casting frame and set them in the cell buffer dam. The electrophoresis buffer was added into the inner chamber and kept pouring after overflow until the buffer surface reached the required level in the outer chamber. The samples were prepared mixing them with loading buffer (Laemmli buffer) and boiled for 5-10 min. The loading protein marker was loaded into the first lane and the samples into the additional wells. Generally, the gel run about 1 hour for a 120V voltage.

Electrophoresis buffer for SDS-PAGE (Tris-Glycine-SDS): 25 mM Tris-HCl (pH 8,3); 192 mM glycine and 0,1 % SDS.

Laemmli Buffer (3x): 20 % glycerol; 60 mM Tris-HCl (pH 6,8); 5% β-mercaptoethanol; 0,05 % bromophenol blue and 10 % SDS.

Coomassie stain: 2g of Coomassie Blue was dissolved in 250ml water, and 75ml of glacial acetic acid and 500ml of ethanol were added. It was made up to 1000ml with water.

Distaining solution: Methanol 50%. Acetic acid 10%. Distilled water 40%.

Western Blot: When the running finished, the gel was dislodged from the glass plates and the stacking gel was cut. The sandwich was assembled as follows: Black side of the cassette, sponge, blotting paper, gel, Immobilon membrane, blotting paper, sponge, clear side of the cassette. The transfer buffer was used to keep wet the sandwich. The sandwich was closed and inserted it into the transfer electrode and into the transfer tank. An ice pack was added inside as well and filled the tank with transfer buffer. The transfer was for 1 hour at 100 V. After the transfer is done, the membrane was placed in a clear plastic container and put about 5 mL PBS-T-Milk, keeping it in a rocker for blocking for 1 hour. After that antibodies were added following their protocols and washed using PBS-T. Finally, Horseradish peroxidase (HRP) substrate was added and visualized using ChemiDoc™ MP imaging system by *Biorad*.

Transfer Buffer: 25 mM Tris-HCl (pH 8,3); 192 mM glycine and 20 % methanol.

PBS-T-Milk: PBS-T with 4 % of skim milk powder.

PBS-T: PBS with 0,05 % of Tween-20.

Chapter I: Section 1

UHPLC-MS characterization of peptides

gaga: H-SQSEQPATCP ICYAVIRQSR NLRRHLELRH F-CONH$_2$

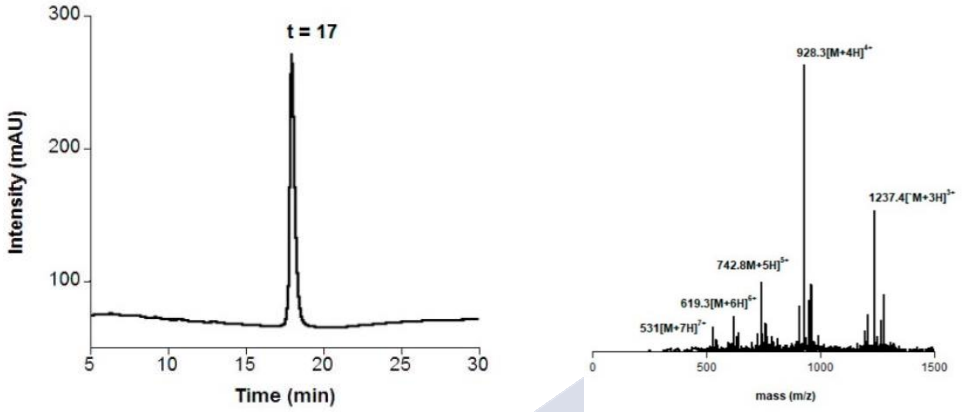


Figure 100. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right:* Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{159}H_{258}N_{54}O_{45}S_2$: 3710. Found: 1237.4 $[M+4H]^{3}$; 928.3 $[M+4H]^{4+}$; 742.8 $[M+5H]^{5+}$; 619.3$[M+6H]^{6+}$; 531 $[M+7H]^{7+}$.

gaga-Hk: H-SQSEQPATCP ICYAVIRQSR NLRRHLELRH FGGGGKPRGR PKK-CONH$_2$

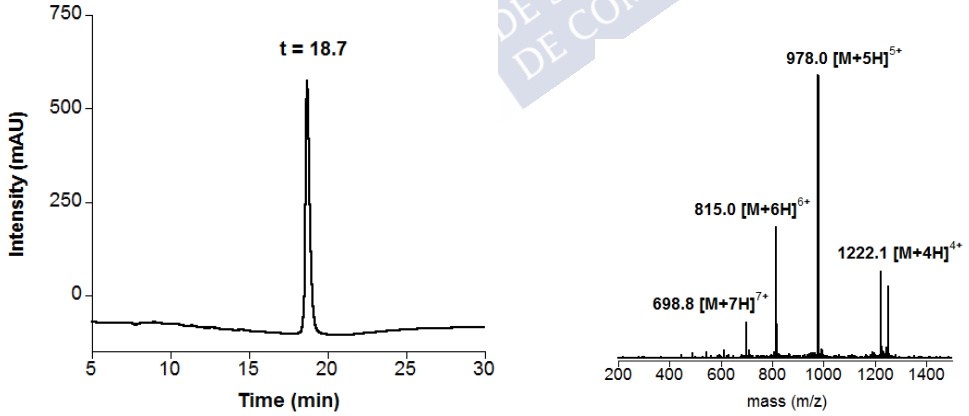


Figure 101. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{209}H_{347}N_{75}O_{57}S_2$: 4883.6. Found: 1222.1 $[M+4H]^{4+}$; 978.0 $[M+5H]^{5+}$; 815.0 $[M+6H]^{6+}$; 698.8 $[M+7H]^{7+}$.

Hk-gaga: H-KPRGRPGKGK GGSQSEQPAT CPICYAVIRQ SRNLRRHLEL RHF-$CONH_2$

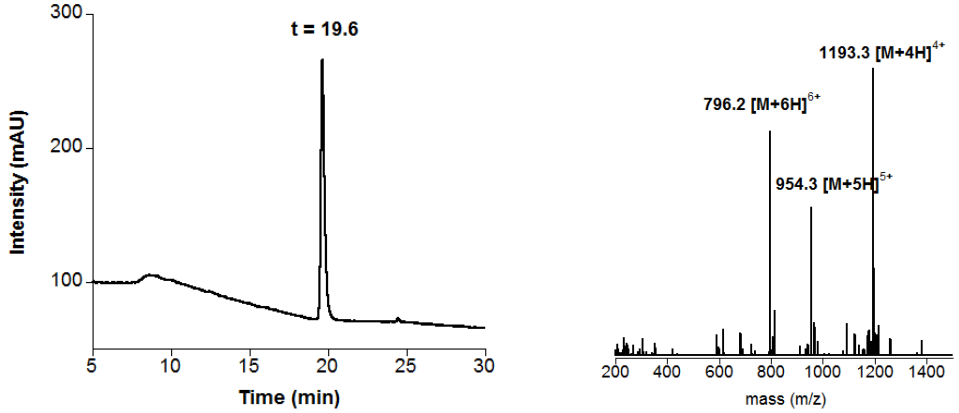


Figure 102. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI^+ (m/z): Calcd. for $C_{206}H_{342}N_{72}O_{55}S_2$: 4768.6. Found: 1193.3 $[M+4H]^{4+}$; 954.3 $[M+5H]^{5+}$; 796.2 $[M+6H]^{6+}$.

gaga-Hk-gaga: H-SQSEQPATCP ICYAVIRQSR NLRRHLELRH FGGGGKPRGR PGKGKGGSQS EQPATCPICY AVIRQSRNLR RHLELRHF-$CONH_2$

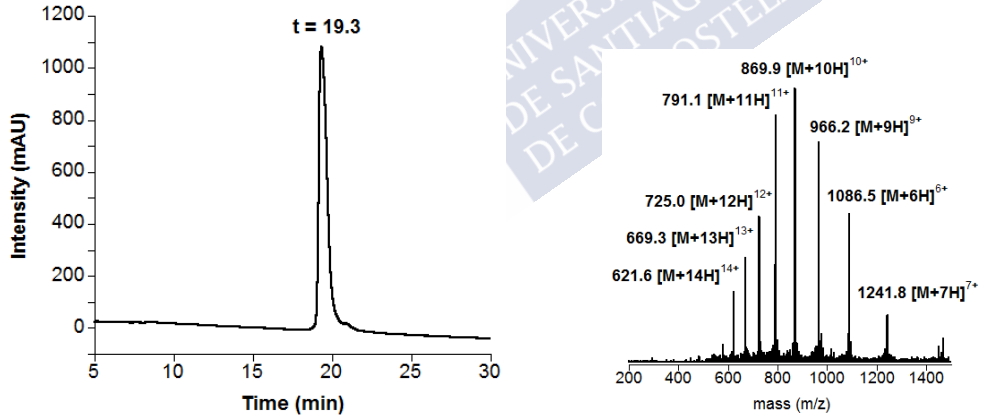


Figure 103. *Left:* HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right:* Mass spectrum of the purified peptide.

EM–ESI^+ (m/z): Calcd. for $C_{373}H_{610}N_{130}O_{103}S_4$: 8686.5. Found: 1241.8 $[M+7H]^{7+}$; 1086.5 $[M+8H]^{8+}$; 966.2 $[M+9H]^{9+}$; 869.9 $[M+10H]^{10+}$; 791.1 $[M+11H]^{11+}$; 725.0 $[M+12H]^{12+}$; 669.3 $[M+13H]^{13+}$; 621.6 $[M+14H]^{14+}$.

Superpositions

This was carried out with *MacPymol* using the references PDB ID: 3UXW for AT-Hook and 1YUI for GAGA.

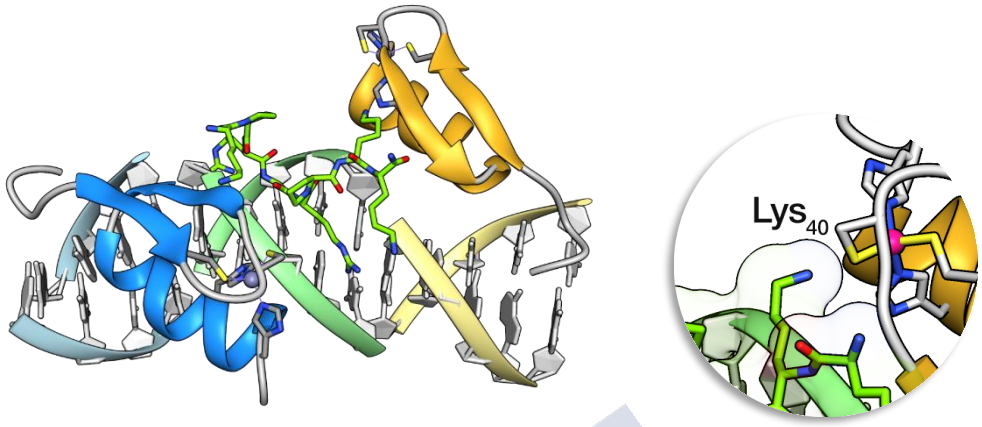


Figure 104. *Left*: Model of the simultaneous interaction of the **AT-Hook** flanked by two **GAGA** fragments (blue and orange). *Right*: Zoom of the model showing the damaging steric clash between Lys^{40} of the AT-hook and the beta-hairpin of the second **GAGA** fragment.

Oligonucleotide sequences

Double stranded (only one strand is shown, consensus sites underlined) oligonucleotides used for EMSA experiments with peptides **gaga-Hk**, **Hk-gaga** and **gaga-Hk-gaga**:

AT•GAG* or *cat•AT•GAG	5´-CGC GTCAT AATT GAGAG CGC-3´
AT•cgc* or *cat•AT•cgc	5´-CGC GTCAT AATT CGCGA CGC-3´
cg•GAG	5´-CGC GTCAT CAGC GAGAG CGC-3´
CTC•AT•GAG	5´-GAG CTCTC AATT GAGAG CGCG-3´
CTC•gc•GAG	5´-GGTT CTCTC GACC GAGAG TTGG-3´
*TMR-**AT•GAG***	5´– (TMR)-CGC GTCAT AATT GAGAG CGC-3´
*TMR-**cg•GAG***	5´– (TMR)-CGC GTCAT CAGC GAGAG CGC-3´
*TMR-**CTC•AT•GAG***	5´– (TMR)-GAG CTCTC AATT GAGAG CGCG-3´

Oligonucleotides used for nanopore experiments were supplied by *Thermo Fischer* and their sequences were:

DNA with the tripartite binding site: 5´-CCCCCCCCCCCCCCCCCCCCAAAAAAAAAAGAG CTCTC*AATT*GAGAGCGCGCACACACGCGCTCTC*AATT*GAGAGCTC-3´

DNA mutated in the 1st GAGAG site: 5´CCCCCCCCCCCCCCCCCCCCAAAAAAAAAAGAG GTCAT*AATT*GAGAGCGCGCACACACGCGCTCTC*AATT*ATGACCTC 3´

DNA mutated in the 2nd GAGAG site: 5´CCCCCCCCCCCCCCCCCCCCAAAAAAAAAAGAG CTCTC*AATT*CTGCCCGCGCACACACGCGGGCAG*AATT*GAGAGCTC 3´

Single-molecule nanopore experiments

Single-molecule measurements were performed in planar lipid bilayers made of 1,2-diphytanoyl-sn-glycero-3-phosphocholine (*Avanti Polar Lipids)* generated with the Montal and Mueller method across an aperture in a Teflon film (0.025 mm thick, Goodfellow) that separated the two compartments of a chamber (*cis* and *trans*). The measurements were recorded at 20 ºC in asymmetric conditions. The *cis* compartment (connected to the ground electrode and where samples were added) was filled with 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.02 mM of $ZnCl_2$ and the *trans* compartment with 5 mM $MgCl_2$, 2 M KCl, 10 mM HEPES pH 7.2 (to increase the ionic current signal). 0.5 µL of previously heptamerized α-hemolysisn (αHL) was added to the *cis* compartment and following the first insertion the chamber was manually perfused to remove remaining αHL. Ionic currents through the αHL pore were measured using previously balanced Ag/AgCl electrodes with agar bridges in 3 M KCl. The electrodes were connected to an amplifier (*Axopatch 200B*, Molecular Devices), the signal was filtered at 5 kHz and recorded at 20 kHz with a digitizer (*Digidata 1440A*, Molecular Devices). Data was obtained in a range of DNA concentrations (0.12-0.15 µM) and peptide concentrations (0.038-0.5 µM) at different voltages (from +90 mV to 120 mV). Raw data was first analysed with *pClamp* software (Molecular Devices) to measure the dwell time in the protein-DNA state (level 1). The fraction of protein bound DNA molecules was estimated dividing the number of signals with an initial step by the total number of signals corresponding to DNA molecules. All dwell times in the protein-DNA state were used to make histograms as shown in Figure 6 (rows 2, 4 and 6). These histograms were fitted using the Curve Fitting routine in Igor Pro (Wavemetrics) and the equation $y = A*\exp(-\lambda*x)$ with A = amplitude, λ= rate parameter and x= dwell time. Alternatively, the natural logarithm of the dwell time in the protein-DNA state was used to make the histograms as shown in Figure 6 (rows 1, 3 and 5). These histograms were fitted using the equation $y = A*\lambda*\exp(x-\lambda*\exp(x))$ with A = area; λ= rate parameter; x= ln(dwell time).

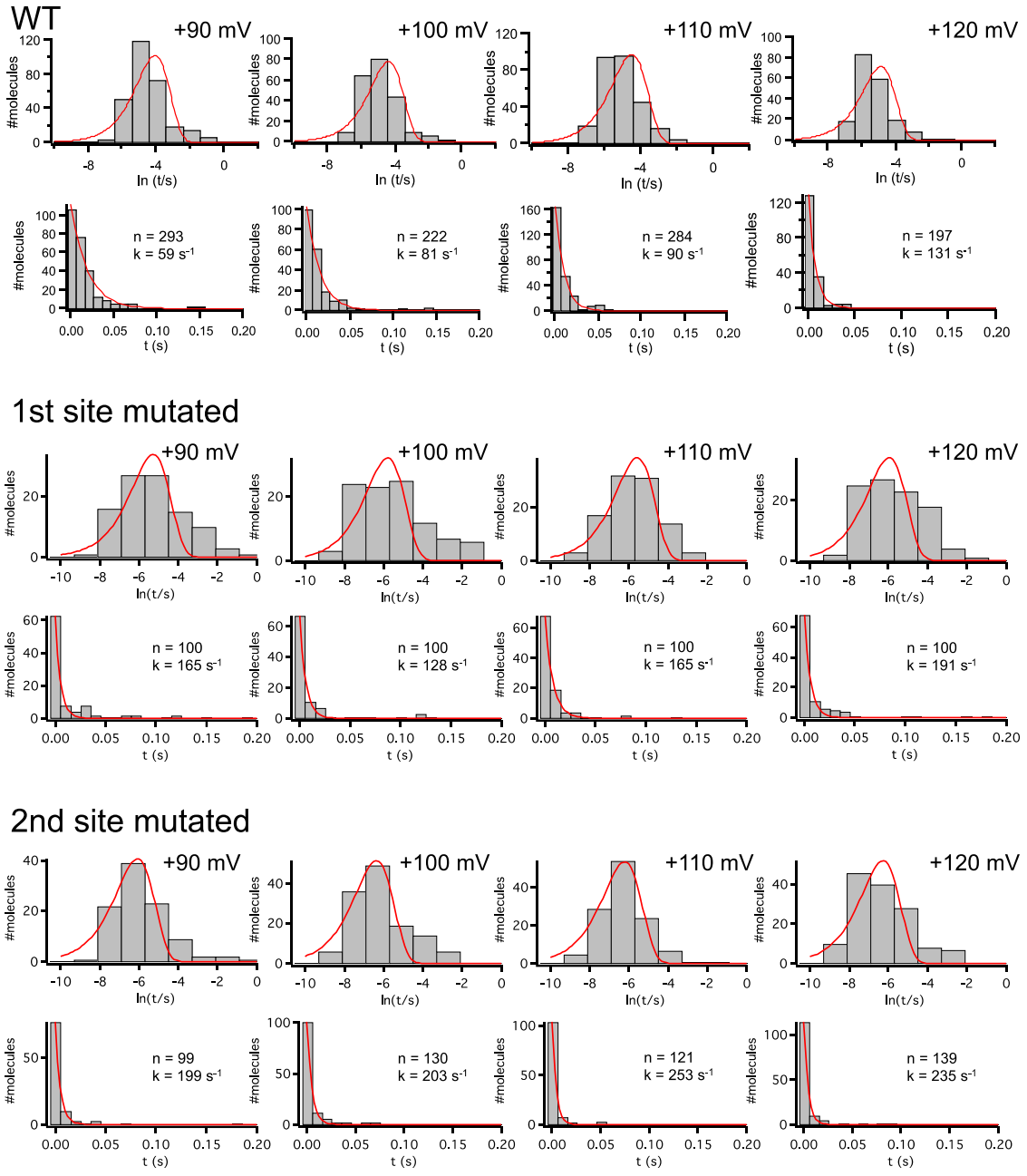


Figure 105. Kinetics (k_{off}) of detachment for the trivalent peptide at different voltages for a DNA that contains the copnsensus tripartite binding site (WT, rows 1 and 2); for a DNA with the 1st GAGAG binding site mutated (rows 3 and 4); and for the DNA with the 2nd GAGAG binding site mutated (rows 5 and 6). The data are displayed both in linear and in logarithmic scale. In red the fit to a single exponential distribution.

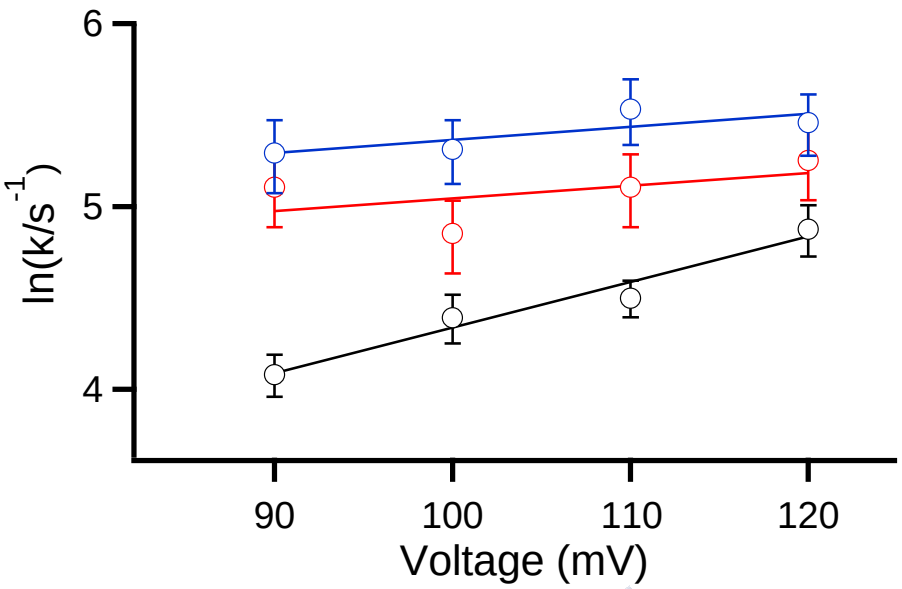


Figure 106. The detachment rates (k_{off}) show a weak dependence on voltage. In black the data for the DNA with the tripartite binding site. In red the data for the DNA with the 1st GAGAG site mutated. In blue the data for the DNA with the 2nd GAGAG site mutated. The lines are the best fit to a straight line. Error bars represent the 95% confidence interval in the rate estimation.

Cell experiments

Plasmid Constructions

All plasmids used were constructed with classic molecular cloning methods or the Gibson assembly method[243] and are listed in the Table 1. The nucleotide and amino acid sequences are provided Table 2.

Amino acid sequences for encoding artificial miniproteins were synthetized "*de novo*" by *GenScript,* amplified with PCR with the correspondent primers. All artificial miniproteins sequence, linkers, tags, the nuclear localization sequence, and activators factors were inserted as part of PCR amplification primers and overlapping sequences for the Gibson assembly. All of them were expressed from the CMV promoter of the pcDNA3 vector.

The binding sites were synthesized by *GenScript*, as an unique copy or as multiple repeats, and PCR amplified with a primer containing a minimal promoter (5´-TAGAGGGTATATAATGGAAGCTCGACTTCCAG-3´) and cloned upstream of the firefly luciferase gene in pGL4.16 (*Promega*), replacing the original inducible promoter. The transfection control plasmid used in the dual luciferase assay was phRL-TK (*Promega*).

[243] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343–345.

Table 1. Plasmids used in the study. All constructs containing the CMV promoter were inserted in the pcDNA3 backbone, and all constructs containing the minimal promoter were inserted in the pGL4.16 backbone.

Backbone/plasmid name	Description
Protein coding plasmids	
pcDNA3/CMV:[gaga:Hk]:gs10:NLS:VP16:Au1	Plasmid ensures constitutive expression of the peptide **gaga:Hk** fused to a flexible GS linker, the SV40 NLS signal, the VP16 activation domain and the Au1 peptide tag at the C-terminus.
pcDNA3/CMV:[gaga:Hk]:gs10:NLS:VPR:Au1	Plasmid ensures constitutive expression of the peptide **gaga:Hk** fused to a flexible GS linker, the SV40 NLS signal, the VPR activation domain and the Au1 peptide tag at the C-terminus.
pcDNA3/CMV:[gaga:Hk(G):gaga]:gs10:NLS:VP16:Au1	Plasmid ensures constitutive expression of the peptide **gaga:Hk(G):gaga** fused to a flexible GS linker, the SV40 NLS signal, the VP16 activation domain and the Au1 peptide tag at the C-terminus.
pcDNA3/CMV:[gaga:Hk(G):gaga]:gs10:NLS:VPR:Au1	Plasmid ensures constitutive expression of the peptide **gaga:Hk(G):gaga** fused to a flexible GS linker, the SV40 NLS signal, the VPR activation domain and the Au1 peptide tag at the C-terminus.
pcDNA3/CMV:[Hk(G):gaga:HK]:gs10:NLS:VP16:Au1	Plasmid ensures constitutive expression of the peptide **Hk(G):gaga:HK** fused to a flexible GS linker, the SV40 NLS signal, the VP16 activation domain and the Au1 peptide tag at the C-terminus.
pcDNA3/CMV:[Hk(G):gaga:HK]:gs10:NLS:VPR:Au1	Plasmid ensures constitutive expression of the peptide **Hk(G):gaga:HK** fused to a flexible GS linker, the SV40 NLS signal, the VPR activation domain and the Au1 peptide tag at the C-terminus.
Reporter plasmids	
pGL4.16_1x[gaga:Hk]bs_Pmin_fLuc	Plasmid ensures expression of the firefly luciferase reporter upon binding of a transcriptional activator to target sites upstream of a minimal promoter. The plasmid contains 1 copy of target sites for **gaga:Hk** DNA-binding domain.
pGL4.16_4x[gaga:Hk]bs_Pmin_fLuc	Plasmid ensures expression of the firefly luciferase reporter upon binding of a transcriptional activator to target sites upstream of a minimal promoter. The plasmid contains 4 copies of target sites for **gaga:Hk** DNA-binding domain.
pGL4.16_1x[gaga:Hk(G):gaga]bs_Pmin_fLuc	Plasmid ensures expression of the firefly luciferase reporter upon binding of a transcriptional activator to target sites upstream of a minimal promoter. The plasmid contains 1 copy of target sites for **gaga:Hk(G):gaga** DNA-binding domain.
pGL4.16_4x[gaga:Hk(G):gaga]bs_Pmin_fLuc	Plasmid ensures expression of the firefly luciferase reporter upon binding of a transcriptional activator to target sites upstream of a minimal promoter. The plasmid contains 4 copies of target sites for **gaga:Hk(G):gaga** DNA-binding domain.
pGL4.16_4x[Mut_GAGA]bs_Pmin_fLuc	Plasmid ensures expression of the firefly luciferase reporter upon binding of a transcriptional activator to target sites upstream of a minimal promoter. The plasmid contains 4 copies of mutated site for **gaga** DNA-binding domain.
pGL4.16_4x[Mut_AT]bs_Pmin_fLuc	Plasmid ensures expression of the firefly luciferase reporter upon binding of a transcriptional activator to target sites upstream of a minimal promoter. The plasmid contains 4 copies of mutated site for **Hk** DNA-binding domain.
pGL4.16_4x[Hk(G):gaga:Hk]bs_Pmin_fLuc	Plasmid ensures expression of the firefly luciferase reporter upon binding of a transcriptional activator to target sites upstream of a minimal promoter. The plasmid contains 4 copies of target sites for **Hk(G):gaga:Hk** DNA-binding domain.

phRL-TK/TK_rLuc	Plasmid ensures constitutive expression of the *Renilla* luciferase, driven from the TK promoter. The plasmid was used as a transfection control in luciferase experiments.
pcDNA3	pcDNA3 plasmid without an insert. The plasmid was used to equalize total amount of transfected DNA or as a mock control in different experiments.

Table 2. Nucleotide and amino acid sequences of the constructs used in the study

Nucleotide/ amino acid sequence
P_{min} Minimal promoter tagagggtatataatggaagctcgacttccag
gaga:Hk 1 x binding site ctctcaatt gcgctctcaattatgacgcg
gaga:Hk 4 x binding site ctctcaatt gcgctctcaattatgacgcggcgctctcaattatgacgcggcgctctcaattatgacgcggcgctctcaattatgacgcg
gaga:Hk:gaga 1 x binding site ctctcaattgagag gagctctcaattgagagcgcg
gaga:Hk:gaga 4 x binding site ctctcaattgagag gagctctcaattgagagcgcggagctctcaattgagagcgcggagctctcaattgagagcgcggagctctcaattgagagcgcg
Mut_GAGA 4 x binding site only aatt gagctctcaattgagagcgcggagctctcaattgagagcgcggagctctcaattgagagcgcggagctctcaattgagagcgcg
Mut_AT 4 x binding site only ctctc__gagag ggttctctcgaccgagagttggggttctctcgaccgagagttggggttctctcgaccgagagttggggttctctcgaccgagagttgg
Hk(G):gaga:HK 4 x binding site aattctctccaatt cgaattctctcaattcgcggcgaattctctcaattcgcggcgaattctctcaattcgcggcgaattctctcaattcgcg
l_1 Sequence between binding sites and P_{min} tggctggtgccgcaccggtaagcagaggatct
l_2 Sequence between P_{min} and Kozak sequence ctcgagggcaatccggtactgttggtaaagccacc
fLuc Firefly luciferase from the plasmid vector pGL4.16 gaagatgccaaaaacattaagaagggcccagcgccattctacccactcgaagacgggaccgccggcgagcagctgcacaaagccatgaagcgctacgccctggtgcccggcaccatcgccttt accgacgcacatatcgaggtggacattacctacgccgagtacttcgagatgagcgttcggctggcagaagctatgaagcgctatgggctgaatacaaaccatcggatcgtggtgtgcagcgaga atagcttgcagttcttcatgcccgtgttgggtgccctgttcatcggtgtggctgtggccccagctaacgacatctacaacgagcgcgagctgctgaacagcatgggcatcagccagcccaccgtcgt attcgtgagcaagaaagggctgcaaaagatcctcaacgtgcaaaagaagctaccgatcatacaaaagatcatcatcatggatagcaagaccgactaccagggcttccaaagcatgtacaccttcg tgacttcccatttgccacccggcttcaacgagtacgacttcgtgcccgagagcttcgaccgggacaaaaccatcgccctgatcatgaacagtagtggcagtaccggattgcccaagggcgtagccc taccgcaccgcaccgcttgtgtccgattcagtcatgcccgcgaccccatcttcggcaaccagatcatccccgacaccgcatcctcagcgtggtgccatttcaccacggcttcggcatgttcaccacgc tgggctacttgatctgcggctttcgggtcgtgctcatgtaccgcttcgaggaggagctattcttgcgcagcttgcaagactataagattcaatctgccctgctggtgcccacactatttagcttcttcgct aagagcactctcatcgacaagtacgacctaagcaacttgcacgagatcgccagcggcggggcgccgctcagcaaggaggtaggtgaggccgtggccaaacgcttccacctaccaggcatccg ccagggctacggcctgacagaaacaaccagcgccattctgatcacccccgaaggggacgacaagcctggcgcagtaggcaaggtggtgcccttcttcgaggctaaggtggtggacttggaca ccggtaagacactgggtgtgaaccagcgcggcgagctgtgcgtccgtggcccccatgatcatgagcggctacgttaacaaccccgaggctacaaacgctctcatcgacaaggacggctggctgc acagcggcgacatcgcctactgggacgaggacgagcacttcttcatcgtggaccggctgaagagcctgatcaaatacaagggctaccaggtagccccagccgaactggagagcatcctgctgc aacaccccaacatcttcgacgccggggtcgccggcctgcccgacgacgatgccggcgagctgcccgccgcagtcgtcgtgctggaacacggtaaaaccatgaccgagaaggagatcgtggac tatgtggccagccaggttacaaccgccaagaagctgcgcggtggtgttgtgttcgtggacgaggtgcctaaaggactgaccggcaagttggacgcccgcaagatccgcgagattctcattaagg ccaagaagggcggcaagatcgccgtgaattctgcttgcaagaactggttcagtagcttaagccactttgtgatccacctaacagccacggcttccctcccgaggtggaggagcaggccgccggc accctgcccatgagctgcgcccaggagagcggcatggatagacaccctgctgcttgcgccagcgccaggatcaacgtc EDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVDITYAEYFEMSVRLAEAMKRYGLNTNHRIVVCSENSLQFFMPVLGALFIGVAVAP ANDIYNERELLNSMGISQPTVVFVSKKGLQKILNVQKKLPIIQKIIIMDSKTDYQGFQSMYTFVTSHLPPGFNEYDFVPESFDRDKTIALIMNSSGSTGLPKGVA LPHRTACVRFSHARDPIFGNQIIPDTAILSVVPFHHGFGMFTTLGYLICGFRVVLMYRFEEELFLRSLQDYKIQSALLVPTLFSFFAKSTLIDKYDLSNLHEIASGGA PLSKEVGEAVAKRFHLPGIRQGYGLTETTSAILITPEGDDKPGAVGKVVPFFEAKVVDLDTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALIDKDGWLH SGDIAYWDEDEHFFIVDRLKSLIKYKGYQVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAAVVVLEHGKTMTEKEIVDYVASQVTTAKKLRGGVVFVDEV PKGLTGKLDARKIREILIKAKKGGKIAVNSACKNWFSSLSHFVIHLNSHGFPPEVEEQAAGTLPMSCAQESGMDRHPAACASARINV

rLuc *Renilla* luciferase from the plasmid vector phRL-TK (*Promega*)

atggcttccaaggtgtacgaccccgagcaacgcaaacgcatgatcactgggcctcagtggtgggctcgctgcaagcaaatgaacgtgctggactccttcatcaactactatgattccgagaagca cgccgag
aacgccgtgatttttctgcatggtaacgctgcctccagctacctgtggaggcacgtcgtgcctcacatcgagcccgtggctagatgcatcatccctgatctgatcggaatgggtaagtccggca agagcgggа
atggctcatatcgcctcctggatcactacaagtacctcaccgcttggttcgagctgctgaaccttccaaagaaaatcatctttgtgggccacgactgggggggcttgtctggccttttcactactcc tacgagcacca
agacaagatcaaggccatcgtccatgctgagagtgtcgtggacgtgatcgagtcctgggacgagtggcctgacatcgaggaggatatcgccctgatcaagagcgaagagggc cgagaaaatggtgcttg
agaataacttcttcgtcgagaccatgctcccaagcaagatcatgcggaaactggagcctgaggagttcgctgcctacctggagccattcaaggagaagggcgaggttagacggcctaccctctcc tggcctc
gcgagatccctctcgttaagggaggcaagcccgacgtcgtccagattgtccgcaactacaacgcctaccttcgggccagcgacgatctgcctaagatgttcatcgag tccgaccctgggttcttttccaacgct
attgtcgagggagctaagaagttccctaacaccgagttcgtgaaggtgaagggcctccactcagccaggaggacgctccagatgaaatgggtaagtacatcaagagcttcgtggagcgcgtgctg
aagaacgagcag

MASKVYDPEQRKRMITGPQWWARCKQMNVLDSFINYYDSEKHAENAVIFLHGNAASSYLWRHVVPHIEPVARCIIPDLIGMGKSGKSGNGSYRLLDHYKYLTAWF
EL

P_{CMV} CMV promoter from plasmid vector pcDNA3

acattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacc cccgcccattga
cgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgcca agtacgcccccattga
cgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagt acatcaatggg
cgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtc gtaacaactccgccccattgacgcaaatggg
cggtaggcgtgtacggtgggaggtctatataagcagagctc

gaga:Hk Sequence of artificial peptide and linker

atgagccagagcgagcagcccgccacctgccccatctgctacgccgtgatcagacagagcagaaacctgagaagacacctggagctgagacacttc

MSQSEQPATCPICYAVIRQSRNLRRHLELRHFGGGGKPRGRPKK

gaga:Hk(G):gaga Sequence of artificial peptide and linkers

atgagccagagcgagcagcccgccacctgccccatctgctacgccgtgatcagacagagcagaaacctgagaagacacctggagctgagacacttcggcggcggcggcaagcccagaggc
agacccggcaagggcaagggcggccagagcgagcagcccgccacctgccccatctgctacgccgtgatcagacagagcagaaacctgagaagacacctggagctgagacacttc

MSQSEQPATCPICYAVIRQSRNLRRHLELRHFGGGGKPRGRPGKGKGGQSEQPATCPICYAVIRQSRNLRRHLELRHF

Hk(G):gaga:HK Sequence of artificial peptide and linkers

atgaagcccagaggcagacccggcaagGgcaagggcggcagccagagcgagcagcccgccacctgccccatctgctacgccgtgatcagacagagcagaaacctgagaagacacctgga
gctgagacacttcggcggcggcggcaagcccagaggcagacccaagaag

MKPRGRPGKGKGGSQSEQPATCPICYAVIRQSRNLRRHLELRHFGGGGKPRGRPKK

gs10 Flexible linker; introduced into constructs with PCR

aagcccagaggcagacccaagaag

GGSGGGSGGSS

NLS SV40 nuclear localization signal; introduced into constructs with PCR

gatccaaaaaagaagagaaaggta

PKKKRKV

VP16 Transcription factor, introduced into constructs with PCR

gcaccccgaccgatgtcagcctgggggacgagctccacttagacggcgaggacgtggcgatggcgcatgccgacgcgctagacgatttcgatctggacatgttgggggacggggattccccgggtcc
gggatttacccccacgactccgcccctacggcgctctggatatggccgacttcgagtttgagcagatgtttaccgatgcccttggaattgacgagtacggtggg

APPTDVSLGDELHLDGEDVAMAHADALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG

VPR Transcription factor, introduced into constructs with PCR

gaggccagcggttcgggacgggctgacgcattggacgattttgatctggatatgctgggaagtgacgccctcgatgattttgaccttgacatgcttggttcggatgcccttgatgactttgacct cg
acatgctcggcagtgacgcccttgatgatttcgacctggacatgctgattaacactagtggcggctctggcggcggctccggaggctctagccagtacctgcccgacaccgacgaccggcaccg
gatcgaggaaaagcggaagcggacctacgagacattcaagagcatcatgaagaagtccccccttcagcggccccaccgaccctagacctccacctagaagaatcgccgtgcccagcagatccag
cgccagcgtgccaaaacctgccccccagccttacccctcaccagcagcctgagcaccatcaactacgacgagttccctaccatggtgttccccagcggccagatctctcaggcctctgctctgg ctc
cagccccctcctcaggtgctgcctcaggctcctgctcctgcaccagctccagccatggtgtctgcactggctcaggcaccagcacccgtgcctgtgctggctcctggacctccacaggctgtggctcc
accagcccctaaacctacacaggccggcgagggcacactgtctgaagctctgctgcagctgcagttcgacgacgaggatctgggagccctgctgggaaacagcaccgatcctgccgtgttcacc
gacctggccagcgtggacaacagcgagttccagcagctgctgaaccagggcatccctgtggcccctcacaccaccgagcccatgctgatggaatacccccgaggccatcacccggctcgtgaca
ggcgctcagaggcctcctgatccagctcctgcccctctgggagcaccaggcctgcctaatggactgctgtctggcgacgaggacttcagctctatcgccgatatggatttctcagccttgctggg ct
ctggcagcggcagccgggattccagggaagggatgtttttgccgaagcctgaggccggctccgctattagtgacgtgtttgagggccgcgaggtgtgccagccaaaacgaatccggccatttc
atcctccaggaagtccatgggccaaccgcccactccccgccagcctcgcaccaacaccaaccggtccagtacatgagccagtcgggtcactgaccccggcaccagtccctcagccactggaccca
gcgcccgcagtgactcccgaggccagtcacctgttggaggaccccgatgaagagacgagccaggctgtcaaagcccttcgggagatggccgatactgtgattccccagaaggaagaggctgc
aatctgtggccaaatggaccttccatccgccccaaggggccatctggatgagctgacaaccacacttgagtccatgaccgaggatctgaacctggactcacccctgaccccggaattgaacg a
gattctggataccttcctgaacgacgagtgcctcttgcatgccatgcatatcagcacaggactgtccatcttcgacacatctctgttt

EASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINTSGGSGGGSGGSSQYLPDTDDRHRIEEKRKRTYETFKSIM
KKSPFSGPTDPRPPPPRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPG
PPQAVAPPAPKPTQAGEGTLSEALLQLQFDDEDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPL
GAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTP
APVPQPLDPAPAVTPEASHLLEDPDEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNLDSPLTPELNEILDTFLNDECLL
HAMHISTGLSIFDTSLF

AU1 Peptide tag, introduced into constructs with PCR

gacacatacaggtatatctaa

DTYRYI

Cell culture and dual luciferase assay

The human embryonic kidney (HEK) 293T cell line was cultured in DMEM medium supplemented with 10% fetal bovine serum (FBS) at 37 ºC in a 5% CO2 environment. 2 × 10^4 cells per well were seeded in a Costar White 96-well plate (*Corning*). At 70% confluence, they were transfected with a mixture of jetPEI (Polyplus transfection) and plasmids as indicated in individual figures, together with the constitutive Renilla luciferase transfection control. Two days after transfection, cells were harvested and lysed with 30 µl of 1×Passive lysis buffer (*Promega*). Firefly luciferase and Renilla luciferase activity was measured using an Orion II microplate reader (*Berthold Technologies*, Pforzheim, Germany). Relative luciferase activity (RLU) was calculated by normalizing each sample's firefly luciferase activity with the constitutive Renilla luciferase activity determined within the same sample. The results represent the mean and standard deviation of four measurements in separate wells. The data are representative of at least three independent experiments.

Chapter I: Section 2

General

Fmoc-O1PenBpy-OH, **1**, was kindly provided by courtesy of David Bouzada[244] and was synthetized according to previously described methods.[245] The bipyridine ligand was synthetized according to previously described methods.[246]

UHPLC-MS characterization of Peptides

brHis$_2$: pABA-DPAA**H**KRA**H**NTEAARRSRARKLQR-CONH$_2$

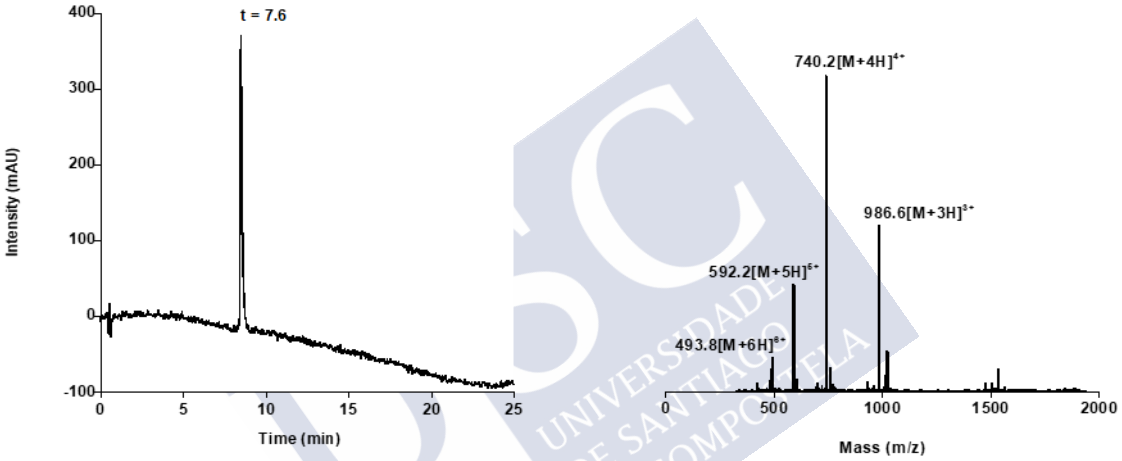


Figure 107. *Left:* HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

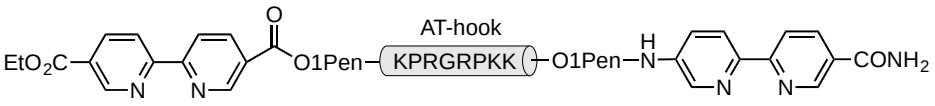
EM–ESI$^+$ (m/z): Calcd. for $C_{123}H_{205}N_{51}O_{35}$: 2956.6. Found: 986.6 $[M+3H]^{3+}$; 740.2 $[M+4H]^{4+}$; 592.2 $[M+5H]^{5+}$; 493.8 $[M+6H]^{6+}$

[244] 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.

[245] I. Gamba, G. Rama, E. Ortega-Carrasco, R. Berardozzi, V. M. Sánchez-Pedregal, L. Di Bari, J.-D. Maréchal, M. E. Vázquez and M. Vázquez-López, *Dalton Trans*. **2016**, *45*, 881-885.

[246] H. Ishida, Y. Maruyama, M. Kyakuno, Y. Kodera, T. Maeda, S. Oishi, *ChemBioChem*. **2006**, *7*, 1567-1570.

Hkbpy$_2$:



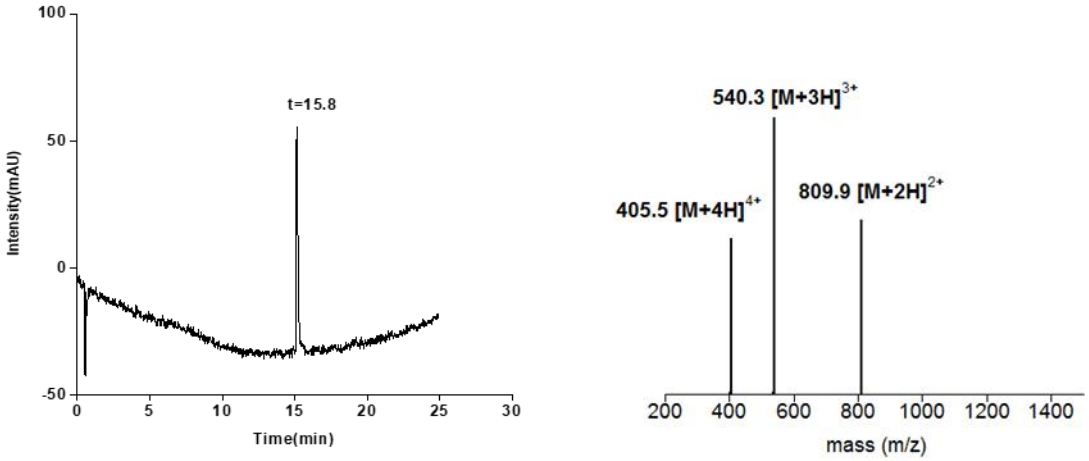


Figure 108. *Left:* HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{75}H_{110}N_{24}O_{17}$: 1618.85. Found: 809.9 $[M+2H]^{2+}$; 540.3 $[M+3H]^{3+}$; 405.5 $[M+4H]^{4+}$.

Bpy^{N}Hk:

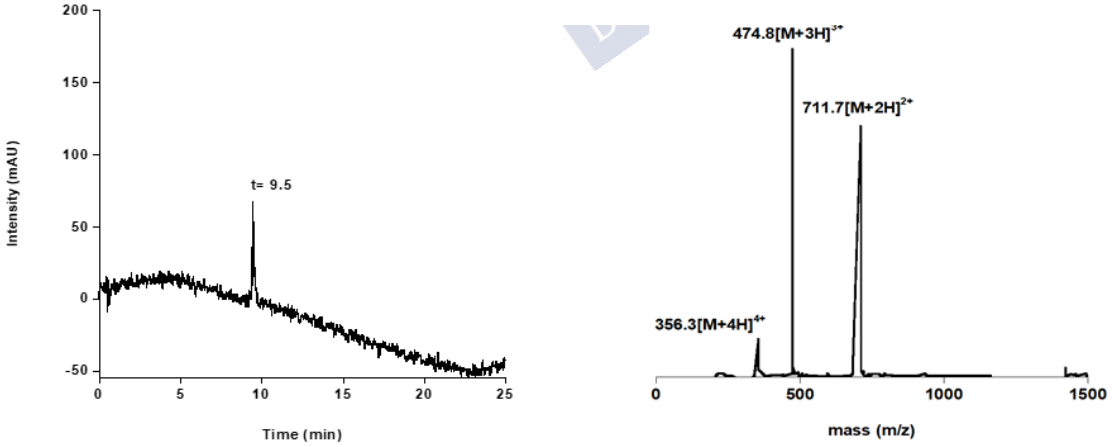
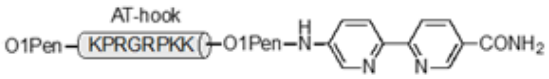


Figure 109. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{64}H_{104}N_{22}O_{15}$: 1421.68. Found: 711.7 $[M+2H]^{2+}$; 474.8 $[M+3H]^{3+}$; 356.3 $[M+4H]^{4+}$.

HkBpy^{C}:



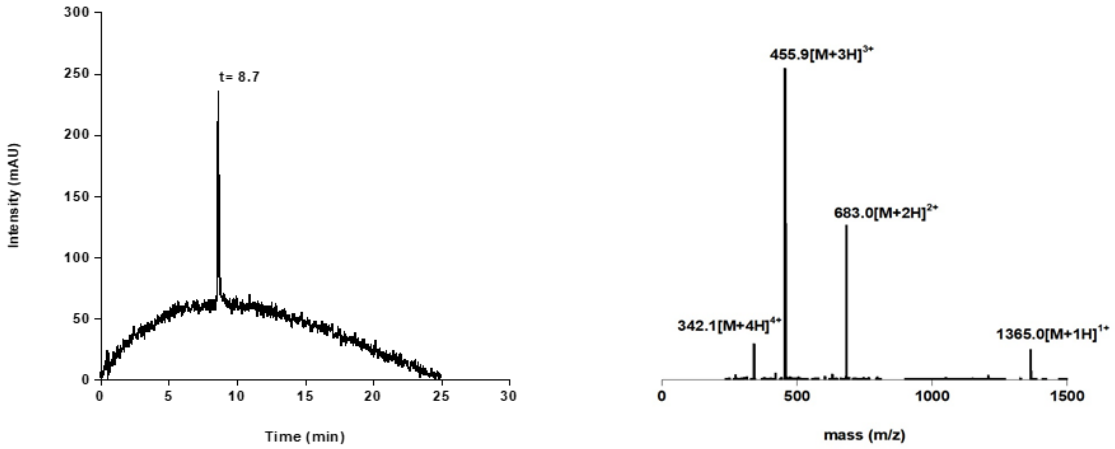


Figure 110. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right:* Mass spectrum of the purified peptide.

EM–ESI^{+} (m/z): Calcd. for $C_{61}H_{101}N_{23}O_{13}$: 1364.63. Found: 1365.0 $[M+1H]^{1+}$; 683.0 $[M+2H]^{2+}$; 455.9 $[M+3H]^{3+}$; 342.1 $[M+4H]^{4+}$.

BrHis: pABA-DPAALKRA**H**NTEAARRSRARKLQR-$CONH_2$

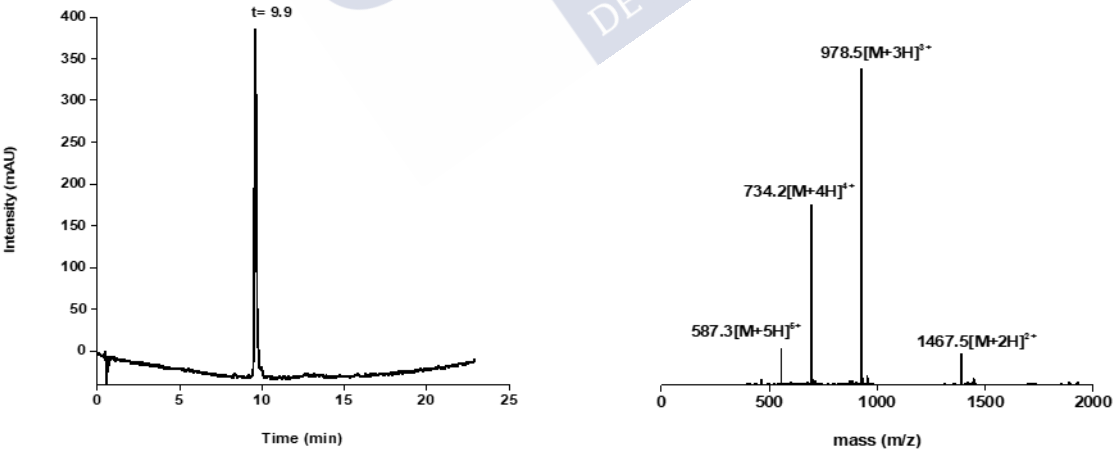


Figure 111. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI+ (m/z): Calcd. for $C_{117}H_{197}N_{45}O_{34}$: 2932.51. Found: 1467.5 $[M+2H]^{2+}$; 978.5 $[M+3H]^{3+}$; 734.2 $[M+4H]^{4+}$; 587.3 $[M+5H]^{5+}$.

Gly_7bpy_2: BPY-GGGGGGGK(BPY)-$CONH_2$

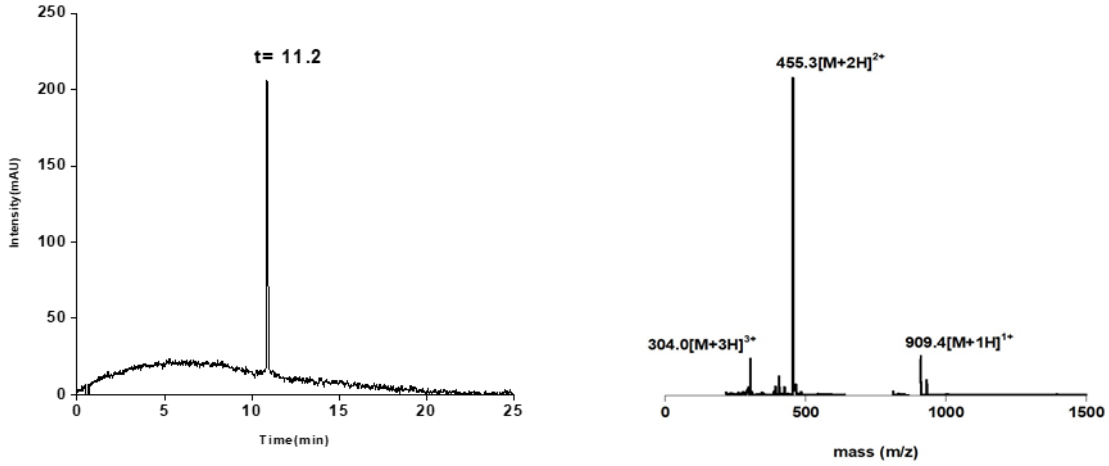


Figure 112. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI^+ (m/z): Calcd. for $C_{42}H_{47}N_{13}O_{11}$: 909.35. Found: 909.4 $[M+1H]^{1+}$; 455.3 $[M+2H]^{2+}$; 304.0 $[M+3H]^{3+}$.

Pro_7bpy_2: BPY-PPPPPPPK(BPY)-$CONH_2$

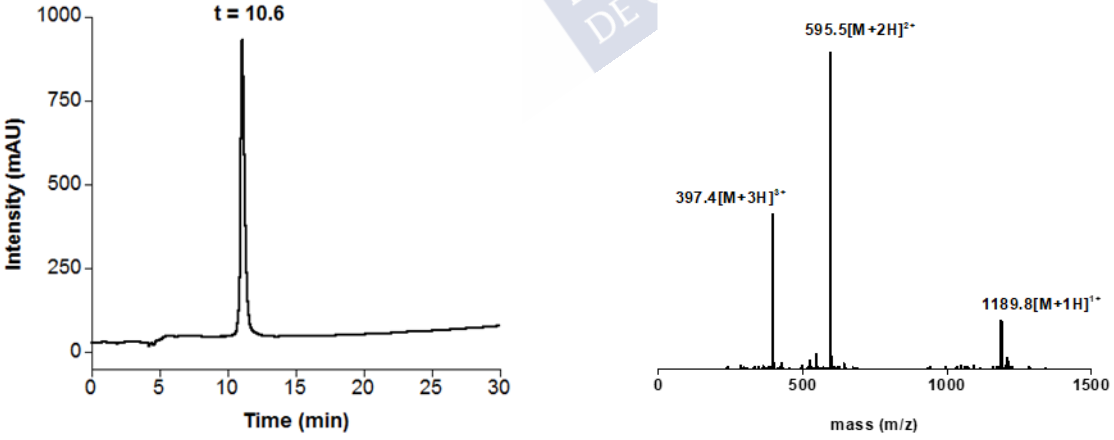


Figure 113. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 75% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI^+ (m/z): Calcd. for $C_{63}H_{75}N_{13}O_{11}$: 1189.57. Found: 1189.8 $[M+1H]^{1+}$; 595.5 $[M+2H]^{2+}$; 397.4 $[M+3H]^{3+}$.

Oligonucleotide sequence

Double stranded (only one strand is shown) oligonucleotides used for EMSA experiments:

CAT•AT•CAT	5´- GCGAG TCAT AATT TCAT AGGCG -3´
ctc•AT•gag	5´- CGCGCTCTC AATT GAGAGCTC-3´
CAT•AT•gag	5´- CGCAG TCAT AATT GAGAGCGC -3´
CAT•cg•CAT	5´- GCGAG TCAT CAGG TCAT AGGCG -3´
CAT•TAT•CAT	5´- GCGAG TCAT TAATT TCAT AGGCG -3´
*TMR-**CAT•AT•CAT***	5´– (TMR)-GCGAG TCAT AATT TCAT AGGCG–3

Chapter I: Section 3

Reagents

$PdCl_2$(en), $Pd(PPh_3)_2Cl_2$, $PdCl_2$(bpy), K_2PdCl_6, K_2PdCl_4, *trans*-Pd(dppe)Cl_2, $PtCl_2$(bpy), $PtCl_2$(en), K_2PtCl_4, $NiCl_2$, $ZnSO_4$, $CuSO_4$, $Co(ClO_4)_2$ and $Fe(NH4)_2(SO_4)_2$ were purchased from *Aldrich*. $CuCl_2$(bpy),[247] $NiCl_2$(bpy),[248] $Pd(bpy)(NO_3)_2$,[249] [$AuCl_2$(en)]Cl,[250] and [$AuCl_2$(bpy)]PF_6,[251] were synthesized according to the literature procedures.

UHPLC-MS characterization of Peptides

The characterization for **brHis$_2$** and **brHis** can find in Chapter I: section 2.

Br: pABA-DPAALKRARNTEAARRSRARKLQR-$CONH_2$

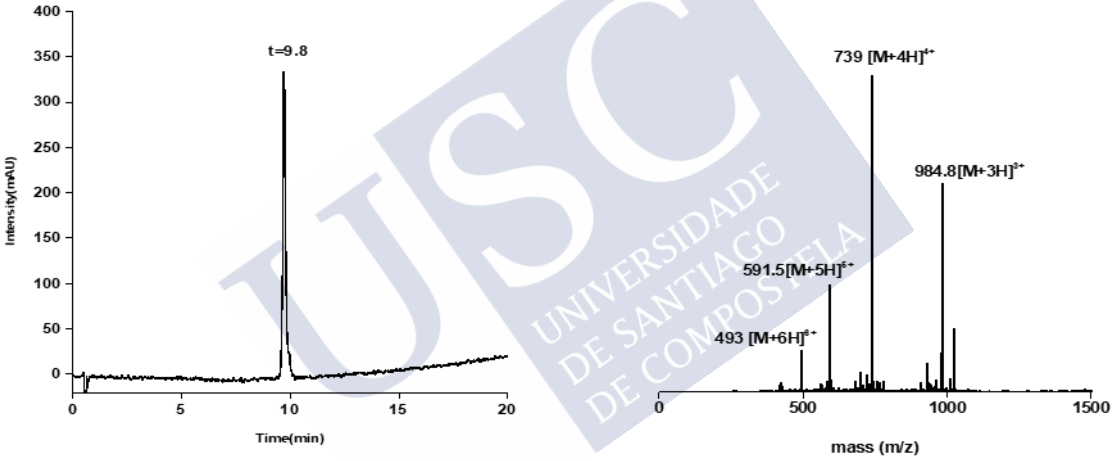


Figure 114. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI^+ (m/z): Calcd. for $C_{123}H_{214}N_{50}O_{35}$: 2951.65. Found: 984.8 $[M+3H]^{3+}$; 739 $[M+4H]^{4+}$; 591.5 $[M+5H]^{5+}$; 493 $[M+6H]^{6+}$

[247] D. A. Martins, L. R. Gouvea, D. da Gama Jean Batista, P. Bernardino da Silva, S. R. W. Louro, M. d. N. C. Soeiro, L. R. Teixeira, *Biometals* **2012**, *25*, 951-960.

[248] B. Brewer, N. R. Brooks, S. Abdul-Halim, A. G. Sykes, *J. Chem. Crystallogr*. **2003**, *33*, 9, 651-662.

[249] R. A. Adrian, S. Zhu, K. K. Klausmeyer, J. A. Walmsley, *Inorg. Chem. Comm*. **2007**, *10*, 1527-1530.

[250] D. M. Motley, Synthesis and characterization of square-planar gold(III) complexes with various ligands, Dissertation for Master of Science in Chemistry, The University of Texas at San Antonio, **2008**.

[251] M. A. Cinellu, G. Minghetti, M. V. Pinna, S. Stoccoro, A. Zucca, M. Manassero, *J. Chem. Soc., Dalton Trans*. **2000**, 1261-1265.

TMR-**brHis$_2$:** TMR-Ahx-DPAA**H**KRA**H**NTEAARRSRARKLQR-CONH$_2$

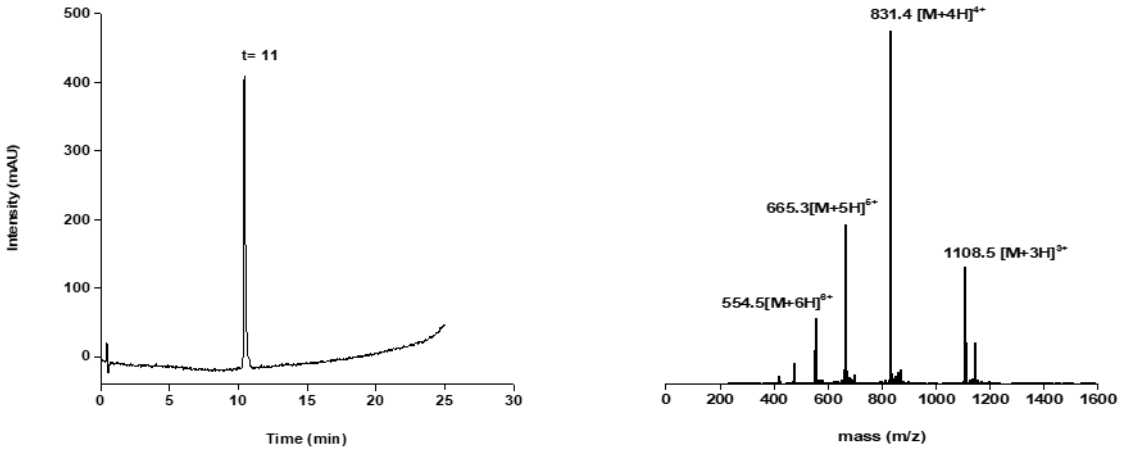


Figure 115. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{146}H_{224}N_{53}O_{38}$: 3324. Found: 1108.5 $[M+3H]^{3+}$; 831.4 $[M+4H]^{4+}$; 665.3 $[M+5H]^{5+}$; 554.5 $[M+6H]^{6+}$.

TMR-**br**: TMR-Ahx-DPAALKRARNTEAARRSRARKLQR-CONH$_2$

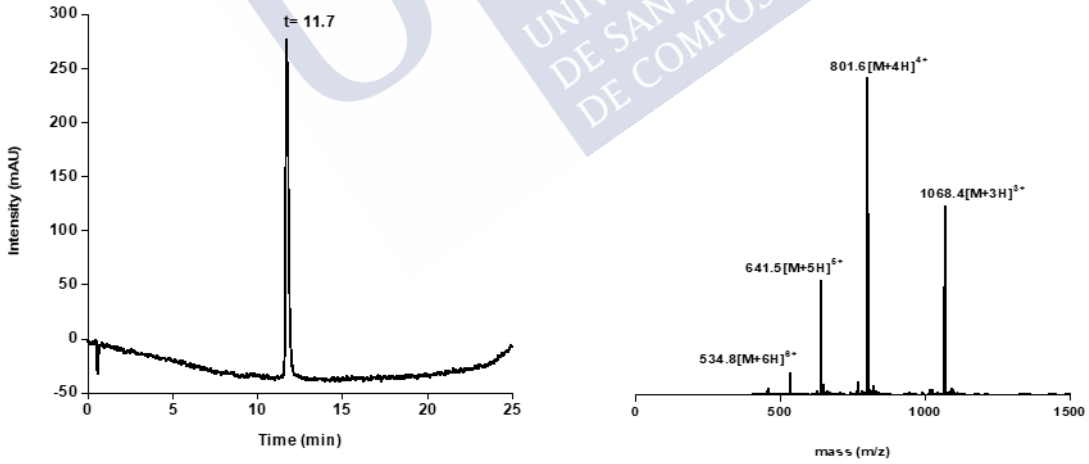


Figure 116. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{146}H_{230}N_{52}O_{38}$: 3317. Found: 1106.6 $[M+3H]^{3+}$; 830.2 $[M+4H]^{4+}$; 664.4 $[M+5H]^{5+}$; 553.8 $[M+6H]^{6+}$.

TMR-**brHis:** TMR-Ahx-DPAALKRA**H**NTEAARRSRARKLQR-$CONH_2$

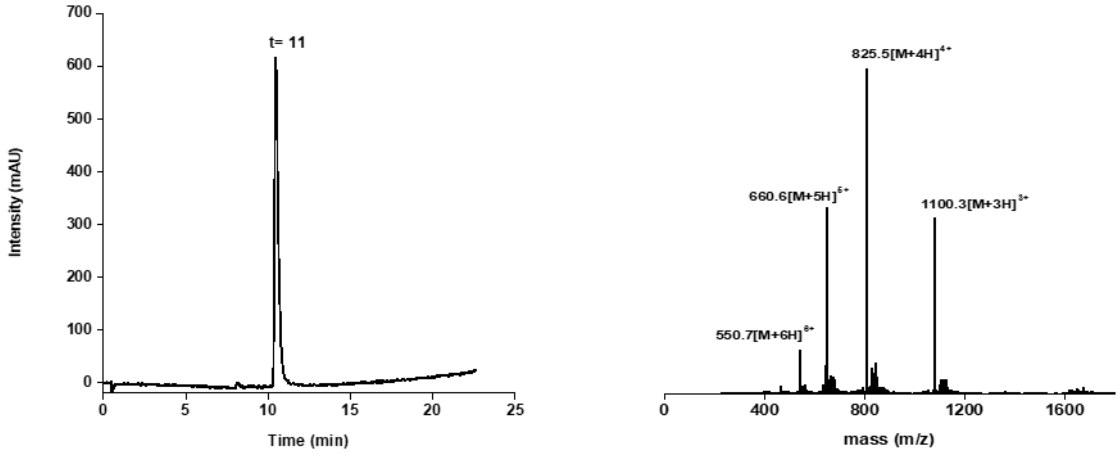


Figure 117. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI^+ (m/z): Calcd. For $C_{146}H_{226}N_{51}O_{38}$: 3298. Found: 1100.3 $[M+3H]^{3+}$; 825.5 $[M+4H]^{4+}$; 660.6$[M+5H]^{5+}$; 550.7 $[M+6H]^{6+}$.

br(S): pABA-DPAALKRARNTEAARRSRARKLQGGC-$CONH_2$

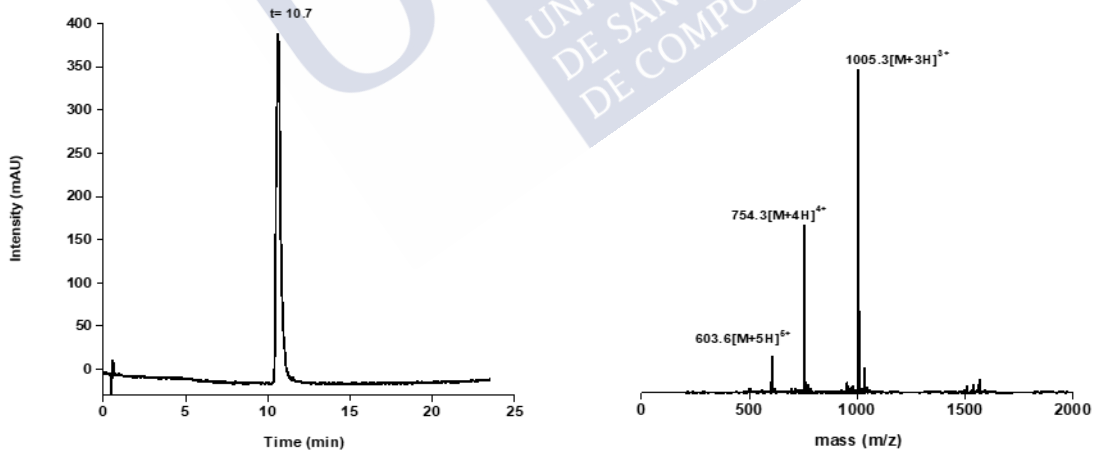


Figure 118. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI^+ (m/z): Calcd. For $C_{124}H_{215}N_{50}O_{36}S$: 3012.6. Found: 1005.3 $[M+3H]^{3+}$; 754.3 $[M+4H]^{4+}$; 603.6 $[M+5H]^{5+}$.

Synthesis of the disulfide dimer br$_2$(SS). The monomeric thiol-containing peptide **br(S)** was dissolved in 100 mM phosphate buffer (pH 8.3, 2 mM), and an air current was passed through the solution. Ellmann's reagent[252] (DTNB, 0.8 mM, 0.4 eq.) was added, and the resulting mixture was stirred for 1.5 h 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 °C, after characterization.

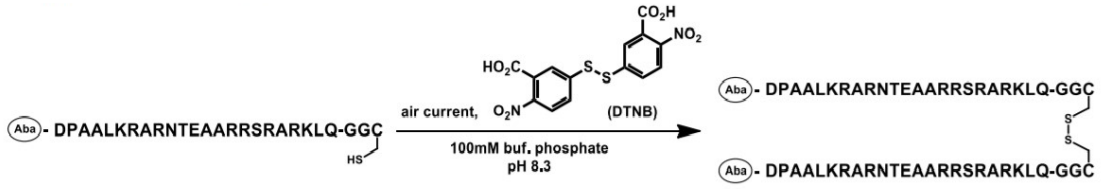


Figure 119. Synthesis of the disulfide dimer **br$_2$(SS).**

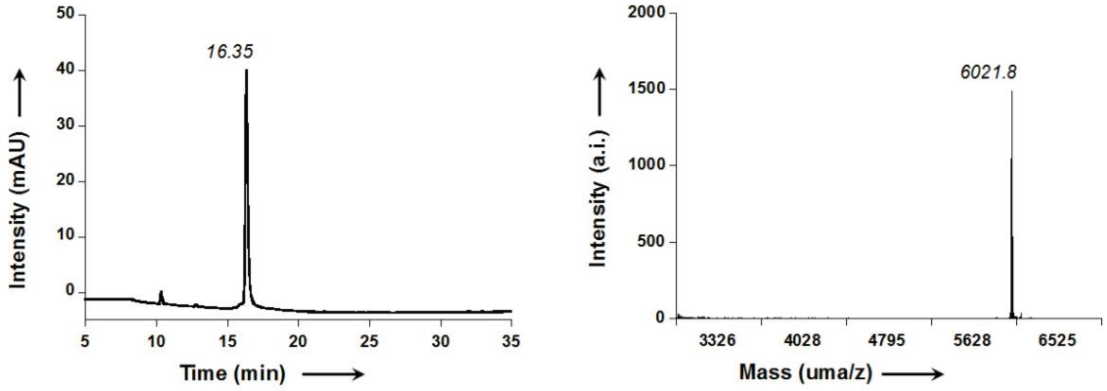


Figure 120. *Left: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. Right: Mass spectrum of the peptide br$_2$(SS) by MALDI-MS.*

EM–ESI$^+$ (m/z): Calcd. For $C_{248}H_{427}N_{100}O_{72}S_2$: 6022.2. Found: 6021.8.

[252] Hermanson, G. T. *Bioconjugate Techniques*. Academic Press; 2nd Ed, **2008**.

UHPLC-MS of Palladium Complexes

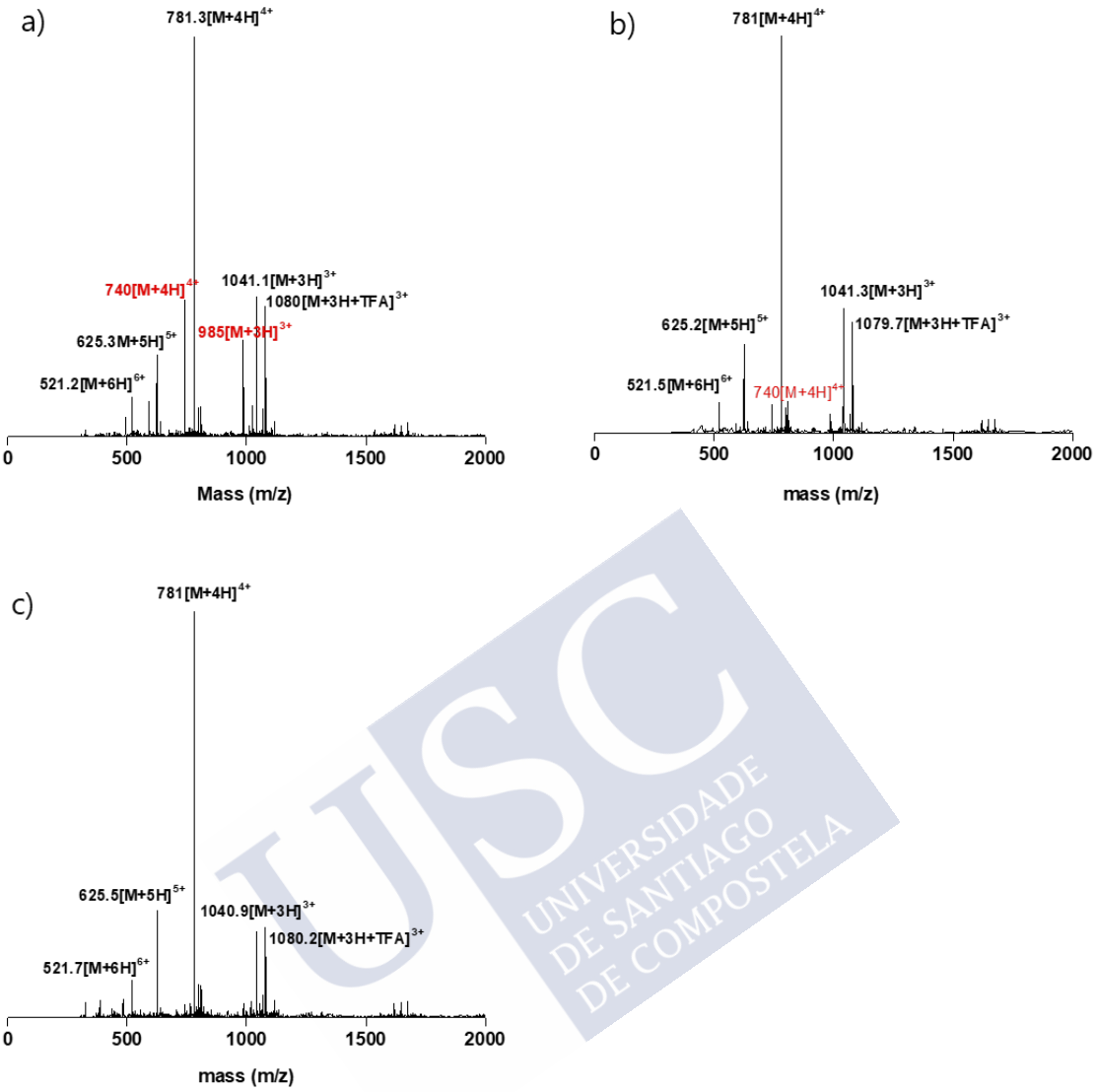


Figure 121. HPLC-MS experiments of **brHis$_2$** + $PdCl_2$(en). A solution containing **brHis$_2$** and $PdCl_2$(en) (1:1) in water was prepared and measured at different time points. a) t = 0 min. after incubation; b) t = 10 min. after incubation; c) t = 20 min. after incubation.

EM–ESI$^+$ (m/z): Calcd. for the complex $C_{125}H_{211}N_{53}O_{35}Pd$: 3122; Calcd. for the peptide $C_{123}H_{205}N_{51}O_{35}$: 2956.

Found a): 1080.2 $[M+4H+TFA]^{3+}$; 1040.9 $[M+3H]^{3+}$; 985 $[M+3H]^{3+}$; 781 $[M+4H]^{4+}$; 740 $[M+4H]^{4+}$; 625.5 $[M+5H]5^+$; 521.7 $[M+6H]^{6+}$ (The red colour corresponds to the peptide´s mass while the black colour it's the complex mass).

Found b): 1079.7 $[M+4H+TFA]^{3+}$; 1041.3 $[M+3H]^{3+}$; 781 $[M+4H]^{4+}$; 740 $[M+4H]^{4+}$;625.2 $[M+5H]5^+$; 521.5 $[M+6H]^{6+}$ (The red colour corresponds to the peptide´s mass while the black colour it's the complex mass).

Found c): 1080.2 $[M+4H+TFA]^{3+}$; 1040.9 $[M+3H]^{3+}$; 781 $[M+4H]^{4+}$; 625.5 $[M+5H]5^+$; 521.7 $[M+6H]^{6+}$

The patterns show the mass for the complex between the peptide and $Pd(en)^{2+}$.

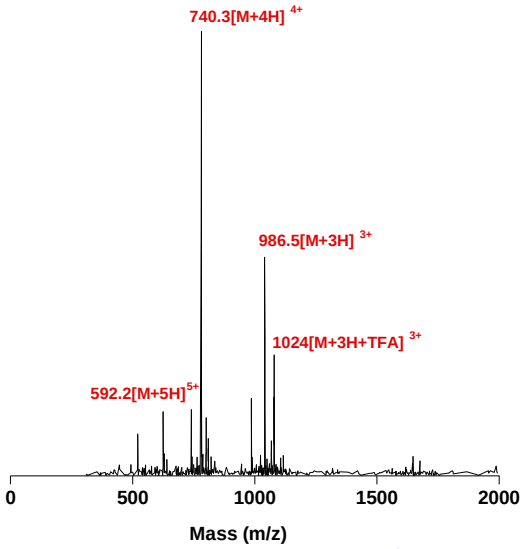


Figure 122. HPLC-MS spectrum of **brHis₂** + $PtCl_2(en)$. A solution containing **brHis₂** and $PtCl_2(en)$ (1:1) in water was prepared and measured at different incubation times (0, 10 and 20 min).

EM–ESI^+ (m/z): Calcd. for the complex $C_{125}H_{213}N_{53}O_{35}Pt$: 3212; Calcd. for the peptide $C_{123}H_{205}N_{51}O_{35}$: 2956.

Found (the same for the different times): 1024 $[M+4H+TFA]^{3+}$; 986.6 $[M+3H]^{3+}$; 740.2 $[M+4H]^{4+}$; 592.2 $[M+5H]^{5+}$ (The red colour corresponds to the peptide´s mass while the black colour it's the complex mass).

Oligonucleotide sequence

Double stranded (only one strand is shown) oligonucleotides used for EMSA experiments:

GTCAT	5'-CGC GTCAT AATTGAGAGCGC-3'
MUT	5'-CGC GTGAT AATTGAGAGCGC-3'
FULL MUT	5'-CGACGCTCTCAAATTCCGTC -3'
TMR-GTCAT	5'-(TMR)-CGC GTCAT AATTGAGAGCGC-3'
AP-1	5'-TGGAG ATGAcgTCAT CTCGT-3'

Fluorescence Anisotropy

Tritation of $PdCl_2(en)$ with TMR-brHis₂

5 µL of a water solution of TMR-**brHis₂** (5 µM) were added to 995 µL of Tris-HCl buffer 20 mM pH 7.5 100 mM NaCl (final concentration 25nM), and the anisotropy was measured at 559 nm, at r.t. Aliquots of a stock solution in water of $PdCl_2(en)$ were successively added to this solution, starting with 1 nM, and the anisotropic value was recorded after each addition. Experimental data correspond to the mean of three independent titration experiments.

UV-Vis Spectra

Measurements were performed in a 2 mm cuvette at 25 ºC. Samples contained 10 mM phosphate buffer pH 7.5 and 100 mM of NaCl, 5 µM peptide, 5 µM of corresponding dsDNA (when present), and one equivalent of either metal complex PdCl2(bpy) or $NiCl_2$(bpy).

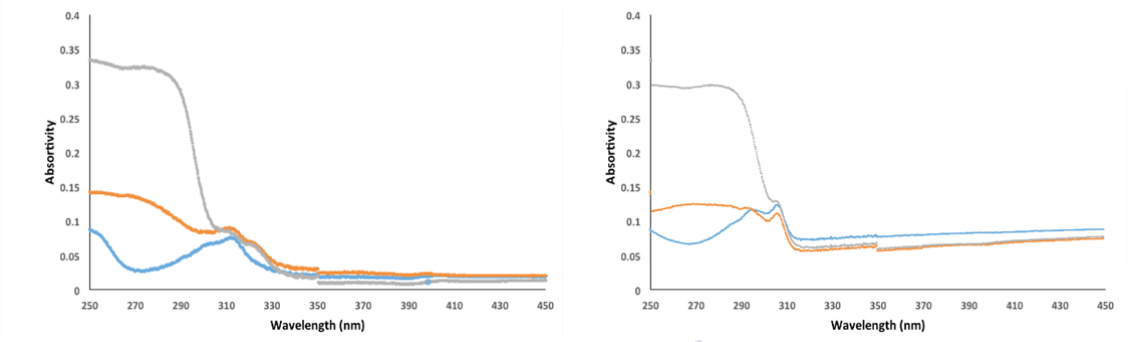


Figure 123. UV-Vis experiments for compounds $PdCl_2$(bpy) (left) and $NiCl_2$(bpy) (right). Blue line: just the metal complex; Orange line: after addition of the peptide **brHis$_2$**. Grey lines: after addition of ***GTCAT***. A new band can be observed when the $PdCl_2$(bpy) is mixed with the peptide; this band shows the presence of the bipy chromophore and is not observed in the case of $NiCl_2$(bpy).

Cell internalization studies

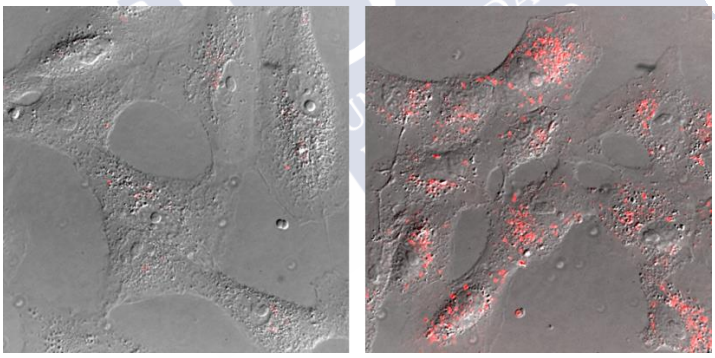


Figure 124. Fluorescence micrographies of Vero cells, brightfield images are superimposed to the red emission channel; after incubation with 5 µM TMR-**brHis$_2$** for 30 min at 37 ºC (left) and 5 µM TMR-**brHis$_2$** premixed with 5 µM $PdCl_2$(en) for 10 min, and then incubated with the cells for 30 min at 37ºC (right).

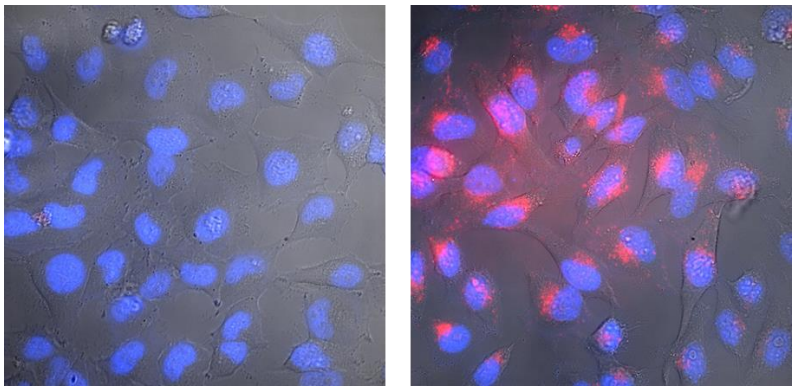


Figure 125. Fluorescence micrographies of Hela cells, brightfield images are superimposed to the red and blue emission channel after incubation with 5 μM TMR-**brHis$_2$** co-incubated with 10 μM Hoechst, for 30 min at 37 ºC. *Left*: without further additives. *Right*: premixed with 5 μM $PdCl_2$(en).

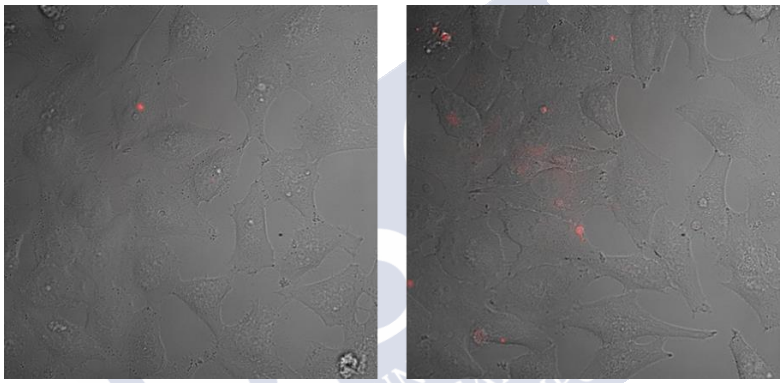


Figure 126. Fluorescence micrographies of Hela cells, brightfield images are superimposed to the red emission channel. *Left:* 5 μM **brHis$_2$** premixed with 5 μM $PdCl_2$(en) for 10 min and incubated with the cells for 30 min at 37 ºC. Then, TMR-**br** (5 μM) was added, and incubated for another 30 min. *Right:* mixture of 5 μM **brHis$_2$**, 5 μM $PdCl_2$(en), and 5 μM TMR-**br** incubated for 10 minutes and then for another 30 min at 37 ºC in the presence of the cells.

Dynamic Light Scattering

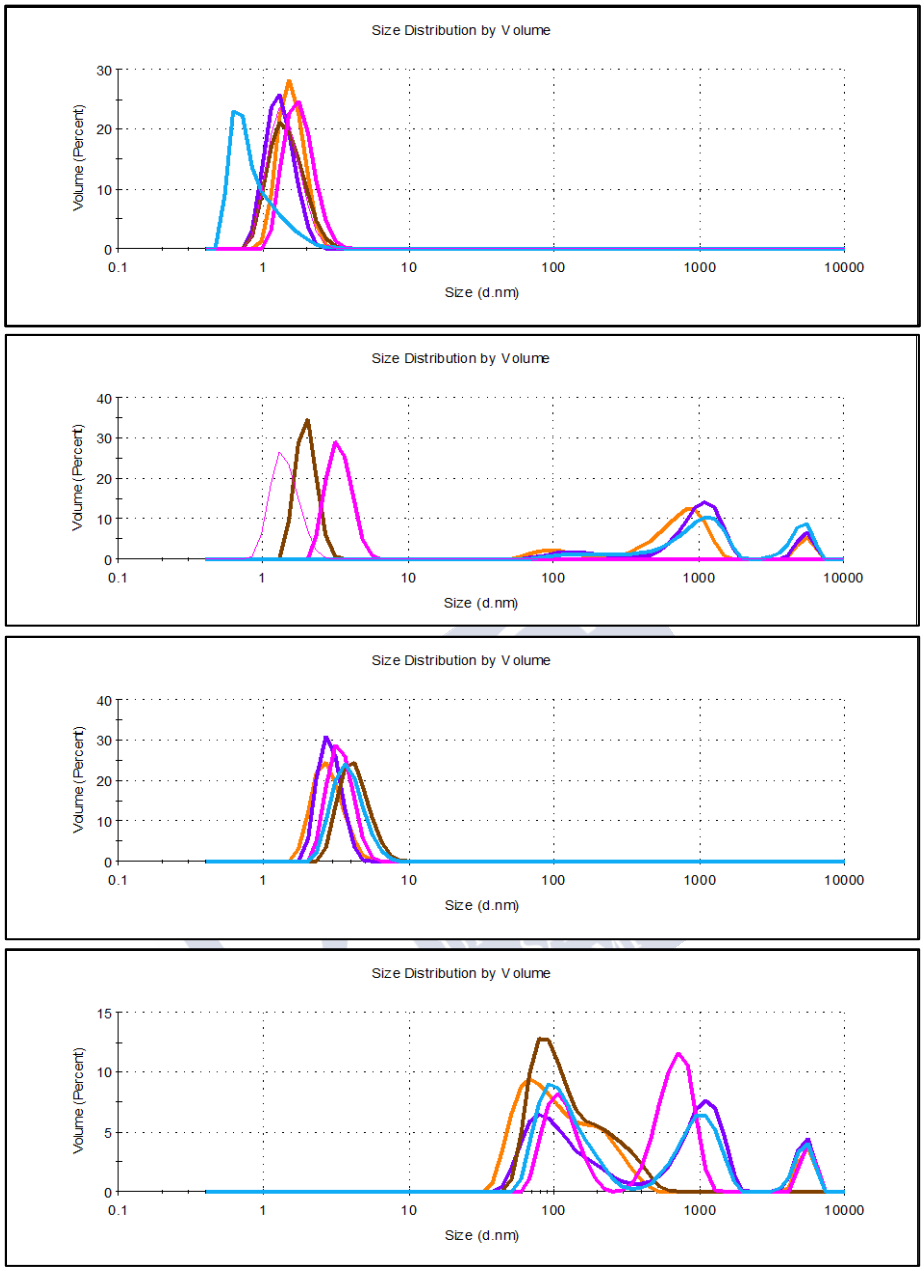


Figure 127. Peptide aggregation monitored by dynamic light scattering. Samples contained: a) 2 mg/ml of **brHis$_2$** in water; b) 2 mg/ml of **brHis$_2$** in water + $PdCl_2$(en) (1:1), without any incubation; c) 2 mg/ml of **brHis$_2$** in water + $PdCl_2$(en) (1:1), pre-mixed and incubated 20 min before measuring; d) control of $PdCl_2$(en) (2mg/ml). Traces in a) at size c.a. 1 nm, indicate that the peptide does not aggregate; although the mixture with $PdCl_2$(en) showed in b) initially appears to contain some aggregates, those are already present in the $PdCl_2$(en) solution in d), and disappear after a few minutes, when the complexation takes place, as shown in trace c).

The same experiments were performed at 5 μM of **brHis$_2$** with similar results.

Chapter II

General

$PdCl_2$(en) were purchased from *Aldrich*. The antibodies ERK2 and phospho-ERK1/2 was kindly provided by Dr. Costoya.[253]

UHPLC-MS characterization of Peptides

αH-His$_2$: TMR-Ahx-R-R-F-F-G-I-**H**-L-T-N-**H**-L-K-T-E-E-G-N

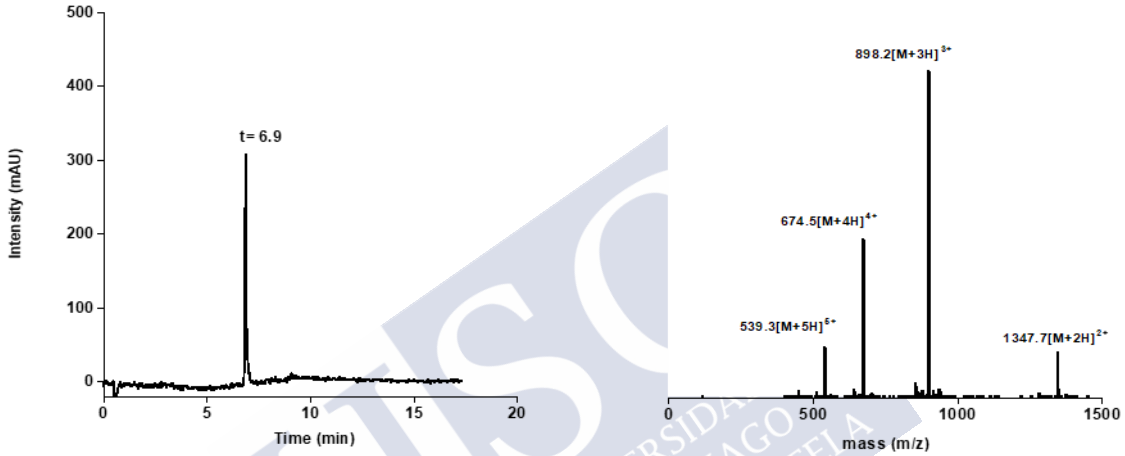


Figure 128. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. for $C_{127}H_{181}N_{35}O_{31}$: 2692.4. Found: 1347.7 $[M+2H]^{2+}$; 898.2 $[M+3H]^{3+}$; 674.5 $[M+4H]^{4+}$; 539.3 $[M+5H]^{5+}$

UHPLC-MS of the metallopeptide

The peptide was mixed with $PdCl_2$(en) (1:1) in water for 10 min, and the resulting crude was analyzed by UHPLC-MS in order to corroborate the formation of metallopeptide.

[253] CiMUS- Centro Singular de Investigación en Medicina Molecular y Enfermedades Crónicas. Avenida Barcelona, s/n, 15782, Santiago de Compostela. Spain.

αH-His$_2$[Pd]

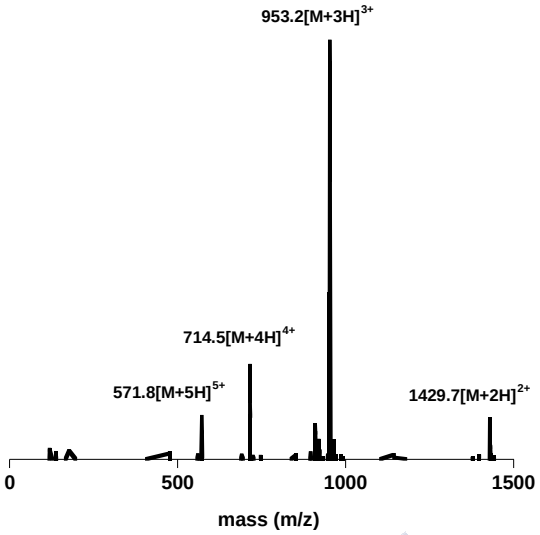


Figure 129. Mass spectrum of the crude resulting from mixing **αH-His$_2$** + $PdCl_2$(en) (1:1), consistent with the formation of a complex containing Pd and the ethylenediamine ligand.

EM–ESI$^+$ (m/z): Calcd. for the metallopeptide $C_{129}H_{189}N_{37}O_{31}Pd$: 2858.34. Found: 1429.7 $[M+2H]^{2+}$; 953.2 $[M+3H]^{3+}$; 714.5 $[M+4H]^{4+}$; 571.8 $[M+5H]^{5+}$

Circular Dichroism Measurements (CD)

The samples contained 10 mM phosphate buffer pH 7.5 and 100 mM of NaCl, 10% 2,2,2-Trifluoroethanol (TFE), 20 μM peptide and different equivalents of metal complex (when present).

Protein Expression and Purification

The constructs *pDNA2.0 6H-TEV-KRAS*, *pDNA2.0 6H-TEV-KRAS G12V* and *pDNA2.0 6H-FLAG-TEV-KRAS G12C* were purchased from *Addgene*.

BL21 cells (*Invitrogen*) were transform with the constructs and grown in Luria broth (LB) to OD_{600} ~0.7 and induced with 250 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 h at 16 °C. Cells were pelleted and resuspended in lysis buffer [20 mM sodium phosphate (pH 8.0), 500 mM NaCl, 10 mM imidazole, 1 mM 2-mercaptoethanol (BME), 5% (vol/vol) glycerol]. After centrifugation, the pellet was frozen until use. Then, protease inhibitor (EDTA free) and lysozyme (1mg/mL) were added and the pellet was sonicated for 6´ (35"on /10"off).

Protein was purified over an IMAC (immobilized metal affinity chromatography) following HisPur™ Ni-NTA protocols (*Thermo fisher*). Protein was concentrated in a 10-kDa Amicon Ultra-15 (*Millipore*) aliquoted with 15% glycerol, and then flash-frozen and stored at -20ºC.

Fluorescence Anisotropy

αH-His$_2$ or **αH-His$_2$**[Pd] (3 µL, 5 µM) was added to 995 µL of Tris-HCl buffer 20 mM [pH 7.5], 100 mM NaCl (final concentration 15nM), and the anisotropy was measured. Aliquots of a stock solution of the corresponding purified protein were successively added to this solution, and the anisotropic value was recorded after each addition. Experimental data correspond to the mean of three independent titration experiments and are representative of at least three biological replicates performed with independent preparations of recombinant proteins. All experiments carried out at 25 °C.

Before use the corresponding purified protein, the glycerol was removed using a Zeba™ Spin Desalting Columns (*Thermo fisher*) that had been equilibrated with anisotropy buffer (Tris-HCl buffer 20 mM [pH 7.5], 100 mM NaCl).

Switch experiment: After performing the titration following the above steps and achieve a saturated **αH-His$_2$**[Pd]/KRAS complex, 50 eq. of DEDTC (respect to the metallopeptide) was added to the solution and the anisotropic value was recorded. After that, 50 eq. of $PdCl_2$(en) was added (respect to the metallopeptide) and the anisotropy was recorded again.

Nucleotide Release Experiment

Loading of fluorescent-labeled GDP (mantGTP) to KRAS^{wt} was conducted following previous reports.[254,255] Purified KRAS^{wt} was buffer-exchanged in Zeba™ Spin Desalting Columns (*Thermo fisher*) in loading buffer (20 mM Tris-HCl [pH 7.5], 50 mM NaCl, 4 mM EDTA and 1 mM DTT). The resulted eluted KRAS^{wt}, after measure their concentration, was incubated with 10-fold molar excess of mantGTP (*Abcam*) for 1.5 h at 20 °C in the dark. Reactions were supplemented with 10 mM $MgCl_2$ and incubated for 1h on ice. Free nucleotide was removed using a Zeba™ Spin Desalting Columns (*Thermo fisher*) that had been equilibrated with the reaction buffer (20 mM Tris-HCL [pH 7.5], 50 mM NaCl, 1 mM MgCl2 and 1 mM DTT).

The effect of peptides on the intrinsic rate of nucleotide release was monitored using the decrease in fluorescence with time as mantGTP dissociates from KRAS in a 100 µL reaction mixture (96-well plate) of 1 µM KRAS^{wt}/mantGPT complex and 10 µM of **αH-His$_2$** or **αH-His$_2$**[Pd]. Fluorescence was excited at 370 nm and emission was monitored at 430 nm during 30 min, using a *Tecan Infinite M Plex* plate reader. The data was processed using the program *Prism*.

[254] S. M. Margarit, H. Sondermann, B. E. Hall, B. Nagar, A. Hoelz, M. Pirruccello, D. Bar-Sagi, J. Kuriyan, *Cell.* **2003**, *112*, 685-695.

[255] M. McCarthy, C. Pagba, P. Prakash, A. Naji, D. van der Hoeven, H. Liang, A. Gupta, Y. Zhou, K. J. Cho, J. Hancock, A. A. Gorfe, *bioRxiv*. **2018**, 440487.

Inhibition of ERK1/2

A549 cells were seeded on glass-bottom plates (150.000cel/ml) 48 h before treatment. Culture medium was removed, the cells was serum-starved and **αH-His$_2$**, **αH-His$_2$**[Pd] or $PdCl_2$(en) were added, at different concentrations, and incubated for 4 hours. After two washes with PBS, the cells were lysed in Laemmli buffer containing Tris pH 6,8 1M, Glycerol, SDS, Bromophenol Blue and β-mercaptoethanol and heated at 95ºC for 10 min. The samples were separated by SDS-12.5% polyacrylamide gel and transferred to nitrocellulose membrane, to probed by Western blotting. Levels of total ERK2 and phosphorylated ERK were detected with anti-ERK2 (1:1000) and phospho-ERK1/2 (1:1000) antibodies, respectively. The visualization was performed with ChemiDoc^{TM} MP imaging system by *Biorad*.

Chapter III

General

Procedures for the synthesis of precursors were performed under an atmosphere of dry nitrogen using vacuum-line and standard Schlenk techniques.

Compounds **1**,[256] **2**,[256] **3**,[257] **C**,[258] **I**[259] are known compounds and were synthesized according to the literature. Their ^{1}H and ^{13}C NMR data were in complete agreement with the reported values.

All catalytic reactions were carried out without particular precautions to extrude moisture or oxygen and open to air unless otherwise stated. Reaction mixtures were stirred using Teflon-coated magnetic stir bars. The abbreviation "r.t." refers to reactions carried out approximately at 23 °C. Temperature was maintained using Thermowatch-controlled heating blocks. Thin-layer chromatography (TLC) was performed on silica gel plates (*Merck* 60 silica gel F_{254}) and components were visualized by observation under UV light and / or by treating the plates with $KMnO_4$ or *p*-anisaldehyde followed by heating. Flash chromatography was carried out on silica gel (*Merck* Geduran Su 60, 40 - 63 µM silica gel, normal phase). Dryings were performed with anhydrous $MgSO_4$.

Concentration refers to the removal of volatile solvents via distillation using a rotary evaporator *Büchi R-210* equipped with a thermostated bath *B-491*, a vacuum regulator *V-850*, followed by residual solvent removal under high vacuum.

Mass spectra of compounds were acquired using IT-MS Bruker AmaZon SL and also using electrospray (ESI). LC-MS analysis were carried out using Bruker AmaZon IT-MS with C18 column.

UHPLC-MS characterization of Peptides

TMR-**brHis$_2$**, **brHis$_2$**, **brHis** and **br** were previously characterized in Chapter I.

TMR-**brHis$_2$(A238T)**: TMR-Ahx-DPAA**H**KRA**H**NTE**T**ARRSRARKLQR-$CONH_2$

[256] T. Gao, P. Xu, M. Liu, A. Bi, P. Hu, B. Ye, W. Wang, W. Zheng, *Chem. Asian. J.* **2015**, *10*, 1142-1145.

[257] J. Clavadetscher, E. Indrigo, S. V. Chankeshwara, A. Lilienkampf, M. Bradley, *Angew. Chem. Int. Ed.* **2017**, *56*, 6864-6868.

[258] Lebedeva, M. A.; Chamberlain, T. W.; Scattergood, P. A.; Delor, M.; Sazanovich, I. V.; Davies, E. S.; Suyetin, M.; Besley, E.; Schröder, M.; Weinstein, J. A. *Chem. Sci.* **2016**, *7*, 5908-5921.

[259] F. Turtaut, S. Ouahrani-Bettache, J. -L. Montero, S. Köhler, J.-Y. Winum, *Med. Chem. Commun.* **2011**, *2*, 995.

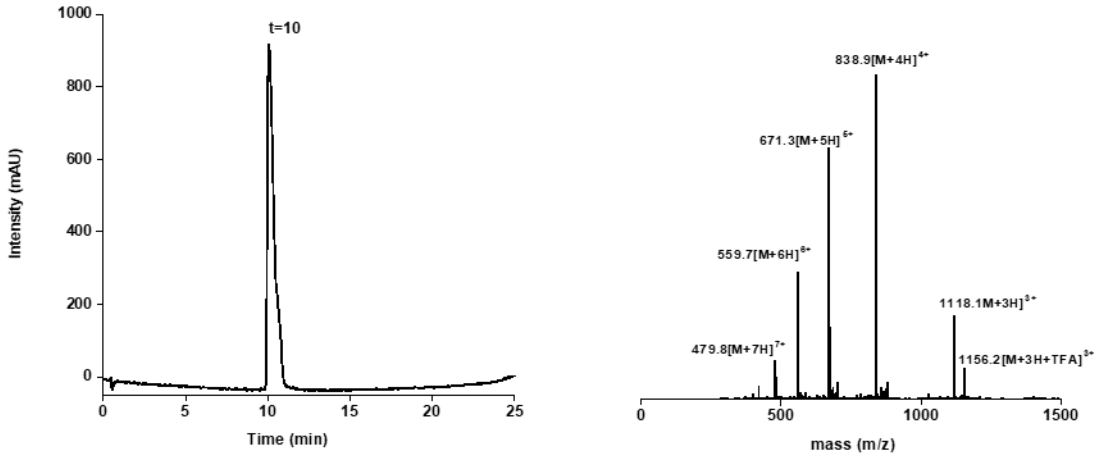


Figure 130. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{147}H_{226}N_{53}O_{39}$: 3354. Found: 1156.2[M+4H+TFA]$^{3+}$; 1118.1[M+3H]$^{3+}$; 838.9 [M+4H]$^{4+}$; 671.3[M+5H]$^{5+}$; 559.7 [M+6H]$^{6+}$; 479.8 [M+7H]$^{7+}$

TMR-**brHis$_2$(T236A, A238T)**: TMR-Ahx-DPAA**H**KRA**H**N**A**E**T**ARRSRARKLQR-CONH$_2$

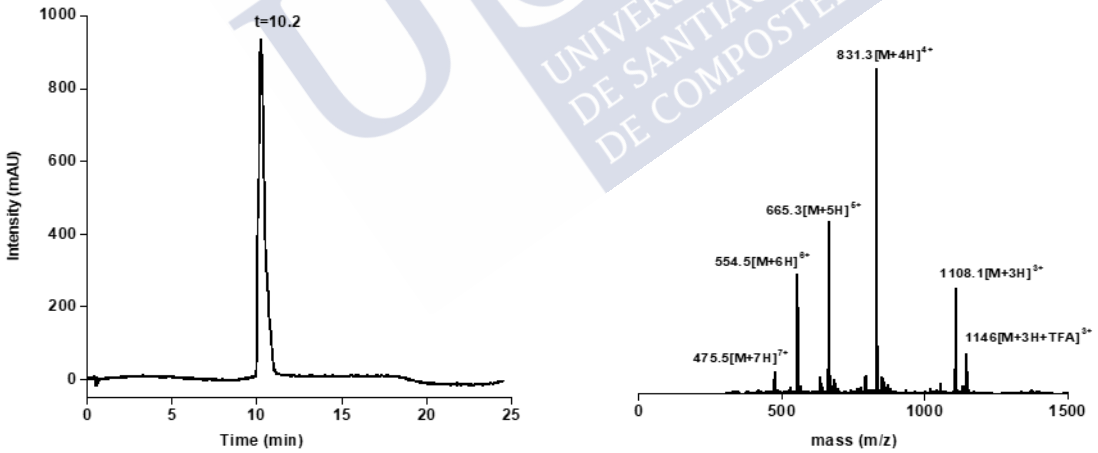


Figure 131. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{146}H_{224}N_{53}O_{38}$: 3324. Found: 1146[M+4H+TFA]$^{3+}$; 1108.1 [M+3H]$^{3+}$; 831.3 [M+4H]$^{4+}$; 665.3 [M+5H]$^{5+}$; 554.5 [M+6H]$^{6+}$; 475.5 [M+7H]$^{7+}$

TMR-**brHis$_2$(Q248A, L247A)**: TMR-Ahx-DPAA**H**KRA**H**NTEAARRSRARK**AA**R-CONH$_2$

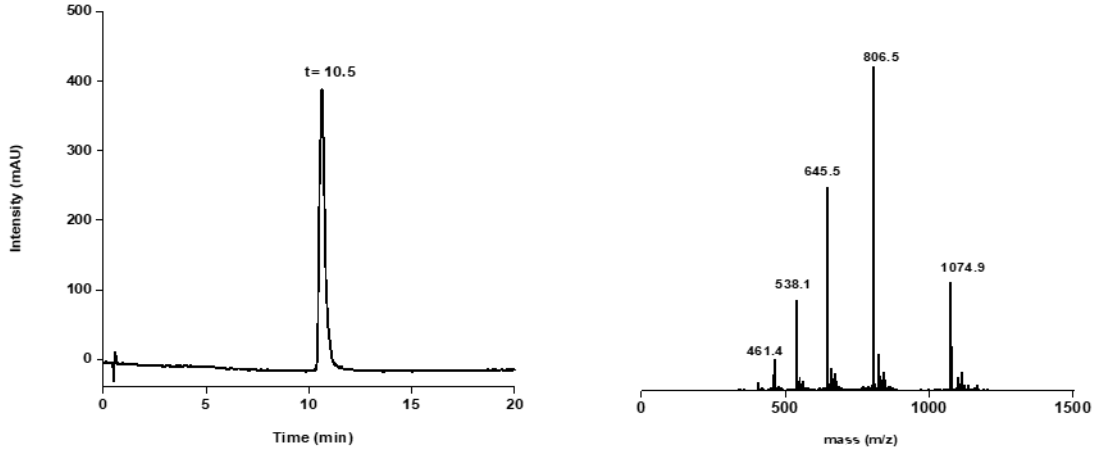


Figure 132. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide**.**

EM–ESI$^+$ (m/z): Calcd. For $C_{141}H_{215}N_{52}O_{37}$: 3222. Found: 1074.9 $[M+3H]^{3+}$; 806.5 $[M+4H]^{4+}$; 645.5 $[M+5H]^{5+}$; 538.1 $[M+6H]^{6+}$; 461.4 $[M+7H]^{7+}$

TMR-**brHis$_2$(R249A, R245S):** TMR-Ahx-DPAA**H**KRA**H**NTEAARRSRA**S**KLQ**A**-CONH$_2$

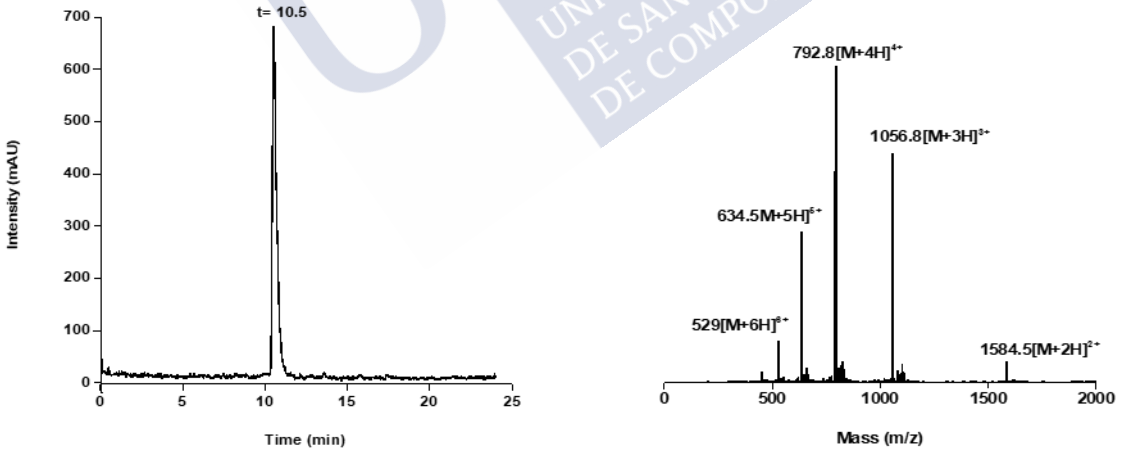


Figure 133. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{140}H_{210}N_{47}O_{39}$: 3168. Found: 1584.5 $[M+2H]^{2+}$; 1056.8 $[M+3H]^{3+}$; 792.8$[M+4H]^{4+}$; 634.5 $[M+5H]^{5+}$; 529 $[M+6H]^{6+}$

TMR-**brHis$_2$(LS)**: TMR-Ahx-DPAA**LS**H**H**KRA**H**NTEAARRSRARKLQR-CONH$_2$

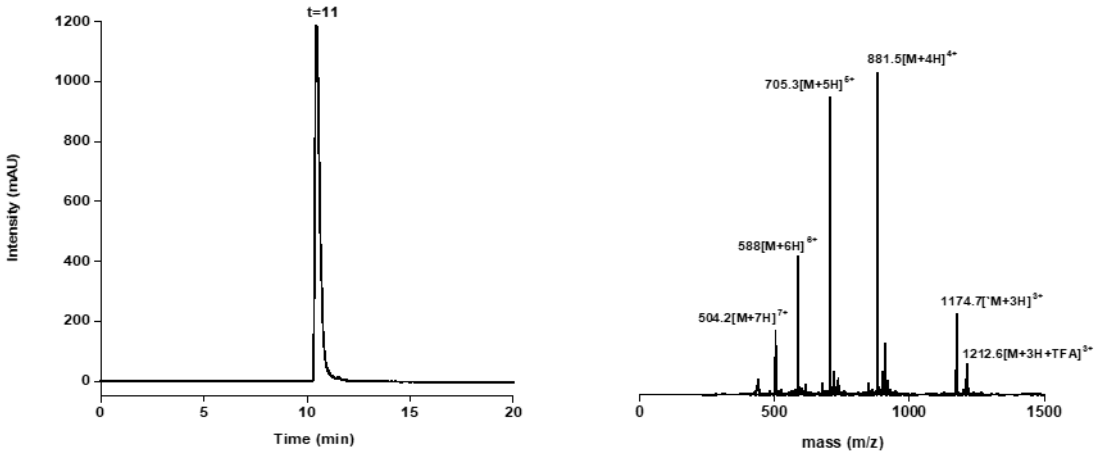


Figure 134. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{155}H_{240}N_{55}O_{41}$: 3524. Found: 1212.6[M+4H+TFA]$^{3+}$; 1174.7[M+3H]$^{3+}$; 881.5 [M+4H]$^{4+}$; 705.3[M+5H]$^{5+}$; 588[M+6H]$^{6+}$; 504[M+7H]$^{7+}$

TMR-**Brhis$_2$(R249A):** TMR-Ahx-DPAA**H**KRA**H**NTEAARRSRARKLQ**A**-CONH$_2$

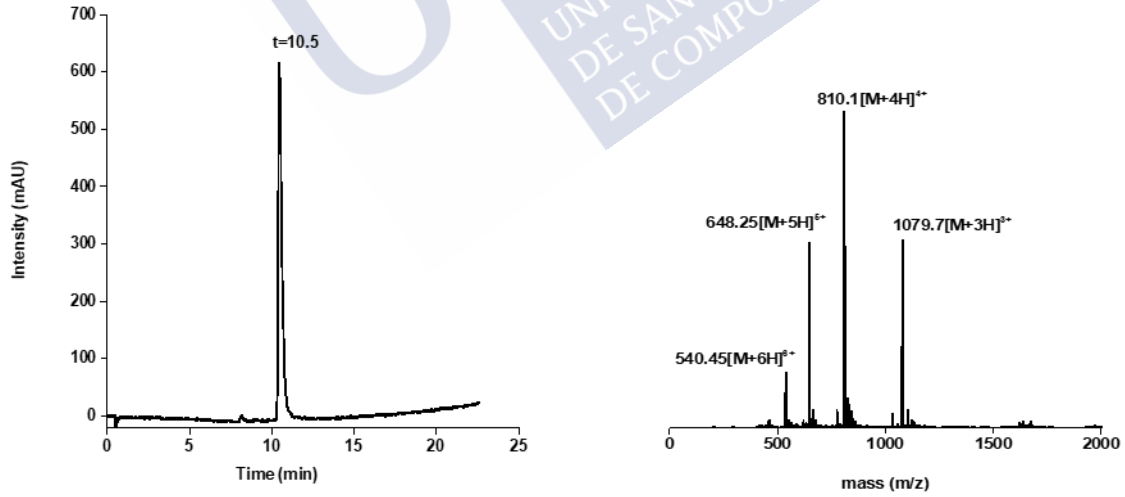


Figure 135. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{142}H_{222}N_{50}O_{38}$: 3236,67. Found: 1079.7 [M+3H]$^{3+}$; 810.1 [M+4H]$^{4+}$; 648.25[M+5H]$^{5+}$; 540.45 [M+6H]$^{6+}$

TMR-**brHis$_2$(R243A):** TMR-Ahx-DPAA**H**KRA**H**NTEAARRS**A**ARKLQR-CONH$_2$

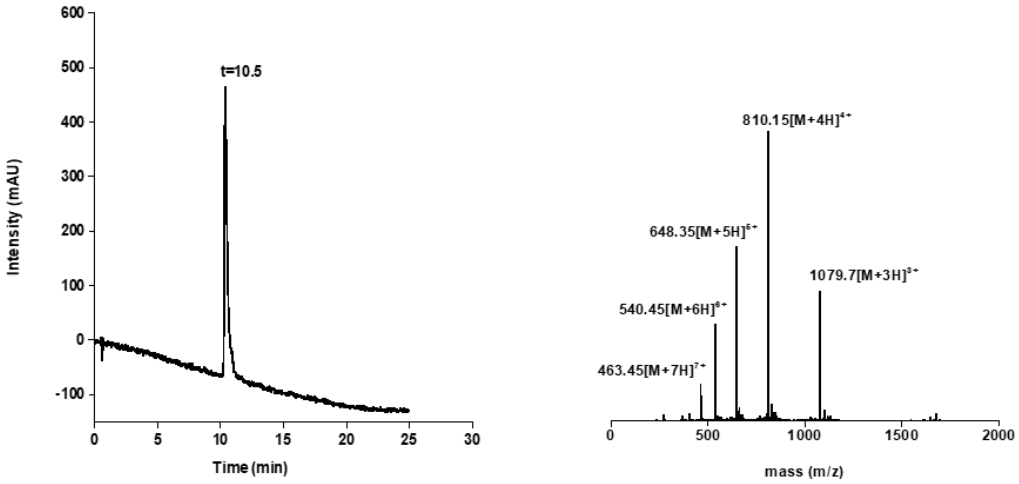


Figure 136. *Left***:** HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{142}H_{222}N_{50}O_{38}$: 3236,67. Found: 1079.7 $[M+3H]^{3+}$; 810.15 $[M+4H]^{4+}$; 648.35$[M+5H]^{5+}$; 540.45 $[M+6H]^{6+}$; 463.45 $[M+7H]^{7+}$

TMR-**brHis$_2$(R241A)**: TMR-Ahx-DPAA**H**KRA**H**NTEAAR**A**SRARKLQR-CONH$_2$

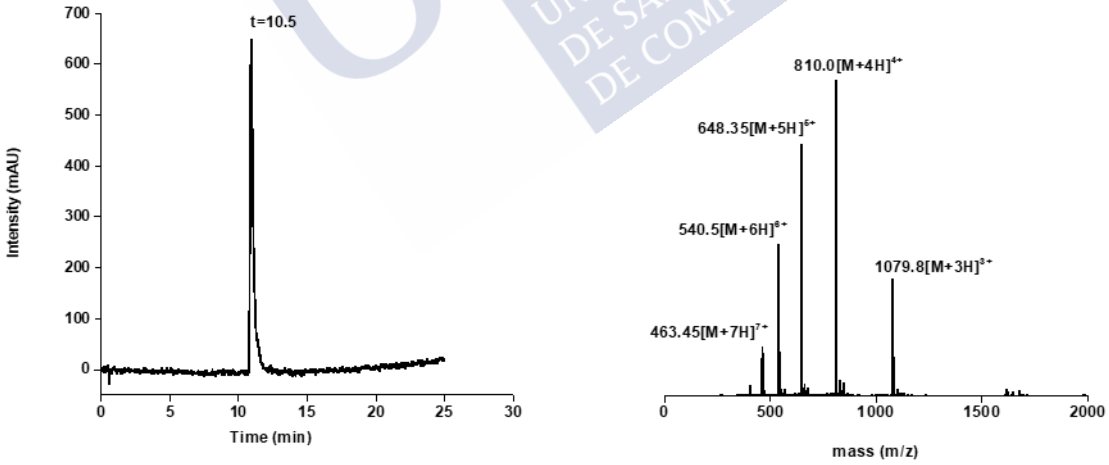


Figure 137. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{142}H_{222}N_{50}O_{38}$: 3236,67. Found: 1079.8 $[M+3H]^{3+}$; 810.0 $[M+4H]^{4+}$; 648.35$[M+5H]^{5+}$; 540.5 $[M+6H]^{6+}$; 463.45 $[M+7H]^{7+}$

TMR-**brHis$_2$(R240A)**: TMR-Ahx-DPAA**H**KRA**H**NTEAA**A**RSRARKLQR-CONH$_2$

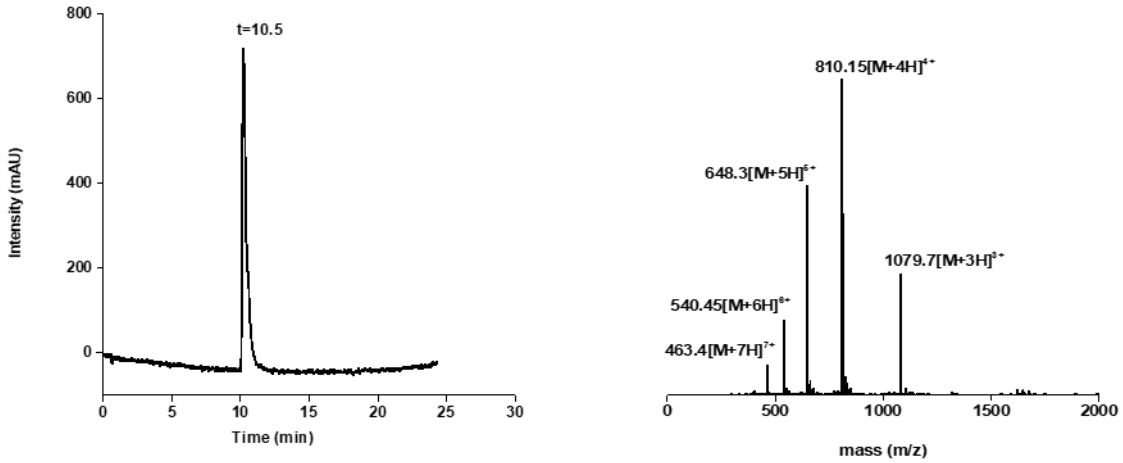


Figure 138. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{142}H_{222}N_{50}O_{38}$: 3236,67. Found: 1079.7 $[M+3H]^{3+}$; 810.15 $[M+4H]^{4+}$; 648.3$[M+5H]^{5+}$; 540.45 $[M+6H]^{6+}$; 463.4 $[M+7H]^{7+}$

TMR-**brHis$_2$(R232A)**: TMR-Ahx-DPAA**H**K**A**A**H**NTEAARRSRARKLQR-CONH$_2$

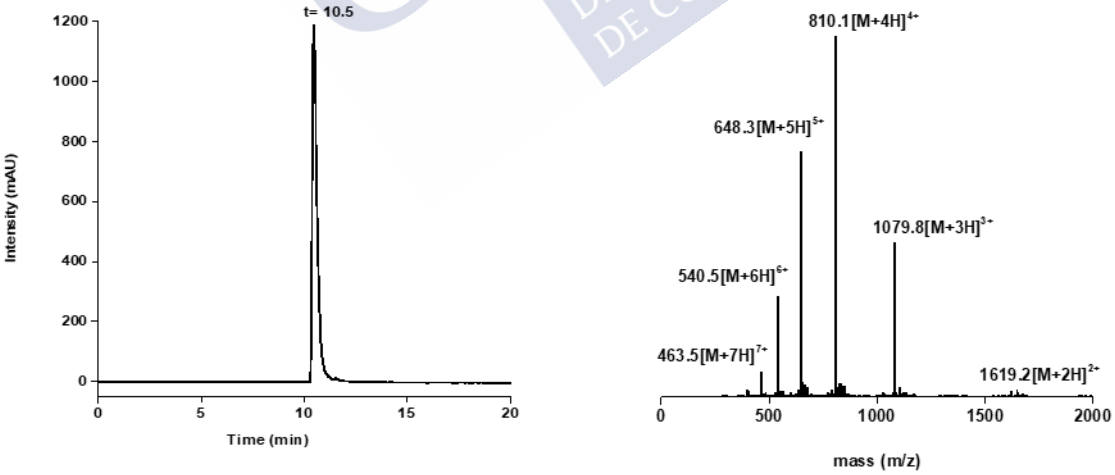


Figure 139. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{142}H_{223}N_{51}O_{37}$: 3236,67. Found: 1619.2 $[M+2H]^{2+}$; 1079.8 $[M+3H]^{3+}$; 810.1 $[M+4H]^{4+}$; 648.3$[M+5H]^{5+}$; 540.5 $[M+6H]^{6+}$; 463.5 $[M+7H]^{7+}$

TMR-**HRGDH**-$CONH_2$

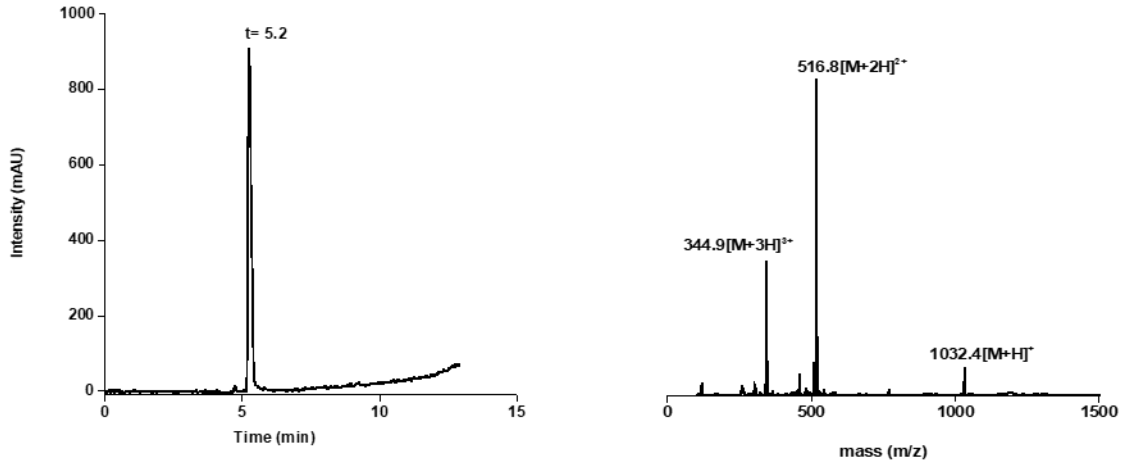


Figure 140. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{49}H_{57}N_{15}O_{11}$: 1031.44. Found: 1032.4 $[M+H]^+$; 516.8$[M+2H]^{2+}$; 344.9$[M+3H]^{3+}$

TMR-**HRGDA**-$CONH_2$

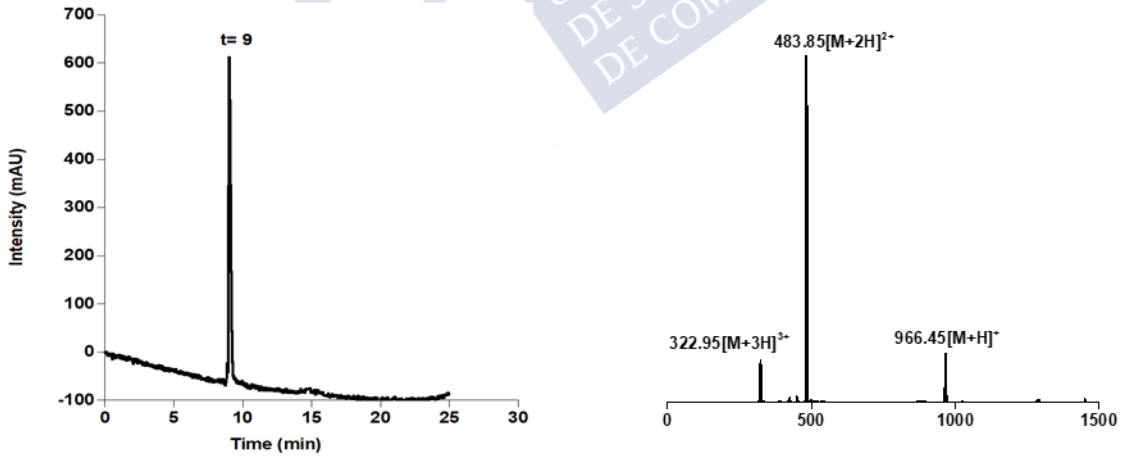


Figure 141. *Left*: HPLC chromatogram of purified peptide. Gradient 5 to 95% B over 30 min. *Right*: Mass spectrum of the purified peptide.

EM–ESI$^+$ (m/z): Calcd. For $C_{46}H_{55}N_{13}O_{11}$: 965.41. Found: 964.45 $[M+H]^+$; 483.85$[M+2H]^{2+}$; 322.95 $[M+3H]^{3+}$

UHPLC-MS of the metallopeptides resulting from mixing the peptides and Pd (II) sources

The peptides were mixed with metal complexes (1:1) in water for 10 min, and the resulting crudes analyzed by UHPLC-MS in order to corroborate the formation of palladopeptides.

TMR-**brHis$_2$**-Pd(en)

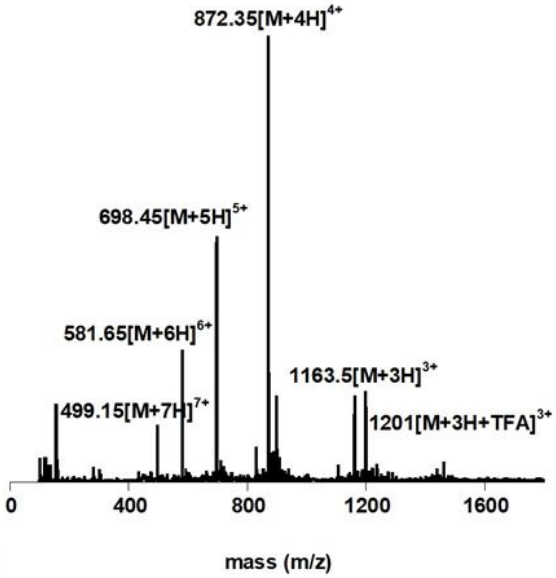


Figure 142. Mass spectrum of the crude resulting from mixing TMR-**brHis$_2$** + $PdCl_2$(en) (1:1), consistent with the formation of a complex containing Pd and the ethylenediamine ligand.

EM–ESI$^+$ (m/z): Calcd. for the metallopeptide $C_{148}H_{230}N_{55}O_{38}Pd$: 3488. Calcd. for the peptide $C_{146}H_{224}N_{53}O_{38}$: 3324.
Found: 1163.5 $[M+3H]^{3+}$; 872.35 $[M+4H]^{4+}$; 698.45 $[M+5H]^{5+}$; 581.65 $[M+6H]^{6+}$; 499.15 $[M+7H]^{7+}$

brHis$_2$-Pd(en)

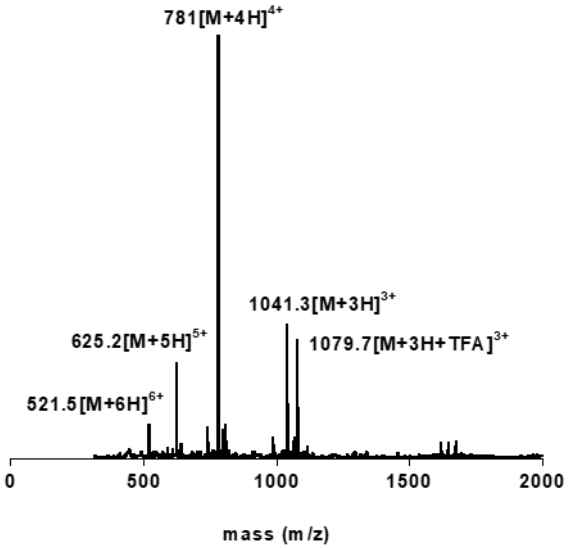


Figure 143. Mass spectrum of the crude resulting from mixing **brHis$_2$** + $PdCl_2$(en) (1:1), consistent with the formation of a complex containing Pd and the ethylendiamine ligand.

EM–ESI$^+$ (m/z): Calcd. for the metallopeptide $C_{125}H_{211}N_{53}O_{35}Pd$: 3122; Calcd. for the peptide $C_{123}H_{205}N_{51}O_{35}$: 2956. Found: 1079.7 $[M+3H+TFA]^{3+}$; 1041.3 $[M+3H]^{3+}$; 781 $[M+4H]^{4+}$; 625.2 $[M+5H]^{5+}$; 521.5 $[M+6H]^{6+}$

TMR-**brHis$_2$(R249A, R245S)**-Pd(en)

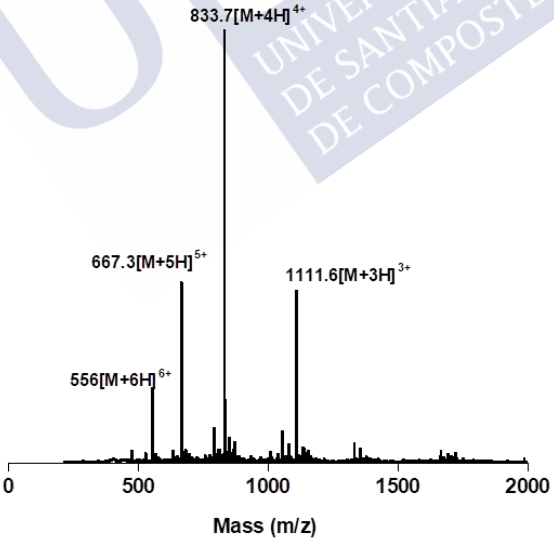


Figure 144. Mass spectrum of the crude resulting from mixing TMR-**brHis$_2$(R249A, R245S)** + $PdCl_2$(en) (1:1), consistent with the formation of a complex containing Pd and the ethylendiamine ligand.

EM–ESI$^+$ (m/z): Calcd. the metallopeptide $C_{142}H_{216}N_{49}O_{39}Pd$: 3334. Calcd. For the peptide $C_{140}H_{210}N_{47}O_{39}$: 3168. Found: 1111.6 $[M+3H]^{3+}$; 833.7 $[M+4H]^{4+}$; 667.3 $[M+5H]5^+$; 556 $[M+6H]^{6+}$

TMR-**brHis$_2$**-Pd

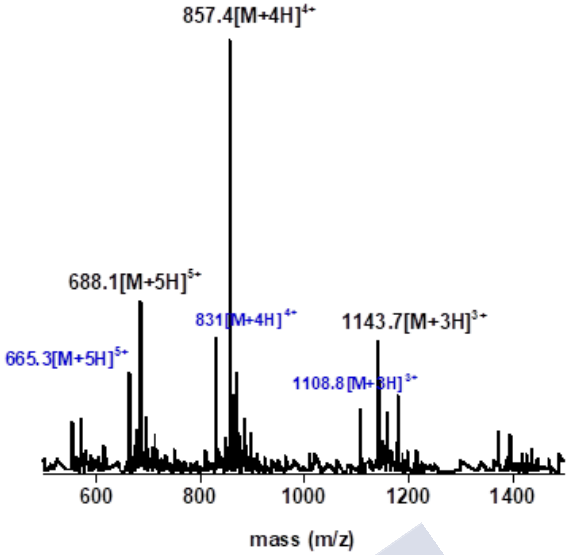


Figure 145. Mass spectrum of the crude resulting from mixing TMR-**brHis$_2$** + $PdCl_2$(COD) (1:1), consistent with the formation of a complex containing Pd. In this case we did not observe the presence of Pd ligands (likely H_2O), which dissociate under the analysis conditions.

EM–ESI$^+$ (m/z): Calcd. for the metallopeptide $C_{146}H_{224}N_{53}O_{38}Pd$: 3429. Calcd. for the peptide: $C_{146}H_{224}N_{53}O_{38}$: 3324.
Found: 1143.7 $[M+3H]^{3+}$; 857.4 $[M+4H]^{4+}$; 688.1 $[M+5H]^{5+}$; Found: 1108.8 $[M+3H]^{3+}$; 831.0 $[M+4H]^{4+}$; 665.3 $[M+5H]^{5+}$ (The blue colour corresponds to the peptide´s mass while the black colour it's the metallopeptide mass).

brHis$_2$-Pd

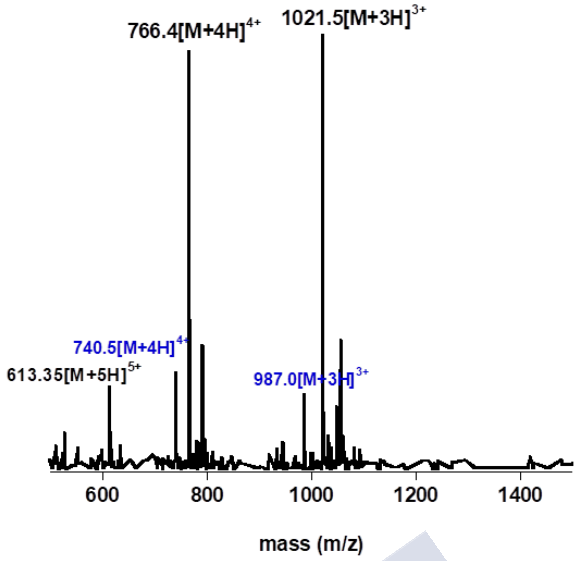


Figure 146. Mass spectrum of the crude resulting from mixing **brHis$_2$** + $PdCl_2$(COD) (1:1), consistent with the formation of a complex containing Pd. We don`t observe the presence of Pd ligands (likely H_2O), which dissociate under the analysis conditions.

EM–ESI$^+$ (m/z): Calcd. for the metallopeptide $C_{123}H_{205}N_{51}O_{35}Pd$: 3063; Calcd. for the peptide $C_{123}H_{205}N_{51}O_{35}$: 2956.
Found: 1021.5 $[M+3H]^{3+}$; 766.4 $[M+4H]^{4+}$; 613.35 $[M+5H]^{5+}$; Found: 987.0 $[M+3H]^{3+}$; 740.5 $[M+4H]^{4+}$ (The blue colour corresponds to the peptide´s mass while the black colour it's the metallopeptide mass).

In addition to this spectrum, we also measured the mass of a complex between **brHis$_2$'** and the same palladium source using ultra high resolution.

brHis$_2$': DPAA**H**KRA**H**NTEAARRSRARKLQR-CONH$_2$

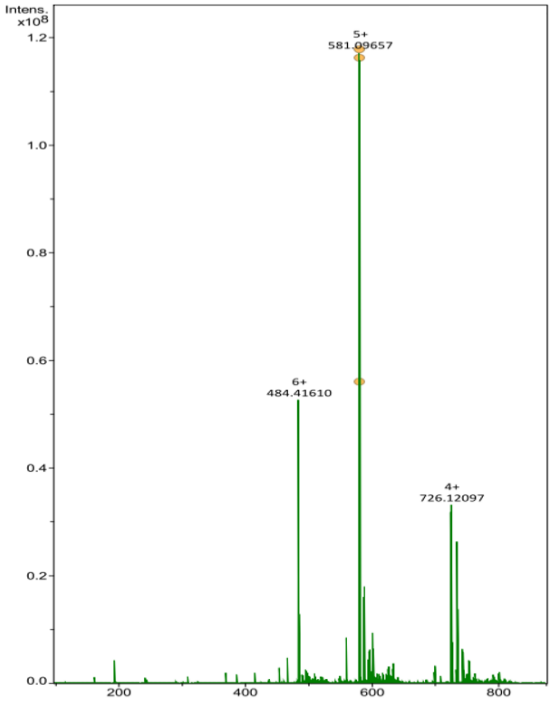


Figure 147. Ultra-high-resolution MS spectrum of the crude resulting from mixing **brHis$_2$**´ + $PdCl_2$(COD) (1:1), consistent with the formation of a complex containing Pd. We do not observe the presence of Pd ligands (likely H_2O), which dissociate under the analysis conditions.

UHR–ESI$^+$ (m/z): Calcd. for the palladopeptide $C_{115}H_{206}N_{51}O_{31}Pd$: 2903; Calcd. for the peptide $C_{115}H_{206}N_{51}O_{31}$: 2797. Found: 726.120 $[M+4H]^{4+}$; 581.096 $[M+5H]^{5+}$;484.416 $[M+6H]^{6+}$

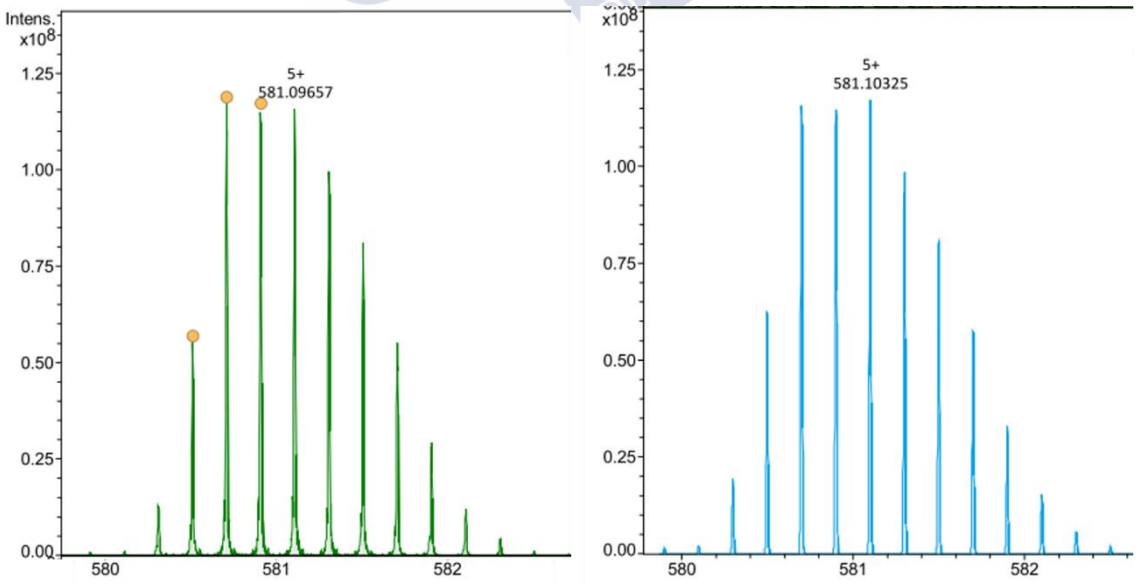


Figure 148. Ultra-high-resolution spectrum of the crude resulting from mixing **brHis$_2$**´ + $PdCl_2$(COD) (1:1) (left), and simulated spectra (right).

TMR-**HRDGH**-Pd

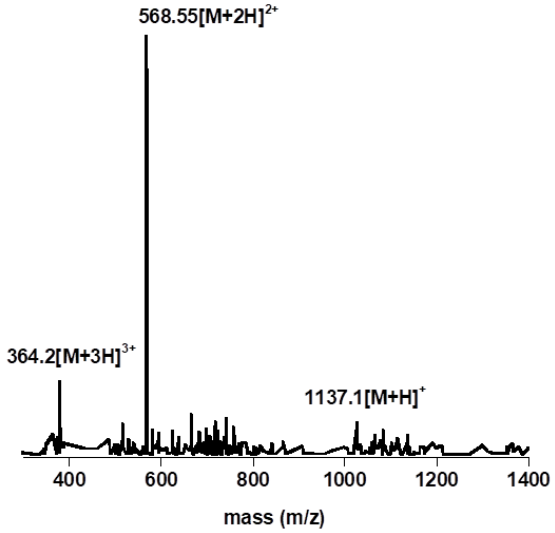


Figure 149. Mass spectrum of the crude resulting from mixing TMR-**HRGDH** + $PdCl_2(COD)$ (1:1), consistent with the formation of a complex containing Pd. We do not observe the presence of Pd ligands (likely H_2O), which dissociate under the analysis conditions.

EM–ESI$^+$ (m/z): Calcd. for the metallopeptide $C_{49}H_{57}N_{15}O_{11}Pd$: 1136; Calcd. for the peptide $C_{49}H_{57}N_{15}O_{11}$: 1031.44. Found: 1137.1 $[M+1H]^{1+}$; 568.55 $[M+2H]^{2+}$; 364.2 $[M+3H]^{3+}$

TMR-**brHis$_2$-Pd1**

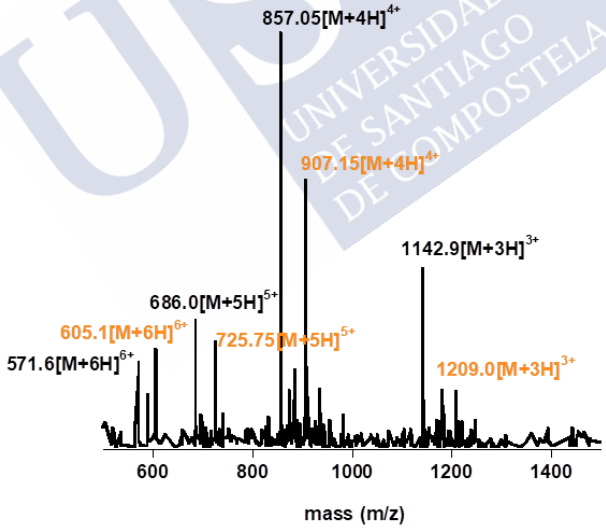
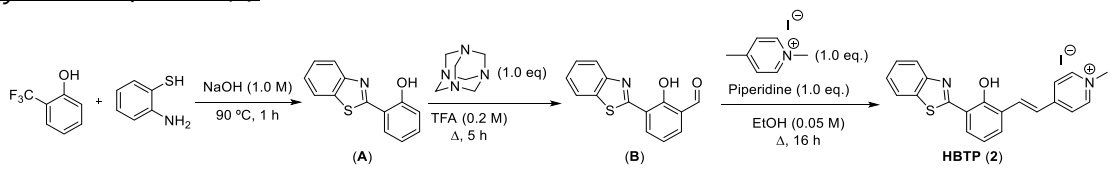


Figure 150. Mass spectrum of the crude resulting from mixing TMR-**brHis$_2$** + **Pd1** (1:1), consonant with the formation of the expected complex.

EM–ESI$^+$ (m/z): Calcd. for the metallopeptide $C_{175}H_{225}N_{59}O_{38}Pd$: . Calcd. for the peptide: $C_{146}H_{224}N_{53}O_{38}$: 3324. Found: Found: 1142.9 $[M+3H]^{3+}$; 857.05 $[M+4H]^{4+}$; 686.0 $[M+5H]^{5+}$; 571.6 $[M+6H]^{6+}$; 1209.0 $[M+3H]^{3+}$; 907.15 $[M+4H]^{4+}$; 725.75 $[M+5H]^{5+}$; 605.1 $[M+6H]^{6+}$ (The black colour corresponds to the peptide containing a coordinated palladium atom [TMR-**brHis$_2$** + Pd^{2+}], while the orange colour corresponds to the palladopeptide where the anthracene moiety of the ligand has broken during the ionization).

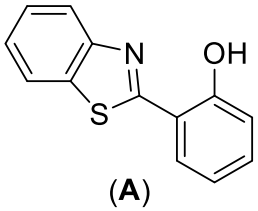
Synthesis of HBTPQ (1), HBTP (2), Substrate (3), and complexes Pd1 and $PdCl_2$(Boc-His)$_2$

Synthesis of HBTP (2)



*2-(benzo[d]thiazol-2-yl)phenol (**A**)*:

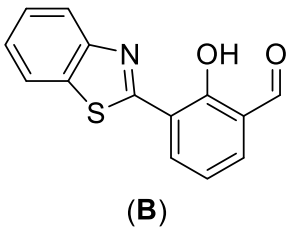
A mixture of 2-(trifluoromethyl)phenol (810.6 mg, 5.0 mmol), 2-aminobenzenethiol (626.0 mg, 5.0 mmol, 1.0 eq.) and 1 N NaOH (20.0 mL, 20.0 mmol, 4.0 eq.) was heated at 90 ºC under nitrogen for 1 h. After cooling to room temperature, the mixture was neutralized with 1 N HCl and the resulting suspension was filtered. The solid was washed with water (3 x 20.0 mL), dried under vacuum and recrystallized from MeOH (reflux and then cooled down to 5 ºC) to give the expected product **A.**


(**A**)

2-(benzo[d]thiazol-2-yl)phenol (A). White solid. **Yield** = 80%. **^{1}H** (300 MHz, $CDCl_3$): 12.53 (s, 1H), 8.00 (d, *J* = 8.1 Hz, 1H), 7.91 (d, *J* = 7.9 Hz, 1H), 7.70 (d, *J* = 7.9 Hz, 1H), 7.51 (t, *J* = 7.6 Hz, 1H), 7.46 - 7.35 (m, 2H), 7.11 (d, *J* = 8.3 Hz, 1H), 6.96 (t, *J* = 7.4 Hz, 1H).

*3-(benzo[d]thiazol-2-yl)-2-hydroxybenzaldehyde (**B**)*:

2-(benzo[d]thiazol-2-yl)phenol (**A**, 900.0 mg, 3.9 mmol) was dissolved in TFA (19.8 mL, 0.2 M), and then hexamethylenetriamine (555.1 mg, 3.9 mmol, 1.0 eq.) was added in one portion. The solution was refluxed until the starting material was consumed. Then, the mixture was cooled to room temperature and neutralized with 1 N NaOH. The resulting solid was washed with saturated brine, and the residue purified by silica gel column chromatography using hexane / EtOAc (95:5). The product **B** was obtained as a yellow solid.

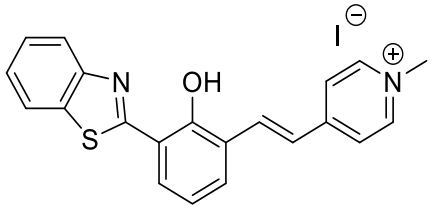

(**B**)

3-(benzo[d]thiazol-2-yl)-2-hydroxybenzaldehyde (B). Yellow solid. **R_f** = 0.6 (Hexane / EtOAc 80:20). **Yield** = 50%. **RP-HPLC** t_r = 8.2 - 8.4 min (Method: from 95% H_2O (0.1%TFA) / 5 % MeCN (0.1% TFA) to 5% H_2O (0.1%TFA) / 95 % MeCN (0.1% TFA) in 12 min at 0.5 mL/min). **^{1}H** (300 MHz, $CDCl_3$): 10.58 (s, 1H), 8.07 (d, *J* = 8.1 Hz, 1H), 7.96 (t, *J* = 8.0 Hz, 2H), 7.58 (t, *J* = 7.6 Hz, 1H), 7.48 (t, *J* = 7.6 Hz, 1H), 7.12 (t, *J* = 7.7 Hz, 1H).

*Synthesis of (E)-4-(3-(benzo[d]thiazol-2-yl)-2-hydroxystyryl)-1-methylpyridin-1-ium (HBTP, **2**)*:

3-(benzo[d]thiazol-2-yl)-2-hydroxybenzaldehyde (**B**, 50.0 mg, 0.2 mmol) and 1,4-dimethylpyridin-1-ium (46.0 mg, 0.2 mmol, 1.0 eq.) was dissolved in EtOH (3.9 mL, 0.05 M). Then, piperidine (19 µL, 0.2 mmol, 1.0 eq.) was added dropwise and the solution was

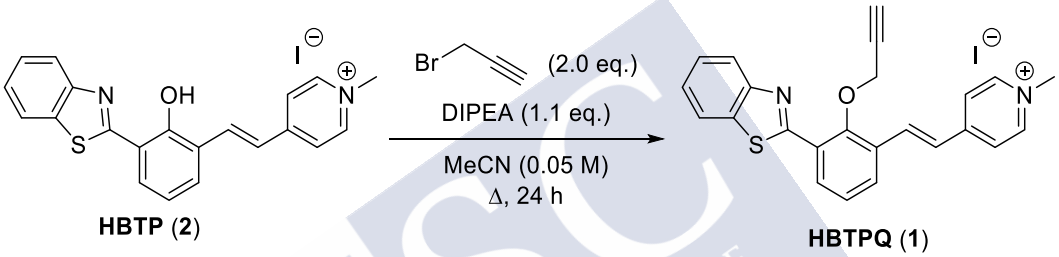
refluxed for vigorous stirring. After completion of the reaction (16 h), the mixture was concentrated under reduced pressure, and the residue was recrystallized from MeOH and Et_2O.



(HBTP, **2**)

(E)-4-(3-(benzo[d]thiazol-2-yl)-2-hydroxystyryl)-1-methylpyridin-1-ium (HBTP, 2).
Pale yellow solid. **Yield** = 78%. **RP-HPLC** t_r = 6.9 - 7.1 min (Method: from 95% H_2O (0.1%TFA) / 5 % MeCN (0.1% TFA) to 5% H_2O (0.1%TFA) / 95 % MeCN (0.1% TFA) in 17 min at 0.5 mL/min). **^{1}H** (300 MHz, DMSO-d_6): 13.35 (s, 1H), 8.86 (d, *J* = 7.7 Hz, 2H), 8.25 (d, *J* = 7.7 Hz, 2H), 8.23 (d, *J* = 7.9 Hz, 1H), 8.18 (d, *J* = 16.9 Hz, 1H), 8.13 (d, *J* = 8.0 Hz, 1H), 8.01 (d, *J* = 8.0 Hz, 1H), 7.93 (d, *J* = 8.0 Hz, 1H), 7.67 (d, *J* = 16.9 Hz, 1H), 7.62 (dd, *J* = 7.1 Hz, 1H), 7.55 (dd, *J* = 7.1 Hz, 1H), 7.19 (dd, *J* = 7.7 Hz, 1H), 4.26 (s, 3H).

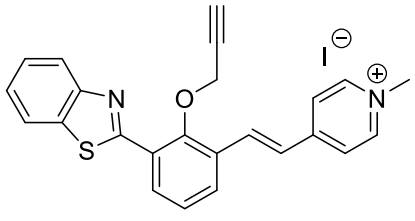
Synthesis of HBTPQ (1)



HBTP (**2**) **HBTPQ** (**1**)

*(E)-4-(3-(benzo[d]thiazol-2-yl)-2-(prop-2-yn-1-yloxy)styryl)-1-methylpyridin-1-ium (**HBTPQ, 1**)*:

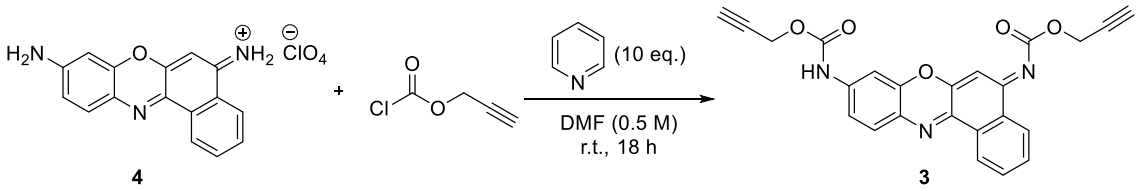
In a flame-dried two necked bottom flask, (E)-4-(3-(benzo[d]thiazol-2-yl)-2-hydroxystyryl)-1-methylpyridin-1-ium (**2**, 60.0 mg, 130.0 µmol), propargyl bromide (28 µL, 260.0 µmol, 2.0 eq.) and ethyldiisopropylamine (24 µL, 140.0 µmol, 1.1 eq.) was dissolved in MeCN (2.5 mL, 0.05 M), and the mixture was refluxed at 100 ºC for 24 h. After evaporation of the solvent under reduced pressure, the residue was purified by RP-HPLC (Method: from 55% H_2O (0.1%TFA) / 45 % MeCN (0.1% TFA) to 5% H_2O (0.1%TFA) / 95 % MeCN (0.1% TFA) in 35 min at 10.0 mL/min).



HBTPQ (**1**)

(E)-4-(3-(benzo[d]thiazol-2-yl)-2-(prop-2-yn-1-yloxy)styryl)-1-methylpyridin-1-ium (HBTPQ, 1).
Red solid. **Yield** = 55. **RP-HPLC** t_r = 9.9 – 9.9 min (Method: from 95% H_2O (0.1%TFA) / 5 % MeCN (0.1% TFA) to 5% H_2O (0.1%TFA) / 95 % MeCN (0.1% TFA) in 17 min at 0.5 mL/min). **^{1}H** (300 MHz, DMSO-d_6): 8.95 (d, *J* = 6.5 Hz 2H), 8.45 (d, *J* = 8.0 Hz, 1H), 8.33 (d, *J* = 6.5 Hz, 2H), 8.20 (d, *J* = 8.0 Hz, 1H), 8.15-8.08 (m, 3H), 7.68 (d, *J* = 16.5 Hz, 1H), 7.60 - 7.57 (m, 1H), 7.55 - 7.49 (m, 2H), 4.87 (d, *J* = 2.0 Hz, 2H), 4.31 (s, 3H), 3.59 (s, 1H).

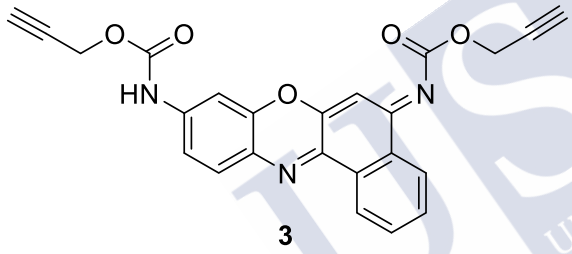
Synthesis of 3



4 3

*prop-2-yn-1-yl (E)-(5-(((prop-2-yn-1-yloxy)carbonyl)imino)-5H-benzo[a]phenoxazin-9-yl)carbamate (**3**):*

Cresyl violet perchlorate (**4**) (100 mg, 0.27 mmol) was dissolved in anhydrous DMF (1.0 mL) and cooled to 0 ºC. Separately, pyridine (140 µL, 2.8 mmol, 10.0 eq.) and propargyl chloroformate (116 µL, 1.4 mmol, 5 eq.) were each dissolved in anhydrous DMF (0.5 mL) and the solutions added dropwise to the cresyl violet. The reaction mixture was stirred at room temperature for 18 h, after which it was diluted with iPrOH/$CHCl_3$ (25:75, 10.0 mL), washed with 5 % HCl (2 × 10.0 mL) and $NaHCO_3$ (2 × 5.0 mL) and dried over $MgSO_4$. The solvent was removed in vacuo and the resulting crude purified by silica gel column chromatography using using hexane / EtOAc (2:1). The compound **3** was obtained as an orange solid.

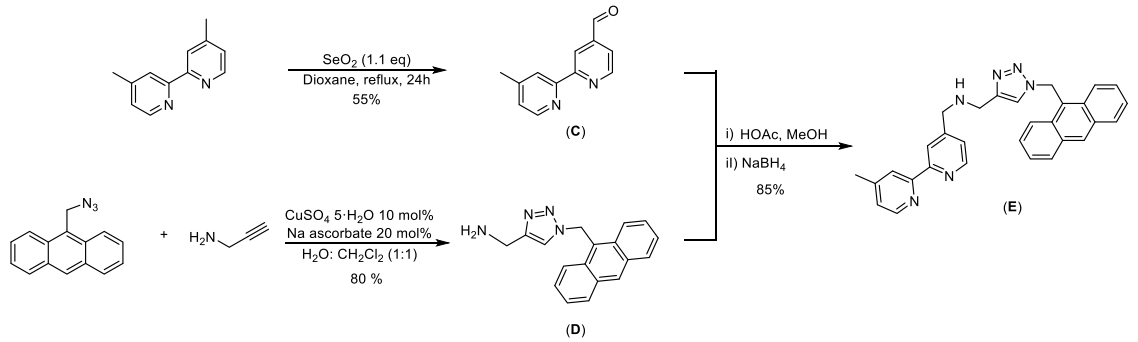


3

prop-2-yn-1-yl(E)-(5-(((prop-2-yn-1-yloxy)carbonyl)imino)-5H-benzo[a]phenoxazin-9-yl)carbamate (3). Orange solid. **Yield** = 21%. $\mathbf{R_f}$ = 0.42 (Hexane/EtOAc 2:1). ^{1}**H** (500 MHz, DMSO-d_6) 10.50 (s, 1H), 8.61 (d, *J* = 8.0 Hz, 1H), 8.31 (d, *J* = 7.8 Hz, 1H), 7.86 (t, *J* = 7.6 Hz, 1H), 7.82 - 7.75 (m, 2H), 7.62 (d, *J* = 2.1 Hz, 1H), 7.43 (dd, *J* = 8.8, 2.2 Hz, 1H), 6.54 (d, *J* = 1.4 Hz, 1H), 4.91 (d, *J* = 2.2 Hz, 2H), 4.83 (d, *J* = 2.3 Hz, 2H), 3.67 (t, *J* = 2.0 Hz, 1H), 3.61 (t, *J* = 2.0 Hz, 1H).

*Synthesis of **Pd1***

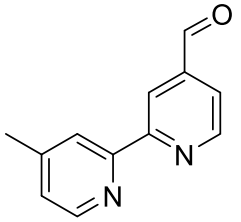
Synthesis of the corresponding ligand (E)



*4′-methyl-2,2′-bipyridine-4-carboxaldehyde (**C**):*

In a round bottom flask under nitrogen, 4,4′-dimethyl-2,2′-bipyridine (276.0 mg, 1.5 mmol) and SeO_2 (191.0 mg, 1.7 mmol) were dissolved in degassed 1,4-dioxane (10.0 mL).

The mixture was heated under reflux for 24 h and the resulting suspension was filtered (while hot). The filtrate was concentrated, re-dissolved in ethyl acetate (200.0 mL) and filtered to remove additional solid material. The filtrate was extracted with 1 M Na_2CO_3 (2 x 100.0 mL), and with 0.3 M $Na_2S_2O_3$ (3 x 100.0 mL) to form the aldehyde bisulfite. The aqueous bisulfite fractions were combined, adjusted to pH 10 with Na_2CO_3 and extracted with CH_2Cl_2 (4 x 100.0 mL). The organic fractions were combined and concentrated to dryness to give the product **C**.

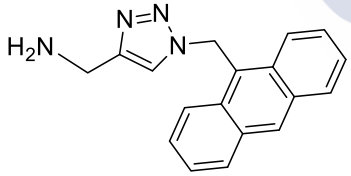


(**C**)

4'-methyl-2,2'-bipyridine-4-carboxaldehyde (C). White powder. **Yield** = 55%.**^{1}H** (300 MHz, $CDCl_3$) 10.18 (s, 1H), 8.90 (d, *J* = 5.0 Hz, 1H), 8.83 (s, 1H), 8.58 (d, *J* = 5.0 Hz, 1H), 8.28 (s, 1H), 7.72 (dt, *J* = 4.9, 1.5 Hz, 1H), 7.20 (d, *J* = 5.0 Hz, 1H), 2.47 (s, 3H).

*(1-(anthracen-9-ylmethyl)-1H-1,2,3-triazol-4-yl)methanamine (**D**):*

In a round bottom flask under nitrogen were added the 9-(azidomethyl)anthracene (100mg, 0.42 mmol, 1.0 eq.), the propargylamine (32 µL, 0.49 mmol, 1.15 eq.) and 6.0 mL of a 1:1 mixture of CH_2Cl_2 : H_2O. Copper sulphate pentahydrate (10.64 mg, 0.042 mmol, 0.1 eq.) and sodium ascorbate (17.0 mg, 0.087 mmol, 0.2 eq.) were added, and the mixture was stirred overnight. The reaction was opened to air and extracted with CH_2Cl_2 (3 x 10.0 mL); the organic phase was washed with brine (30.0 mL) and dried over anhydrous $MgSO_4$. The solvent was removed and the corresponding triazole was purified by flash column chromatography using increasing percentages of CH_2Cl_2 / MeOH (98:2 to 90:10). The compound **D** was obtained as a mustard powder.



(**D**)

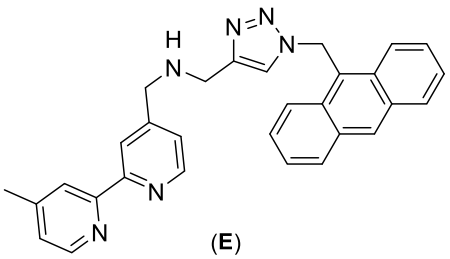
(1-(anthracen-9-ylmethyl)-1H-1,2,3-triazol-4-yl)methanamine (D).
Mustard solid. **R_f** = 0.32 (CH_2Cl_2 / MeOH 9:1). **^{1}H** (500 MHz, MeO*D*-d_4): 8.51 (s, 1H), 8.29 (d, *J* = 8.9 Hz, 2H), 8.01 (d, *J* = 8.4 Hz, 2H), 7.58 - 7.52 (m, 2H), 7.51 - 7.44 (m, 2H), 7.27 (s, 1H), 6.45 (s, 2H), 4.58 (bs, 4H).**^{13}C** (126 MHz, MeO*D*-d_4): 133.3 (C), 132.5 (C), 131.6 (CH), 131.2 (CH), 129.3 (CH), 127.0 (CH), 125.5 (C), 124.61(CH), 123.2 (CH), 48.0 (CH_2). **HRMS-ESI** calculated for $C_{18}H_{17}N_4$: 289.1448, found 289.1447.

*1-(1-(anthracen-9-ylmethyl)-1H-1,2,3-triazol-4-yl)-N-((4'-methyl-[2,2'-bipyridin]-4-yl) methyl)methanamine (**E**):*

In Schlenk flask under nitrogen atmosphere, 4'-methyl-2,2'-bipyridine-4-carboxaldehyde (**C**, 21.7 mg, 0.110 mmol, 1.0 eq.) and the amine **D** (32.1 mg, 0.111 mmol, 1.02 eq.) were dissolved in MeOH (3.0 mL). A drop of glacial acetic acid was added (25.0 µL), and the mixture was heated under reflux for 1 h. The resulting mixture was cooled, and an excess of $NaBH_4$ (17 mg, 0.44 mmol, 4.0 eq.) was added. After the reaction was complete (TLC monitoring), the reaction was quenched with $Na_2CO_{3\,(aq)}$. The methanol was then removed under reduced pressure, and the aqueous phase extracted with CH_2Cl_2 (3 x 10.0

mL). The organic phase was washed with brine and dried over $MgSO_4$, and the crude obtained after removal of the solvent was purified by flash column chromatography using CH_2Cl_2 / MeOH mixture as eluent (92:8 to 90:10). The compound **E** was obtained as a pale-yellow powder.

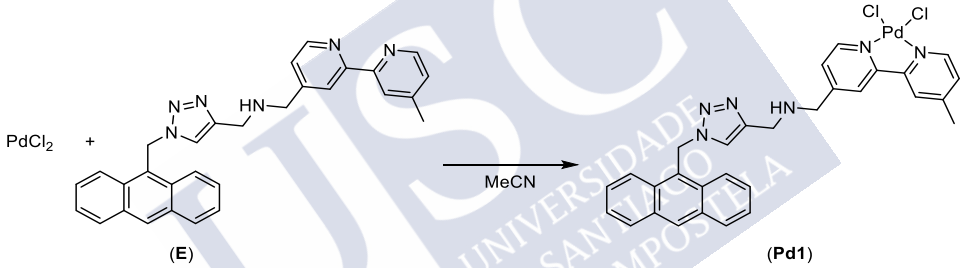
1-(1-(anthracen-9-ylmethyl)-1H-1,2,3-triazol-4-yl)-N-((4'-methyl-[2,2'-bipyridin]-4-yl) methyl)methanamine (**E**).



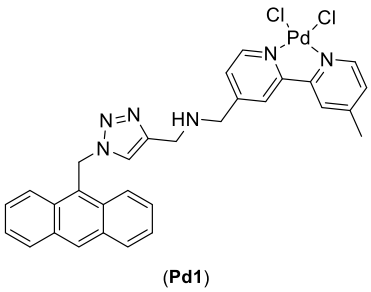
(**E**)

Yellow solid. $\mathbf{R_f}$ = 0.20 (CH_2Cl_2 / MeOH 9:1) **Yield** = 85%. **^{1}H** (300 MHz, $CDCl_3$): 8.56 (s, 1H), 8.49 (dd, *J* = 12.6, 5.0 Hz, 2H), 8.29 (d, *J* = 8.8 Hz, 2H), 8.19 (d, *J* = 11.4 Hz, 2H), 8.06 (d, *J* = 8.4 Hz, 2H), 7.54 (m, 4H), 7.16 (d, *J* = 4.6 Hz, 1H), 7.10 (d, *J* = 4.8 Hz, 1H), 7.04 (s, 1H), 6.51 (s, 2H), 3.79 (s, 2H), 3.75 (s, 2H) 2.41 (s, 3H). **^{13}C** (75 MHz, $CDCl_3$): 156.4 (C), 155.9 (C), 149.9 (C), 149.3 (CH), 149.0 (CH), 148.3 (C), 146.4 (C), 131.54 (C), 130.9 (C), 129.9 (CH), 129.6 (CH), 127.8 (CH), 125.5 (CH), 124.8 (CH), 123.9 (C), 123.1 (CH), 122.1 (CH), 121.4(CH), 120.6 (CH), 52.2 (CH_2), 46.5 (CH_2), 44.2 (CH_2), 21.3 (CH_3). **HRMS-ESI** calculated for $C_{30}H_{27}N_6$: 471.2292, found 471.2294.

*Synthesis of the complex **Pd1:***



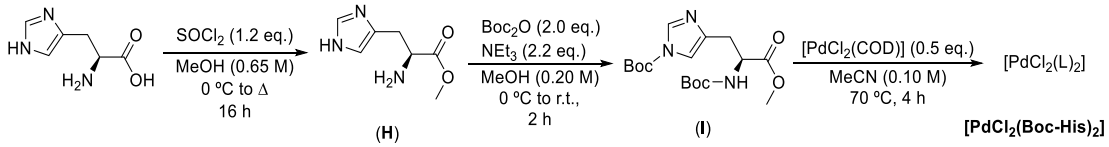
In a Schlenk under nitrogen atmosphere, $PdCl_2$ (11.8 mg, 0.066 mmol, 1.0 eq.) was added, suspended in MeCN (1.0 mL) and refluxed until all the solid dissolved. To the corresponding hot solution was added the ligand **E** (32.0 mg, 0.068 mmol, 1.02 eq.) dissolved in a mixture of MeCN : MeOH (4:1) (500.0 µL). We rapidly observed the formation of a precipitate. The reaction was allowed to cool down to room temperature and the solid to settle. Once cool, the supernatant was decanted, the solid washed twice with diethyl ether and dried over vacuum. The compound **Pd1** was obtained as a yellow powder.



(**Pd1**)

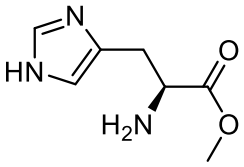
Pd1. yellow powder. **Yield** = 93%. **^{1}H** (300 MHz, DMSO-d_6): 7.82 (dd, *J* = 6.0, 4.0 Hz, 2H), 7.77 (s, 1H), 7.67 (d, *J* = 8.8 Hz, 2H), 7.30 (s, 1H), 7.27 - 7.19 (m, 3H), 6.93 (s, 1H), 6.79 - 6.52 (m, 6H), 5.71 (s, 2H), 2.90 (s, 2H), 2.81 (s, 2H), 1.50 (s, 3H). **^{13}C** (75 MHz, DMSO-d_6): 156.0 (C), 155.4(C), 155.2 (C), 153.1 (C), 148.7 (CH), 148.5 (CH), 145.9 (C), 130.9 (C), 130.2 (C), 129.1 (CH), 129.0 (C), 127.5 (CH), 127.0 (CH), 125.7 (C), 125.4 (CH), 124.2 (C), 123.2 (C), 122.5 (CH), 122.2 (CH), 50.7 (CH_2), 45.4 (CH_2), 43.5 (CH_2), 20.9 (CH_3). **HR-ESI** Calculated for $C_{60}H_{52}Cl_4N_{12}Pd_2$: $[2M-2Cl]^{2+}$ 612.0945, founded 612.0948.

*Synthesis of [$PdCl_2$(Boc-His)$_2$] (**Pd2**)*



*Methyl L-histidinate (**H**):*

To a suspension of *L*-histidine (2.0 g, 12.9 mmol) in MeOH (20.0 mL, 0.65 M), $SOCl_2$ (1.12 mL, 15.5 mmol, 1.2 eq.) was added dropwise at 0 ºC. The reaction was then refluxed for 16 h. After full conversion, the mixture was concentrated to dryness and co-evaporated several times with more MeOH. The solid was triturated with Et_2O to give the product as a white solid.

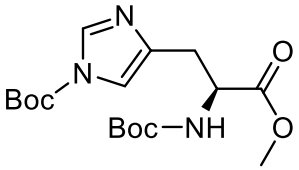


(**H**)

Methyl-*L*-histidinate (H). White solid. $\mathbf{R_f}$ = 0.65 (CH_2Cl_2 / MeOH 7:3). **Yield** = 99%. ^{1}**H** (300 MHz, DMSO-d_6): 9.09 (s, 1H), 7.53 (s, 1H), 4.48 (t, *J* = 6.8 Hz, 1H), 3.72 (s, 3H), 3.32 (d, *J* = 6.8 Hz, 2H).

*Tert-butyl (S)-4-(2-((tert-butoxycarbonyl)amino)-3-methoxy-3-oxopropyl)-1H-imidazole-1-carboxylate (**I**):*

Methyl-*L*-histidinate (**H**, 2.2 g, 12.9 mmol) was dissolved in MeOH (64.5 mL, 0.2 M) and then di-*tert*-butyl-dicarbonate (5.6 g, 25.8 mmol, 2.0 eq.) was added. The solution was cooled to 0 ºC and NEt_3 (3.9 mL, 28.2 mmol, 2.2 eq.) was added dropwise. After that, the solution was stirred vigorously at room temperature for 2 h. The mixture was then diluted with EtOAc and washed with water. The organic layer was dried over $MgSO_4$, concentrated under reduced pressure and the residue was purified by silica gel column chromatography using hexane / Et_2O (1:1 to 0:1). The product **I** was obtained as a white solid.



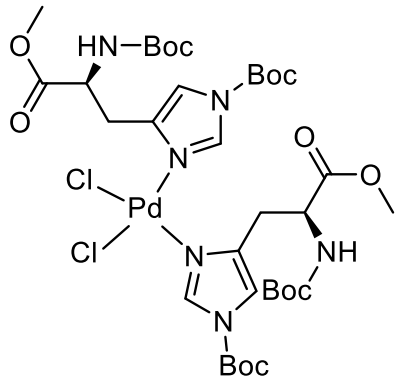
(**I**)

tert-butyl (S)-4-(2-((tert-butoxycarbonyl)amino)-3-methoxy-3-oxopropyl)-1H-imidazole-1-carboxylate (I). White solid. $\mathbf{R_f}$ = 0.4 (Et_2O). **Yield** = 83%. ^{1}**H** (300 MHz, DMSO-d_6): 8.12 (s, 1H), 7.25 (s, 1H), 7.20 (d, *J* = 8.1 Hz, 1H), 4.26 (dq, *J* = 8.4, 5.3 Hz, 1H), 3.63 (s, 3H), 3.01 - 2.68 (m, 2H), 1.56 (s, 9H), 1.35 (s, 9H).

Synthesis of [$PdCl_2$(Boc-His)$_2$] (**Pd2**):

[$PdCl_2$(COD)] (150.0 mg, 525.0 µmol) was added to a flame-dried Schlenk tube and dissolved in dry MeCN (5.2 mL, 0.1 M). Then, the ligand **I** (388.2 mg, 1.05 mmol, 2.0 eq.) was added under nitrogen atmosphere. The solution was refluxed for 3 h. After cooling down to room temperature, the solution was concentrated *in vacuo*. Dry Et_2O (5.0 mL) was added, the residue triturated and the solvent evaporated. Dry CH_2Cl_2 (3.0 mL) was added and hexane was added until a precipitated was observed. The solvents were

decanted and the solid was washed with Et_2O (3 x 15.0 mL). The complex [$PdCl_2$(Boc-His)$_2$] was obtained as a yellow solid.



[$PdCl_2$(Boc-His)$_2$]

[$PdCl_2$(Boc-His)$_2$] (Pd2). Yellow solid. **Yield** = 75%. **^{1}H** (300 MHz, DMSO-d_6): 8.12 (s, 1H), 7.25 (s, 1H), 7.20 (d, J = 8.1 Hz, 1H), 4.26 (dq, J = 8.4, 5.3 Hz, 1H), 3.63 (s, 3H), 3.01 – 2.68 (m, 2H), 1.56 (s, 9H), 1.35 (s, 9H). **HRMS-ESI** Calculated for $C_{34}H_{54}Cl_2N_6O_{12}Pd$: 937.21, found 937.39.

Characterization of Products

ESI-MS of the Pd1 and [$PdCl_2$(Boc-His)$_2$]

ESI-MS of the Pd1

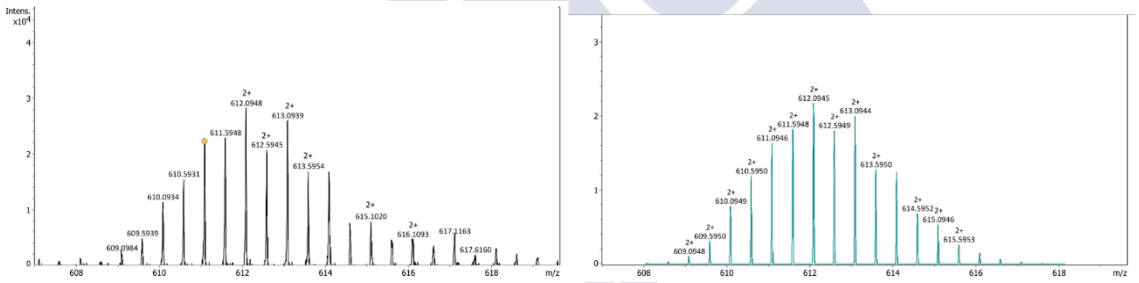


Figure 151. ESI-MS of the complex **Pd1**.

HR–ESI$^+$ (m/z): Calcd. for $C_{60}H_{52}Cl_4N_{12}Pd_2$: $[2M-2Cl]^{2+}$ 612.0945, founded 612.0948

ESI-MS of the [$PdCl_2$(boc-His)$_2$] (Pd2)

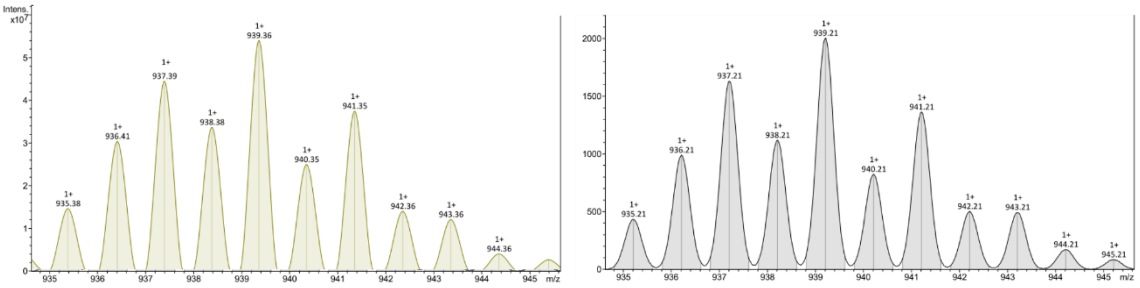


Figure 152. ESI-MS of the complex [**$PdCl_2$(boc-His)$_2$]** (**Pd2**).

HRMS-ESI Calculated for $C_{34}H_{54}Cl_2N_6O_{12}Pd$: 937.21, found 937.39.

*UV and fluorescence of compounds **1, 2, 3, 4** and **Pd1.***

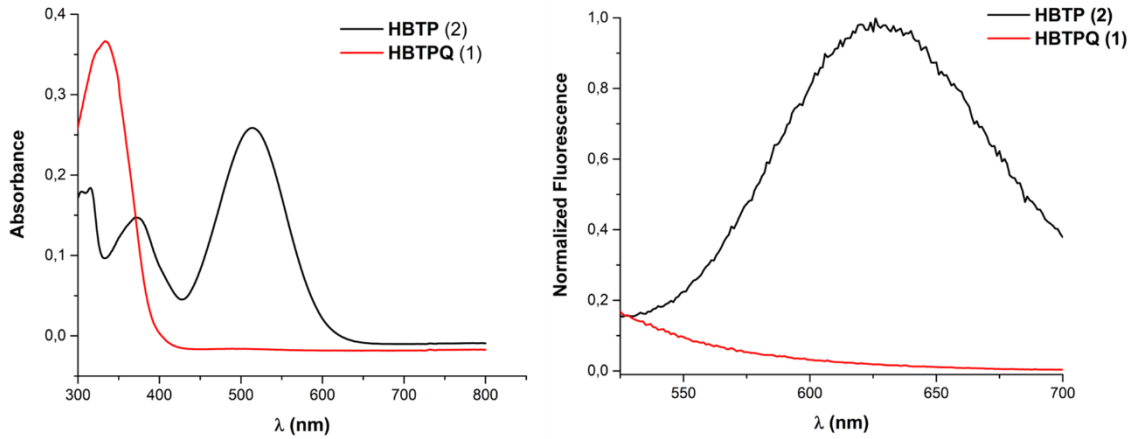


Figure 153. *Left*: Absorption spectra of **1** and **2** (20 µM, 7:3 v/v DMSO / H_2O). *Right*: Fluorescence spectra of **1** and **2** (6 µM, 7:3 v/v DMSO / H_2O, λ_{exc} = 373 nm).

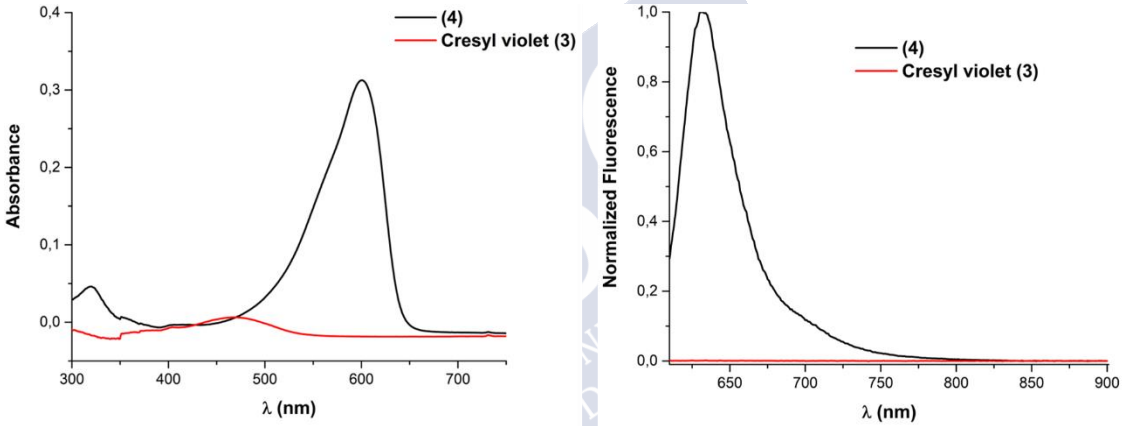


Figure 154. *Left*: Absorption spectra of **3** and 4 (20 µM, 7:3 v/v DMSO / H_2O). *Right*: Fluorescence spectra of **3** and 4 (5 µM, 7:3 v/v DMSO / H_2O, λ_{exc} = 600 nm).

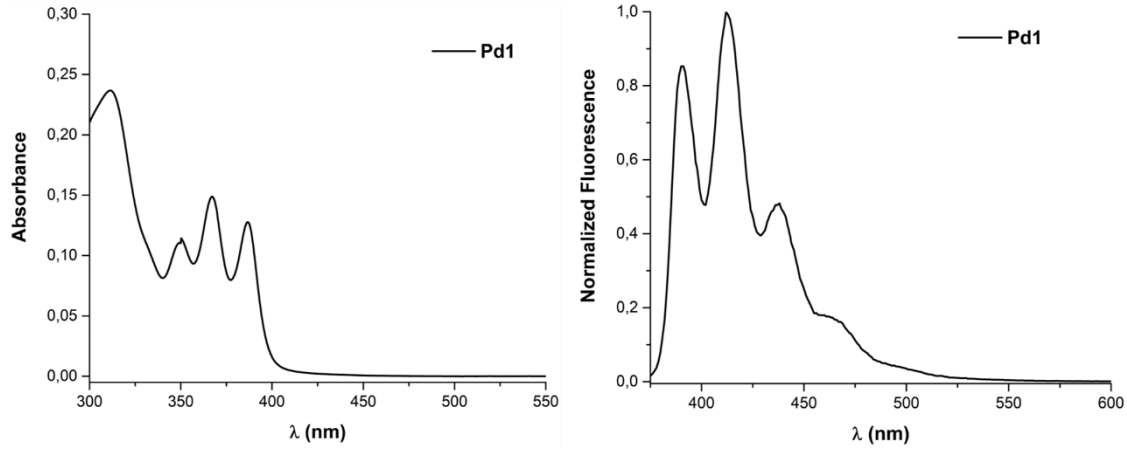


Figure 155. *Left*: Absorption spectra of **Pd1** (20 µM, 7:3 v/v DMSO / H_2O). *Right*: Fluorescence spectra of **Pd1** (5 µM, 7:3 v/v DMSO / H_2O, λ_{exc} = 350 nm).

NMR Spectra

NMR Spectra of compound D

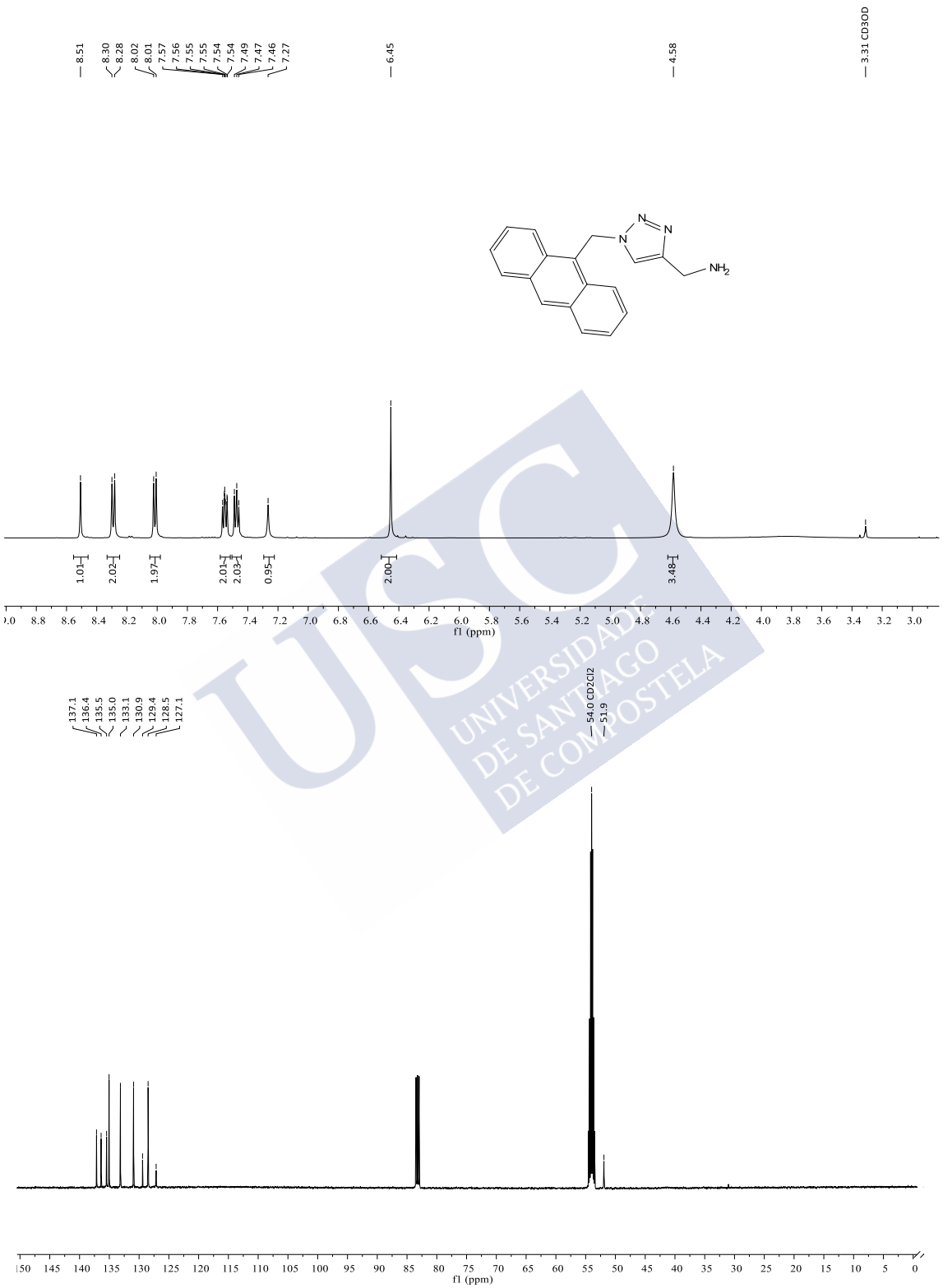


Figure 156. NMR spectra of compound **D.** ^{1}H (top) and ^{13}C (bottom).

NMR Spectra of compound E

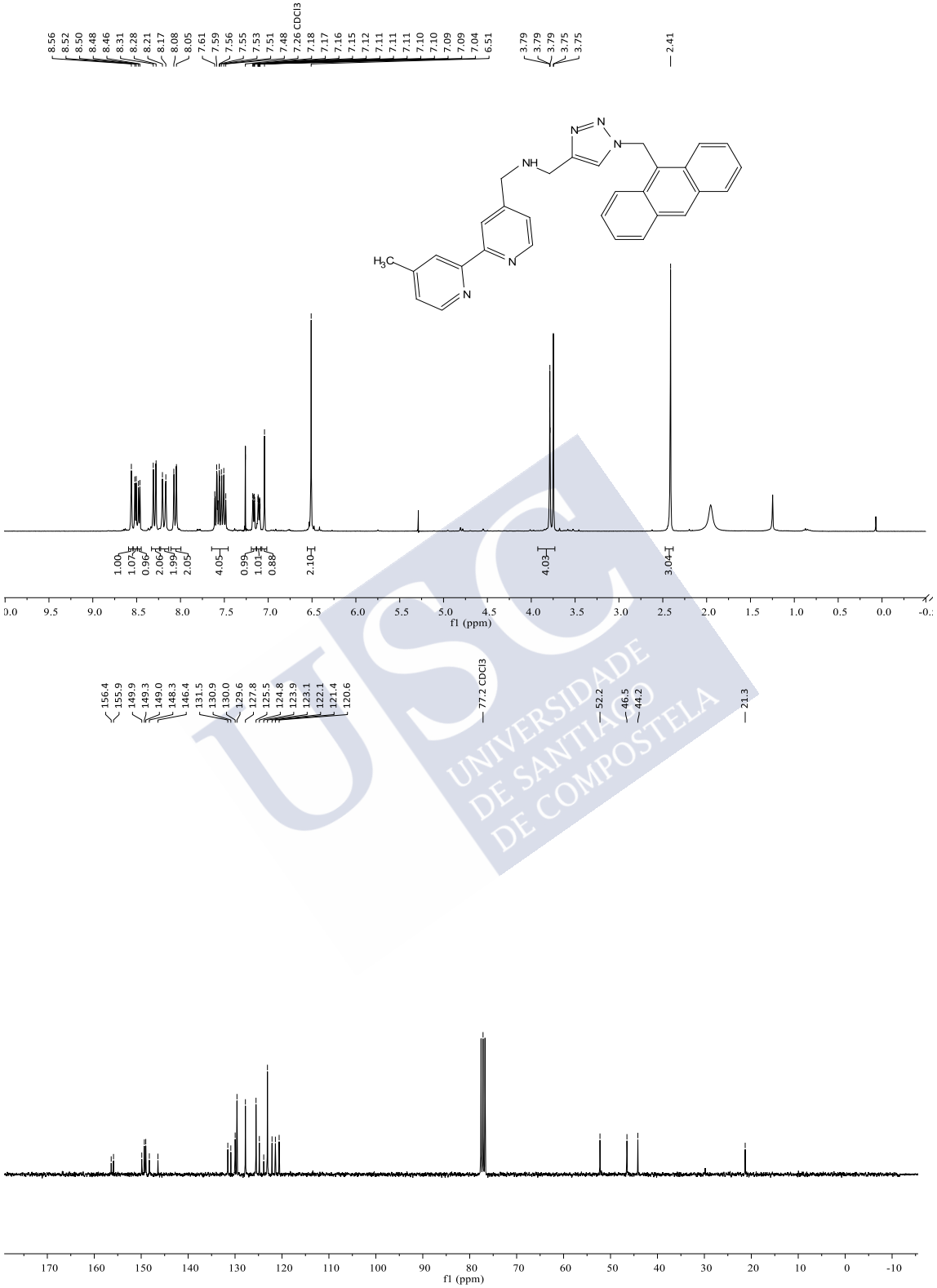


Figure 157. NMR spectra of compound **E.** ^{1}H (top) and ^{13}C (bottom).

NMR Spectra of complex Pd1

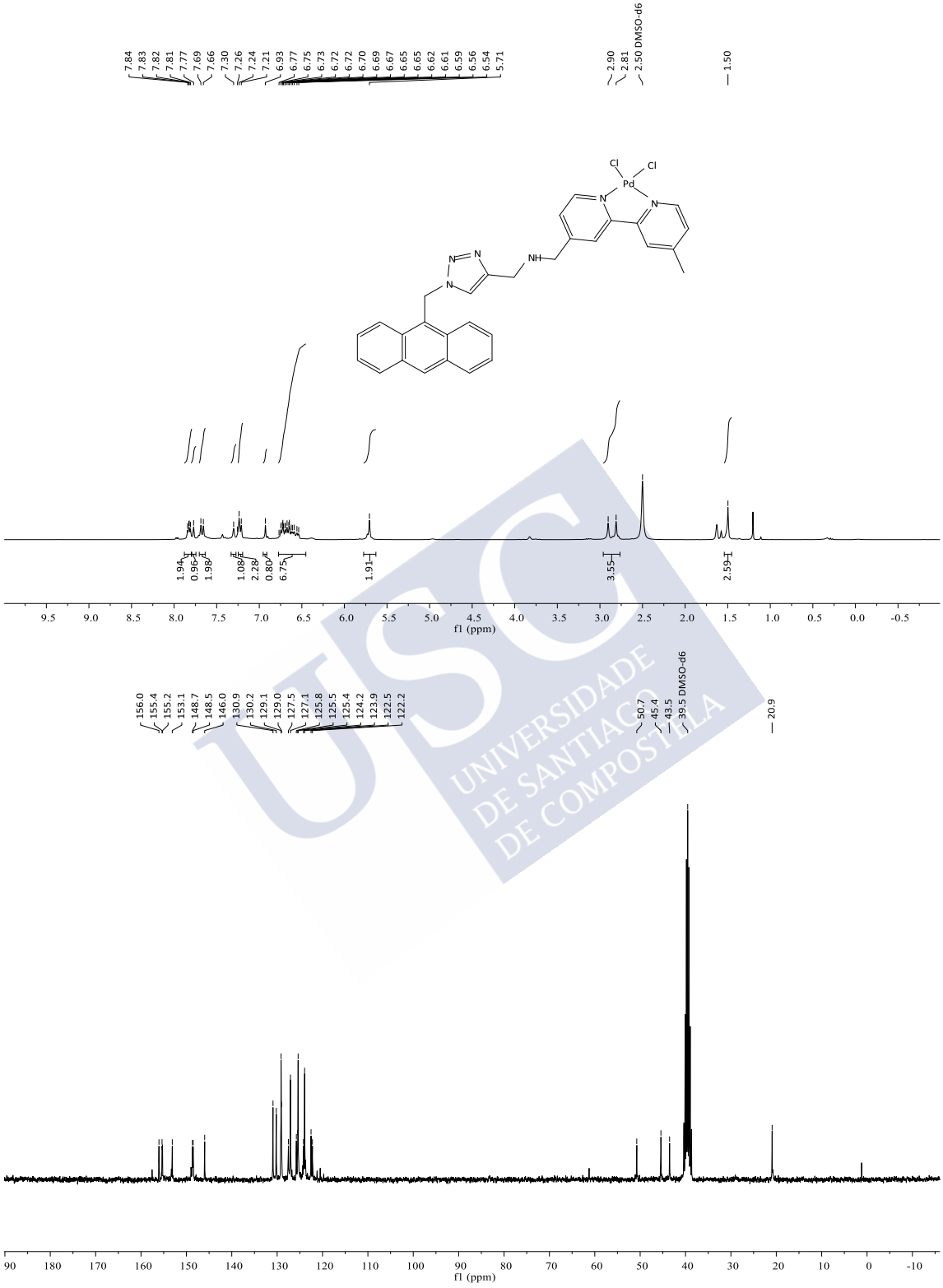
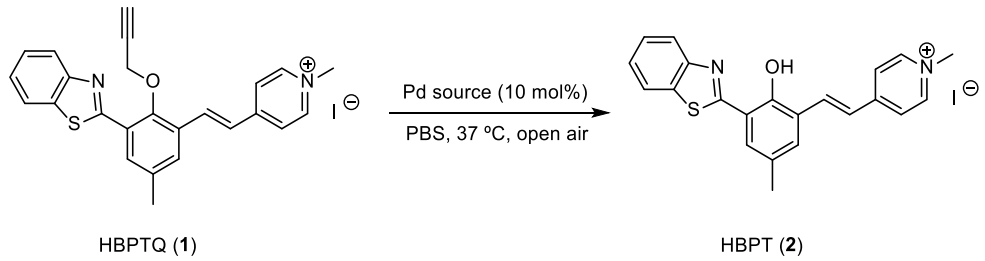


Figure 158. NMR spectra of compound **Pd1.** ^{1}H (top) and ^{13}C (bottom).

In vitro reactivity



HBPTQ (**1**) HBPT (**2**)

The catalytic deprotection of HBPTQ (**1**) to release the uncaged product **2** was performed in a 2.0 mL HPLC-vial with screw cap. For this purpose, a fresh solution of **1** (10 µL, 20 mM in DMSO, 1.0 eq.) was added to PBS (990 µL), and to the resulting solution was added another solution of the palladium complex (1 µL, 20 mM in DMSO, 0.1 eq.). The reaction mixture was kept for 24 h at 37 °C under stirring at 1000 rpm. After that time, 50 µL of the reaction were taken, diluted to 100 µL with MeOH, and analyzed by RP-HPLC-MS. The results were treated according to the calibration curve, in which coumarin was used as internal standard. Every value is the average value of two independent measurements.

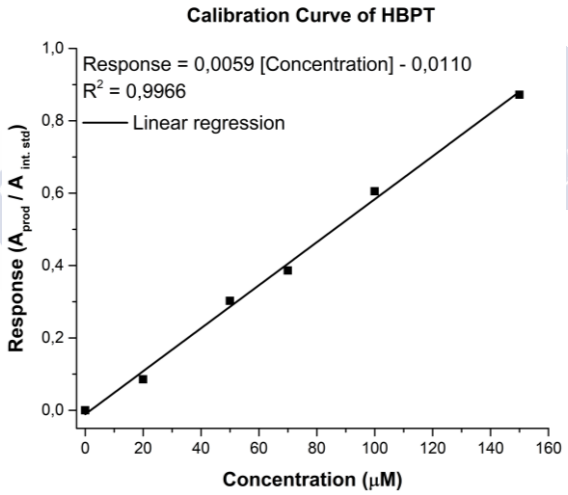


Figure 159. Calibration curve for compound **2.**

Intracellular reactions

HeLa, A549, Vero or MCF-7 cells were seeded on glass-bottom plates 48 h before treatment. Culture medium was removed and FBS-DMEM containing probe **1** (50 µM) or **3** (75 µM) was added. After incubation for 30 minutes, cells were washed twice with FBS-DMEM and a solution of palladium sources or palladopeptides (peptides were premixed with the palladium sources (1:1) in water for 10 min before the addition to cells) in FBS-DMEM was added. After 1-hour incubation, cells were washed twice in PBS and replaced with fresh FBS-DMEM to observe under the microscope with adequate filters. Digital

pictures of the different samples were taken under identical conditions of gain and exposure.

CTFC measurements

We have processed the different images of internalization experiments with Image J program, obtaining a mean value of corrected total cell fluorescence (CTFC) considering the TMR fluorophore. In addition, we have also processed the different images of intracellular reactions experiments, obtained a mean value of corrected total cell fluorescence (CTFC) regarding HBTP (**2**).

ICP-MS

1.000.000 cells per well were seeded in 100 mm plated two days before treatment with the palladopeptide resulting from brHis_2 and $PdCl_2$(COD) or only with $PdCl_2$(COD), in FBS-DMEM for 2 hours. Cells were then washed twice with PBS and lysed in 70% HNO_3. The obtained lysates were digested in duplicate with HNO_3 / H_2O_2 by heating with microwave energy before being analyzed.

Chapter IV

Plasmid Constructions

All constructs use in the Chapter are listed in the Table 1, and the nucleotide and amino acid sequences are provided in Table 2.

The mammalian expression HaloTag® pHT2 Vector was purchased from *Promega*. Using classic molecular cloning methods (PCR amplification, digestions and ligations) were introduced the different targeting sequences to organelles. All of them were expressed from the CMV promoter of the pHT2 vector.

pETDuet_1 Vector used for obtaining purified proteins from bacteria expression, was acquired from *AddGene*. Using molecular cloning methods, the HALO-tag sequence was added and also the Poly-His region, in order to have a tag to purify the protein.

Both pHT2-HALO-mutated(D106) and pETDuet_1-HALO-mutated(D106) were obtained by mutagenesis of the sequence by PCR, from the pHT2-HALO vector and pETDuet_1-HALO, respectively.

Backbone/plasmid name	Description
Protein coding plasmids	
pHT2/CMV:HALO	Plasmid ensures expression of the HALO protein in the cytosol of mammalian cells.
pHT2/CMV:HALO(D106A)	Plasmid ensures expression of the HALO-mutated(D106A) protein in the cytosol of mammalian cells.
pHT2/CMV:HALO:ER	Plasmid ensures expression of the HALO protein in the endoplasmic reticulum (ER) of mammalian cells.
pHT2/CMV:HALO:Nucleus	Plasmid ensures expression of the HALO protein in the nucleus of mammalian cells.
pHT2/CMV:HALO:Membrane	Plasmid ensures expression of the HALO protein in the plasma membrane of mammalian cells.
pETDuet_1/T7:HALO:Histag	Plasmid ensure expression of the HALO protein fused to His-tag in bacteria, under T7 promoter and lacO regulator.
pETDuet_1/T7:HALO(D106A):Histag	Plasmid ensure expression of the HALO-mutated(D106A) protein fused to His-tag in bacteria, under T7 promoter and lacO regulator.

Table 3. Plasmids used in this chapter.

Nucleotide/ amino acid sequence	
P_{CMV}	CMV promoter from plasmid vector pHT2
	tcaatattggccattagccatattattcattggttatatagcataaatcaatattggctattggccattgcatacgttgtatctatatcataatatgtacatttatattggctcatgtccaatat gaccgccatgttgg cattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgccc aacgacccccgcccatt gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtc cgccccc tattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttacgggactttcctacttggcagtacatctacgtattagtcatcgctatta ccatggtgatgcggttttggcagtaca ccaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaataaccccgcc ccgttg acgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatc
HALO	Halo protein from plasmid pHT2
	ggatccgaaatcggtacaggcttccccttcgaccccattatgtggaagtcctgggcgagcgtatgcactacgtcgatgttggaccgcgggatggcacgcctgtgctgttcctgcacggtaac ccgacctcgtcctacctgtggcgcaacatcatcccgcatgtagcaccgagtcatcggtgcattgctccagacctgatcgggatgggaaaatcggacaaaccagacctcgattat ttcttcgacg accacgtccgctacctcgatgccttcatcgaagccttgggtttggaagaggtcgtcctggtcatccacgactggggctcagctctcggattccactgggccaagcgcaatccggaacgggtca aaggtattgcatgtatggaattcatccggcctatcccgacgtgggacgaatggccagaattcgcccgtgagaccttccaggccttccggaccgccgacgtcggccgagagttgatcatcgatc agaacgctttcatcgagggtgcgctcccgatgggggtcgtccgtccgcttacggaggtcgagatggaccactatcgcgagcccttcctcaagcctgttgaccgagagccactgtggcgattcc ccaacgagctgcccatcgccggtgagcccgcgaacatcgtcgcgctcgtcgaggcatacatgaactggctgcaccagtcacctgtcccgaagttgttgttctggggcacacccggcgtactg atccccccggccgaagccgcgagacttgccgaaagcctccccaactgcaagacagtggacatcggcccgggattgttcttgctccaggaagacaacccggaccttatcggcagtgagatcg cgcgctggctccccgggctggccggctaa GSEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYLWRNIIPHVAPSHRCIAPDLIGMGKSDKPDLDYFFDDHVRYLDAFIEALGLEEVV LVIHDWGSALGFHWAKRNPERVKGIACMEFIRPIPTWDEWPEFARETFQAFRTADVGRELIIDQNAFIEGALPMGVVRPLTEVEMDHYREPFLKPVDREP LWRFPNELPIAGEPANIVALVEAYMNWLHQSPVPKLLFWGTPGVLIPPAEAARLAESLPNCKTVDIGPGLFLLQEDNPDLIGSEIARWLPGLAG
HALO(D106A)	Halo mutated protein (D106A) after mutagenesis
	ggatccgaaatcggtacaggcttccccttcgaccccattatgtggaagtcctgggcgagcgtatgcactacgtcgatgttggaccgcgggatggcacgcctgtgctgttcctgcacggtaac ccgacctcgtcctacctgtggcgcaacatcatcccgcatgtagcaccgagtcatcggtgcattgctccagacctgatcgggatgggaaaatcggacaaaccagacctcgattatttcttcgacg accacgtccgctacctcgatgccttcatcgaagccttgggtttggaagaggtcgtcctggtcatccacgCctggggctcagctctcggattccactgggccaagcgcaatccggaacgggtca aaggtattgcatgtatggaattcatccggcctatcccgacgtgggacgaatggccagaattcgcccgtgagaccttccaggccttccggaccgccgacgtcggccgagagttgatcatcgatc agaacgctttcatcgagggtgcgctcccgatgggggtcgtccgtccgcttacggaggtcgagatggaccactatcgcgagcccttcctcaagcctgttgaccgagagccactgtggcgattcc ccaacgagctgcccatcgccggtgagcccgcgaacatcgtcgcgctcgtcgaggcatacatgaactggctgcaccagtcacctgtcccgaagttgttgttctggggcacacccggcgtactg atccccccggccgaagccgcgagacttgccgaaagcctccccaactgcaagacagtggacatcggcccgggattgttcttgctccaggaagacaacccggaccttatcggcagtgagatcg cgcgctggctccccgggctggccggctaa GSEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYLWRNIIPHVAPSHRCIAPDLIGMGKSDKPDLDYFFDDHVRYLDAFIEALGLEEVV LVIH**A**WGSALGFHWAKRNPERVKGIACMEFIRPIPTWDEWPEFARETFQAFRTADVGRELIIDQNAFIEGALPMGVVRPLTEVEMDHYREPFLKPVDREP LWRFPNELPIAGEPANIVALVEAYMNWLHQSPVPKLLFWGTPGVLIPPAEAARLAESLPNCKTVDIGPGLFLLQEDNPDLIGSEIARWLPGLAG
ER	Igk_secretion peptide; introduced into construct with PCR in the N-terminal of HALO protein
	atggagacagacactcctgctatgggtactgctgctctgggttccaggttccactggtgac METDTLLLWVLLLWVPGSTGD
NLS	SV40 nuclear localization signal; introduced in the N-terminal of HALO protein by PCR
	gatccaaaaaagaagagaaaggta PKKKRKV
MEMBRANE	Igk_secretion peptide (N-terminal) and Transmembrane fragment (C-terminal), introduced both into construct by PCR and fused with HALO protein
	Transmembrane fragment → aatgctgtgggccaggacacgcaggaggtcatcgtggtgccacactccttgcccttaaggtggtggtgatctcagccatcctggccctggtggtg ctcaccatcatctcccttatcatcctcatcatgctttggcagaagaagccacgttag NAVGQDTQEVIVVPHSLPFKVVVISAILALVVLTIISLIILIMLWQKKPR
T7	T7 promoter from plasmid vector pETDuet_1
	taatacgactcactatag
lacO	Lac Operator, regulator binding site
	gggaattgtgagcggataacaattcccc
His-tag	Histidine sequence introduced in the N-terminal of HALO protein by PCR
	catcaccatcaccatcac HHHHHH

Table 4. Nucleotide and amino acid sequences of the constructs used in the study.

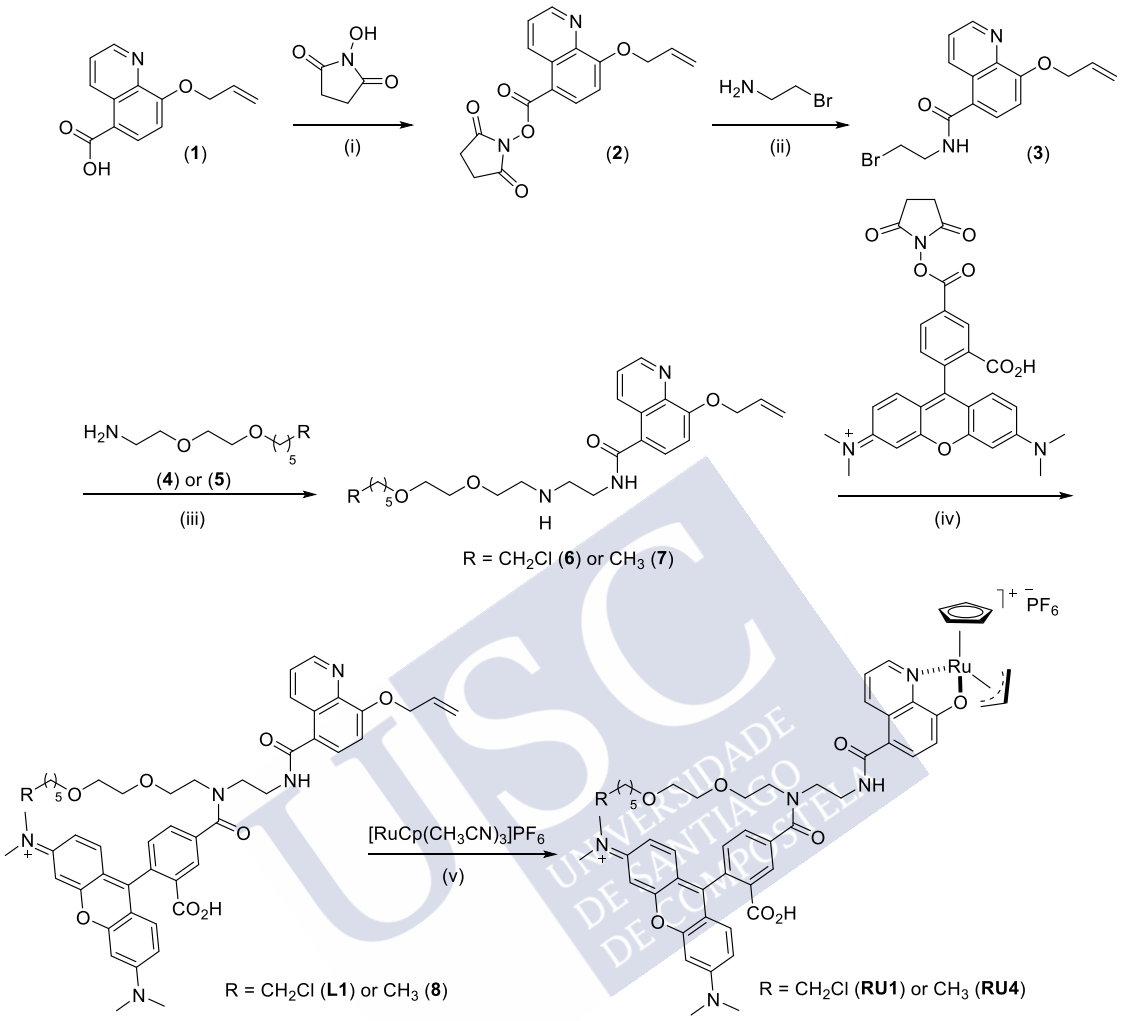
Protein Expression and Purification

BL21 CodonPlus-RP (DE3) bacteria transformed with the expression plasmid were diluted and incubated at 37 °C with shaking to reach at OD_{600}~ 0.4-0.6. Expression was induced with 1 mM IPTG and incubated at 37 °C with shaking for 3 h. Induced cultures were centrifuged at 3200 g and the pellet washed twice with PBS. Then, it resuspended in 1/10 volume of lysis buffer and after centrifugation, the pellet was frozen until use. Then, protease inhibitor (EDTA free) and lysozyme (1mg/mL) were added and the pellet was sonicated for 6´ (35"on /10"off).

Protein was purified over an IMAC (immobilized metal affinity chromatography) following HisPur™ Ni-NTA protocols (*Thermo fisher*). Protein was concentrated in a 10-kDa Amicon Ultra-15 (*Millipore*) aliquoted with 15% glycerol, and then flash-frozen and stored at -20ºC.

Ligands synthesis

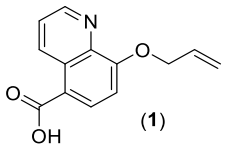
*Synthesis of the ruthenium complexes **RU1**, **RU4** and **L1***



Scheme 1. Synthesis of the ruthenium complexes **RU1** and **RU4, and L1**. Reagents: (i) **1** (1.0 eq.), N-hydroxysuccinimide (1.2 eq.), EDC (1.2 eq.), N,N-dimethylformamide, 0 °C, 1 h, then 12 h, room temperature; (ii) 2-Bromoethane-1-amine (1.1 eq.), DIEA (2.0 eq.), N,N-dimethylformamide, 24 h, room temperature; (iii) **4** or **5** (1.0 eq.), K_2CO_3 (1.0 eq.) in water, dichloromethane, 24 h, room temperature; (iv) **6** or **7** (1.0 eq.), 5-TAMRA Se (1.3 eq.), DIEA (9.0 eq.), N,N-dimethylformamide, 48 h, room temperature; (v) **L1** or **8** (1.0 eq.), $[RuCp(CH_3CN)_3]PF_6$ (1.0 eq.), dichloromethane, 12 h, room temperature.

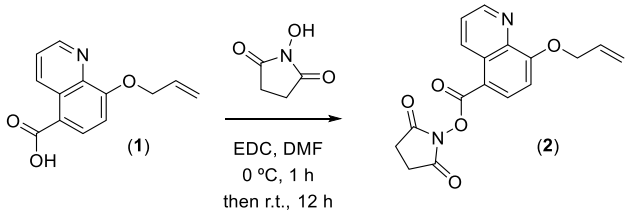
- *Synthesis of the ligands* **L1** *and* **8***:*

*Synthesis of 8-(allyloxy)quinoline-5-carboxylic acid (**1**)*



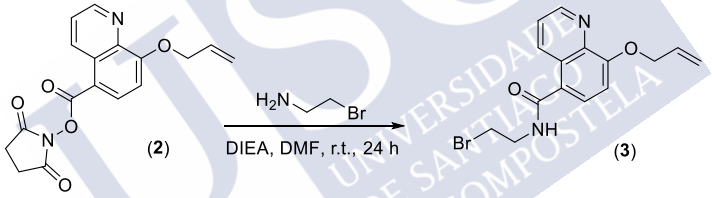
Compound **1** was synthesized according to procedure described by Meggers *et al.*[260] and spectroscopic data agreed with the reported values.

Synthesis of 2,5-dioxopyrrolidin-1-yl 8-(allyloxy)quinoline-5-carboxylate (**2**)



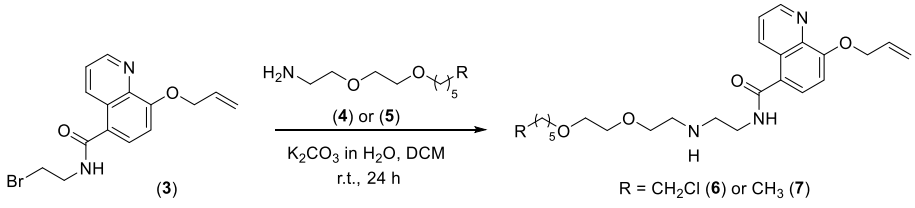
8-(Allyloxy)quinoline-5-carboxylic acid (**1**) (150.0 mg, 0.65 mmol) was dissolved in N,N-dimethylformamide (4.0 mL). 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (150.5 mg, 1.2 eq.) and N-hydroxysuccinimide (92.4 mg, 1.2 eq.) were subsequently added to the solution at 0 °C and stirred for 1 h. The reaction mixture was left stirring overnight at room temperature and the solvent was removed under reduced pressure. The crude was diluted with 20.0 mL of water and the product was extracted with dichloromethane (2 x 30.0 mL). The organic phases were dried and concentrated in vacuum. The crude was used in the next step without further purification.

Synthesis of 8-(allyloxy)-N-(2-bromoethyl)quinoline-5-carboxamide (**3**)



2,5-Dioxopyrrolidin-1-yl 8-(allyloxy)quinoline-5-carboxylate (**2**) (125.0 mg, 0.38 mmol) was dissolved in N,N-dimethylformamide (4.0 mL). 2-Bromoethane-1-amine hydrobromide (86.4 mg, 1.1 eq.) and diisopropylethylamine (33.4 µL, 2.0 eq.) were subsequently added to the solution. The reaction mixture was left stirring 24 h and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel.

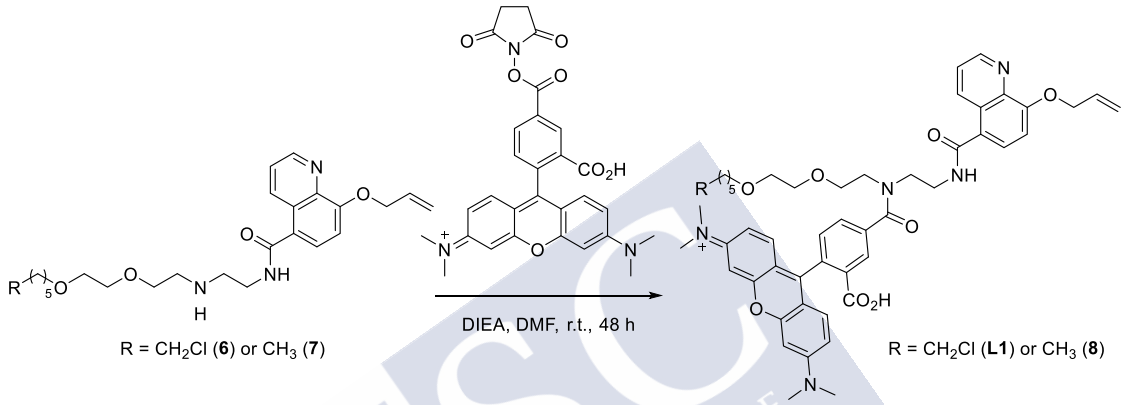
Synthesis of 8-(allyloxy)-N-(2-((2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)amino)ethyl) quinoline-5-carboxamide (**6**) *and 8-(allyloxy)-N-(2-((2-(2-(hexyloxy)ethoxy)ethyl)amino) ethyl)quinoline-5-carboxamide* (**7**)



[260] T. Völker, E. Meggers, *ChemBioChem.* **2017**, *18*, 1083-1086.

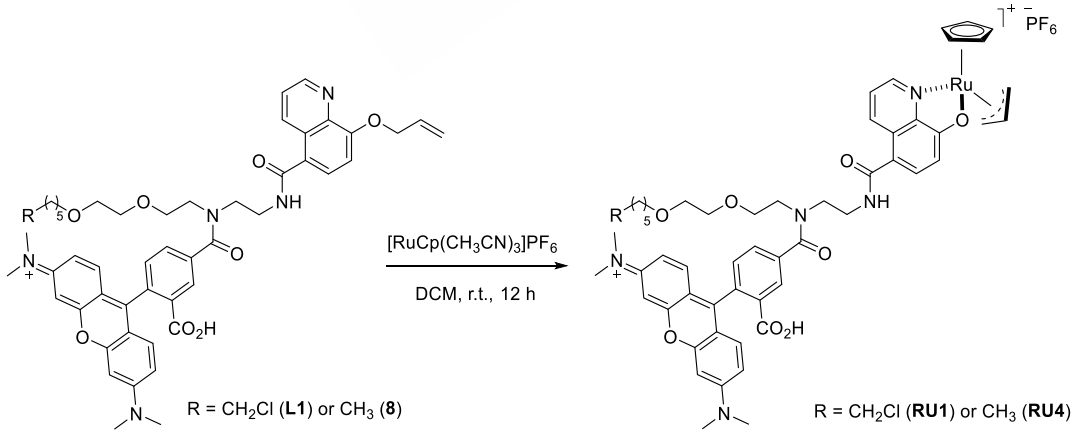
8-(Allyloxy)-N-(2-bromoethyl)quinoline-5-carboxamide (**3**) (64.0 mg, 0.19 mmol) was dissolved in dichloromethane (2.0 mL). Substrate **4** or **5** (1.0 eq.) was added, followed by a suspension of K_2CO_3 in water (26.3 mg, 1 eq.). The reaction mixture was left stirring 24 h and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (DCM/MeOH 9:1).

Synthesis of N-(9-(4-((2-(8-(allyloxy)quinoline-5-carboxamido)ethyl)(2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)carbamoyl)-2-carboxyphenyl)-6-(dimethylamino)-3H-xanthen-3-ylidene)-N-methylmethanaminium (**L1**) *or N-(9-(4-((2-(8-(allyloxy)quinoline-5-carboxamido)ethyl)(2-(2-(hexyl)oxy)ethoxy)ethyl)carbamoyl)-2-carboxyphenyl)-6-(dimethylamino)-3H-xanthen-3-ylidene)-N-methylmethanaminium* (**8**)



Substrate **6** or **7** (0.04 mmol) was dissolved in N,N-dimethylformamide (1.5 mL). 5-Carboxy-tetramethylrhodamine *N*-succinimidyl ester (28.1 mg, 1.3 eq.) and diisopropylethylamine (63.0 µL, 9 eq.) were added. The reaction mixture was left stirring for two days and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (DCM/MeOH 9:1).

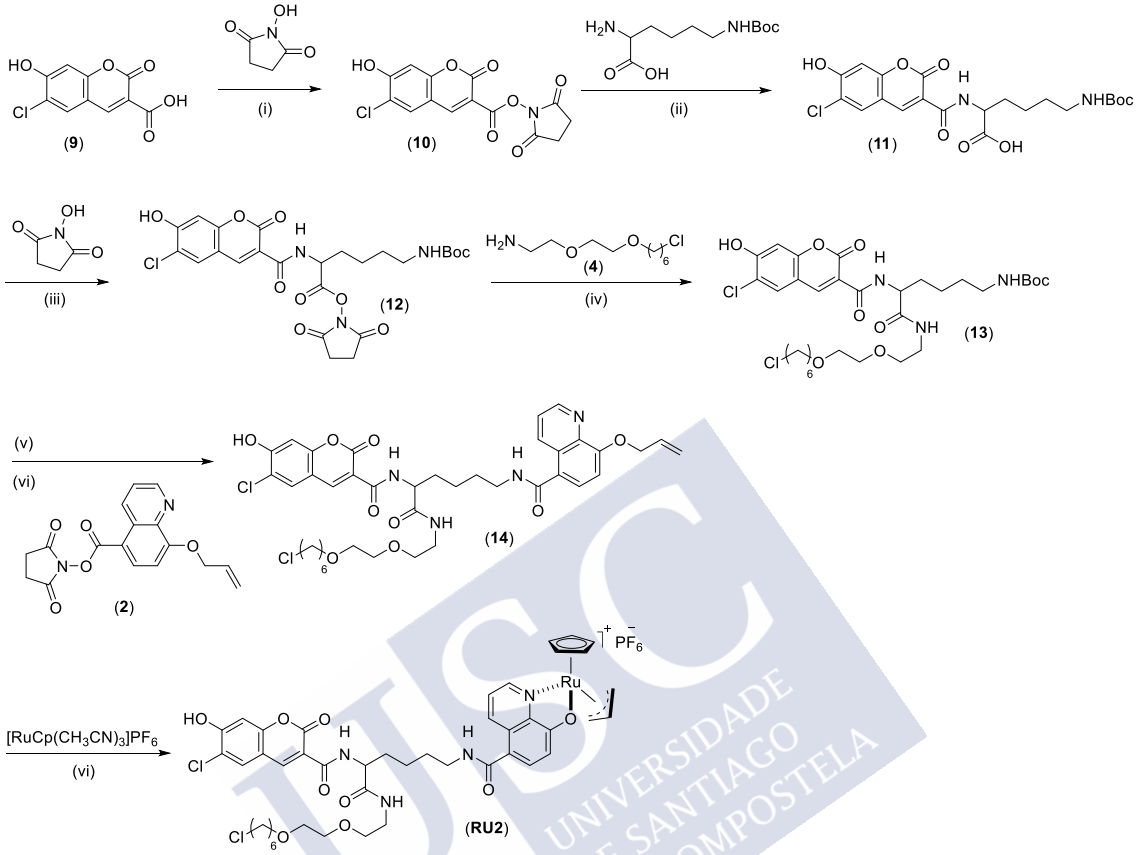
- *Synthesis of the ruthenium complexes* **RU1** *and* **RU4**



To a 10 mM solution of **L1** or **8** (0.02 mmol) in anhydrous dichloromethane (0.17 mL) was added 1.0 eq. of $[RuCp(CH_3CN)_3]PF_6$ (0.02 mmol). The reaction mixture was left stirring overnight at room temperature. After that time, it was centrifuged and the supernatant

was discarded. The solid was washed with cold dichloromethane and dried under reduced pressure to afford the final complex **RU1** or **RU4** as a dark solid.

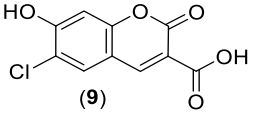
Synthesis of ruthenium complex **RU2**



Scheme 2. Synthesis of the ruthenium complex **RU2**. Reagents: (i) **9** (1.0 eq.), N-hydroxysuccinimide (1.2 eq.), EDC (1.2 eq.), N,N-dimethylformamide, 0 °C, 1 h, then 12 h, room temperature; (ii) H-Lys(Boc)-OH (1.1 eq.), DIEA (2.0 eq.), N,N-dimethylformamide, 24 h, room temperature; (iii) **11** (1.0 eq.), N-hydroxysuccinimide (1.2 eq.), EDC (1.2 eq.), N,N-dimethylformamide, 0 °C, 1 h, then 12 h, room temperature; (iv) **4** (1.0 eq.), K_2CO_3 (1.0 eq.) in water, dichloromethane, 24 h, room temperature; (v) TFA (7.0 eq.), dichloromethane, 2 h; (vi) **2** (1.0 eq.), DIEA (9 eq.), N,N-dimethylformamide, 48 h, room temperature; (vii) **14** (1.0 eq.), $[RuCp(CH_3CN)_3]PF_6$ (1.0 eq.), dichloromethane, 12 h, room temperature.

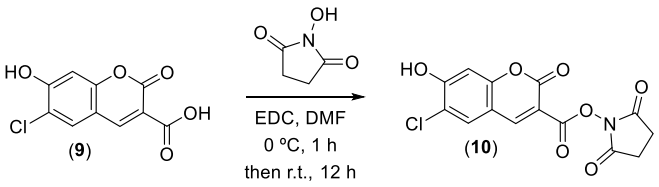
- *Synthesis of the ligand* **14***:*

Synthesis of 6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carboxylic acid (**9**)



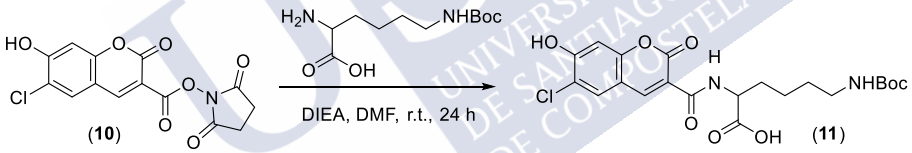
Compound **9** was synthesized according to procedure described by Goeldner *et al.*[261] and spectroscopic data agreed with the reported values.

Synthesis of 2,5-dioxopyrrolidin-1-yl 6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carboxylate (**10**)



6-Chloro-7-hydroxy-2-oxo-2H-chromene-3-carboxylic acid (**9**) (150.0 mg, 0.62 mmol) was dissolved in N,N-dimethylformamide (4.0 mL). 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (150.5 mg, 1.2 eq.) and N-hydroxysuccinimide (92.4 mg, 1.2 eq.) were subsequently added to the solution at 0 °C and stirred for 1 h. The reaction mixture was left stirring overnight at room temperature and the solvent was removed under reduced pressure. The crude was diluted with 20.0 mL of water and the product was extracted with dichloromethane (2 x 30.0 mL). The organic phases were dried and concentrated in vacuum. The crude was used in the next step without further purification.

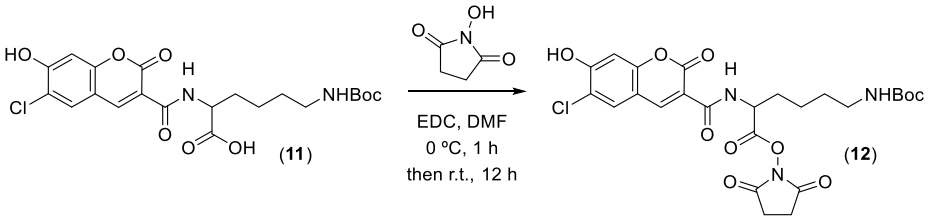
Synthesis of N^6-(tert-butoxycarbonyl)-N^2-(6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carbonyl)lysine (**11**)



2,5-Dioxopyrrolidin-1-yl 6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carboxylate (**10**) (209.4 mg, 0.62 mmol) was dissolved in N,N-dimethylformamide (4.0 mL). H-Lys(Boc)-OH (168.0 mg, 1.1 eq.) and diisopropylethylamine (216.0 µL, 2.0 eq.) were subsequently added to the solution. The reaction mixture was left stirring 24 h and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel.

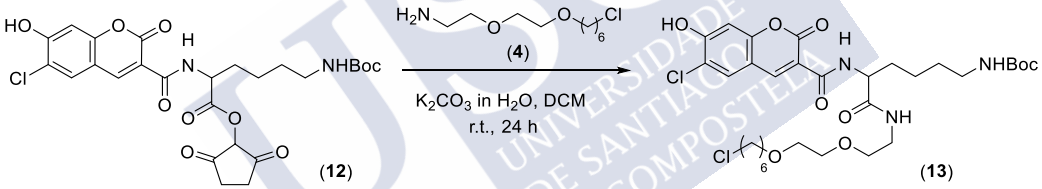
[261] C. Orange, A. Specht, D. Puliti, E. Sakr, T. Furuta, B. Winsor, M. Goeldner, *Chem. Commun.* **2008**, *10*, 1217-1219.

Synthesis of 2,5-dioxopyrrolidin-1-yl N^6-(tert-butoxycarbonyl)-N^2-(6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carbonyl)lysinate (**12**)



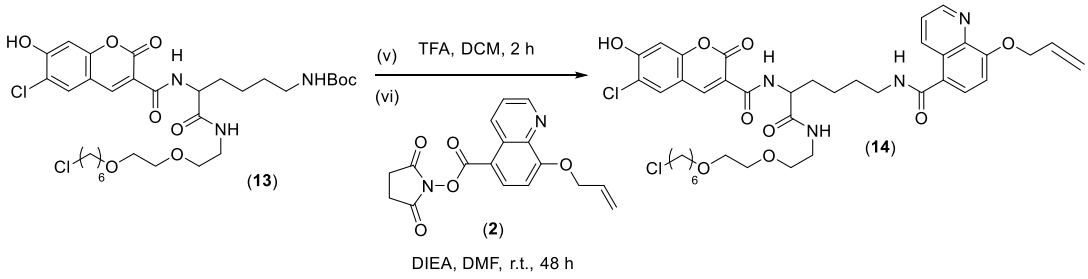
N^6-(tert-butoxycarbonyl)-N^2-(6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carbonyl)lysine (**11**) (154.7 mg, 0.33 mmol) was dissolved in N,N-dimethylformamide (4.0 mL). 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (75.3 mg, 1.2 eq.) and N-hydroxysuccinimide (46.2 mg, 1.2 eq.) were subsequently added to the solution at 0 °C and stirred for 1 h. The reaction mixture was left stirring overnight at room temperature and the solvent was removed under reduced pressure. The crude was diluted with 20.0 mL of water and the product was extracted with dichloromethane (2 x 30.0 mL). The organic phases were dried and concentrated in vacuum. The crude was used in the next step without further purification.

Synthesis of tert-butyl (5-(6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carboxamido)-6-((2-(2-(chloromethoxy)ethoxy)ethyl)amino)-6-oxohexyl)carbamate (**13**)



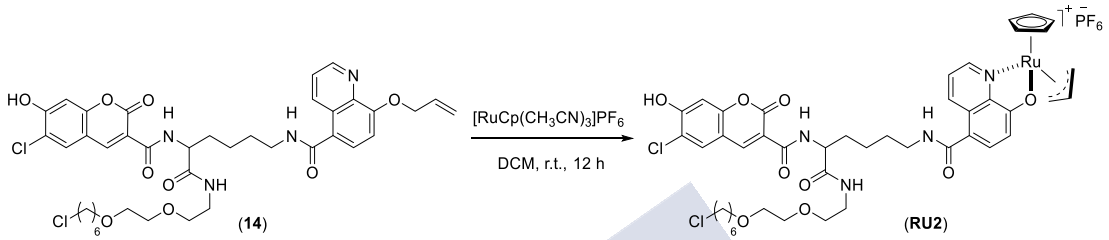
2,5-Dioxopyrrolidin-1-yl N^6-(tert-butoxycarbonyl)-N^2-(6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carbonyl)lysinate (**12**) (100.0 mg, 0.18 mmol) was dissolved in dichloromethane (3.0 mL). Substrate **4** (1.0 eq.) was added, followed by a suspension of K_2CO_3 in water (24.9 mg, 1 eq.). The reaction mixture was left stirring 24 h and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (DCM/MeOH 9:1).

Synthesis of 8-(allyloxy)-N-(5-(6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carboxamido)-6-((2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)amino)-6-oxohexyl)quinoline-5-carboxamide(**14**)



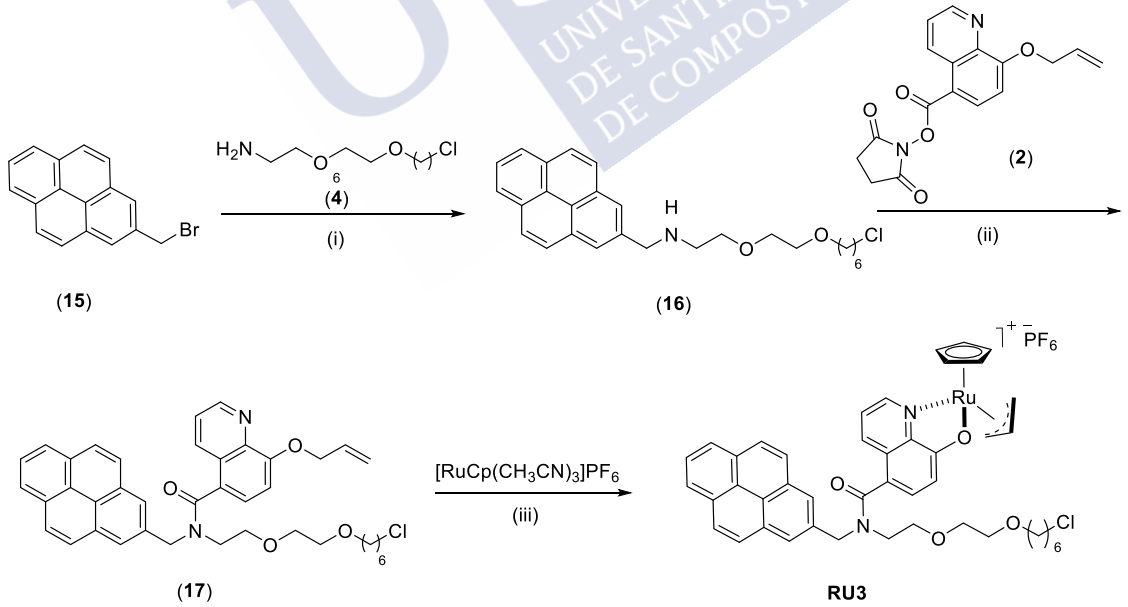
tert-Butyl (5-(6-chloro-7-hydroxy-2-oxo-2H-chromene-3-carboxamido)-6-((2-(2-(chloromethoxy)ethoxy)ethyl)amino)-6-oxohexyl)carbamate (**13**) (86.4 mg, 0.14 mmol, 1.1 eq.) was dissolved in dichloromethane and TFA (7 eq.) was added and the mixture was stirred for 2 h. The crude of reaction was dissolved in N,N-dimethylformamide (2.0 mL) and 2,5-dioxopyrrolidin-1-yl 8-(allyloxy)quinoline-5-carboxylate (**2**) (42.4 mg, 0.13 mmol) and diisopropylethylamine (45.3 µL, 2.0 eq.) were subsequently added to the solution. The reaction mixture was left stirring for 48 h and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel.

- *Synthesis of the ruthenium complex* **RU2**



To a 10 mM solution of **14** (0.02 mmol) in anhydrous dichloromethane (0.17 mL) was added 1.0 eq. of $[RuCp(CH_3CN)_3]PF_6$ (0.02 mmol). The reaction mixture was left stirring overnight at room temperature. After that90 time, it was centrifuged, the supernatant was discarded. The solid was washed with cold dichloromethane and dried under reduced pressure to afford the final complex **RU2** as a dark solid.

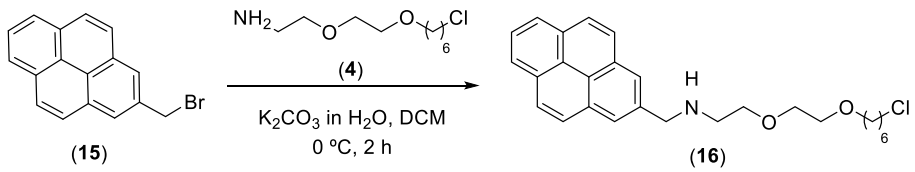
Synthesis of the ruthenium complex **RU3**



Scheme 3. Synthesis of the ruthenium complex **RU3**. Reagents: (i) **4** (1.0 eq.), K_2CO_3 (1.0 eq.) in water, dichloromethane, 2 h, 0 °C; (ii) **2** (1.0 eq.), DIEA (9 eq.), N,N-dimethylformamide, 48 h, room temperature; (iii) **17** (1.0 eq.), $[RuCp(CH_3CN)_3]PF_6$ (1.0 eq.), dichloromethane, 12 h, room temperature.

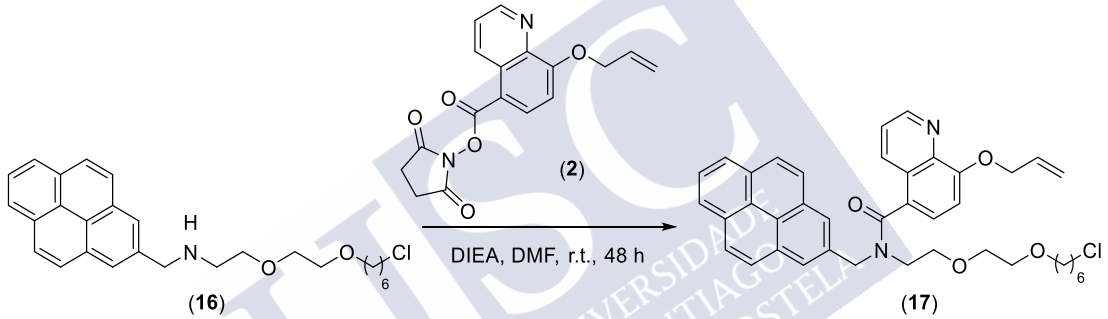
- *Synthesis of the ligand* **17**

*Synthesis of 2-(2-((6-chlorohexyl)oxy)ethoxy)-N-(pyren-2-ylmethyl)ethan-1-amine (***16***)*



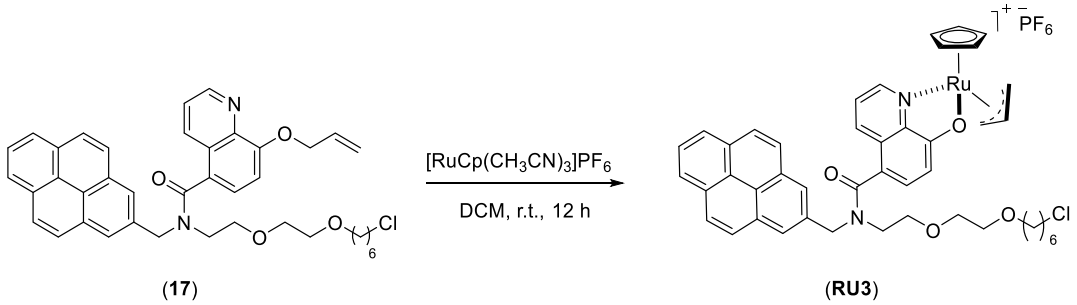
1-(Bromomethyl)pyrene (1) (56.1 mg, 0.19 mmol) was dissolved in dichloromethane (2.0 mL). Substrate **4** (1.0 eq.) was added at 0 °C, followed by a suspension of K_2CO_3 in water (26.3 mg, 1 eq.). The reaction mixture was left stirring 2 h at 0 °C and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (DCM/MeOH 9:1).

*Synthesis of 8-(allyloxy)-N-(2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)-N-(pyren-2-ylmethyl)quinoline-5-carboxamide (***17***)*



2-(2-((6-Chlorohexyl)oxy)ethoxy)-N-(pyren-2-ylmethyl)ethan-1-amine (**16**) (56.8 mg, 0.13 mmol) was dissolved in N,N-dimethylformamide (2.0 mL) and 2,5-dioxopyrrolidin-1-yl 8-(allyloxy)quinoline-5-carboxylate (**2**) (42.4 mg, 0.13 mmol) and diisopropylethylamine (45.3 µL, 2.0 eq.) were subsequently added to the solution. The reaction mixture was left stirring for 48 h and the solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (DCM/MeOH 12:1).

- *Synthesis of the ruthenium complex* **RU3**



To a 10 mM solution of **17** (0.02 mmol) in anhydrous dichloromethane (0.17 mL) was added 1.0 equiv of $[RuCp(CH_3CN)_3]PF_6$ (0.02 mmol). The reaction mixture was left stirring overnight at room temperature. After that time, it was centrifuged and the supernatant was discarded. The solid was washed with cold dichloromethane and dried under reduced pressure to afford the final complex **RU3** as a dark solid.

In vitro assays

Before use the corresponding purified protein, the glycerol was removed using a Zeba™ Spin Desalting Columns (*Thermo fisher*) that had been equilibrated before with PBS.

Binding reaction:

To carry out the binding reaction, the proteins HALO and HALO-Mutated(D106A) were quantified using a Nanodrop Spectrophotometer and mixed at a concentration of 10 μM with the ligands (20 μM) in PBS at 37ºC in an Eppendorf ThermoMixer (*Fisher scientific*). After one hour with shaking, the proteins were concentrated using *Amicon®* Ultra Centrifugal Filters (*Millipore*), carrying out two wash steps with PBS to ensure the removal of excess ligand.

After that, the samples were measured again in the Nanodrop Spectrophotometer and mixed with Laemmli buffer getting a concentration of 20 μg of each protein/complex, and then heated 5min at 95 ºC. The samples were resolved on a 12,5% denaturing polyacrylamide gel (with SDS) using Tris-glycine-SDS buffer, over 40 min a r.t. The analysis was performed directly using a ChemiDoc MP Imaging System (*Bio-rad*) for fluorescence visualization or by staining with Coomassie Brilliant Blue for 5 min.

Reactivity assay:

After performed the biding reaction and measured the protein concentration in a Nanodrop Spectrophotometer, we carried out the reactivity assays using $Rho(alloc)_2$ as caged probe. We performed these reactions in PBS during 4h at 37ºC with shaking in the Eppendorf ThermoMixer, using 10 μM of protein-complex conjugate and 50 μM of probe. We assumed that all the protein is bound to the ruthenium complex, although the proportion of the metal is probably lower. The reactions were measured in a microtiter plate reading spectrophotometer by recording the fluorescence (Excitation Wavelength: 485nm, Emission Wavelength: 535nm). The conversion was calculated according to the calibration curve, in which the rhodamine uncaged was used as a patron.

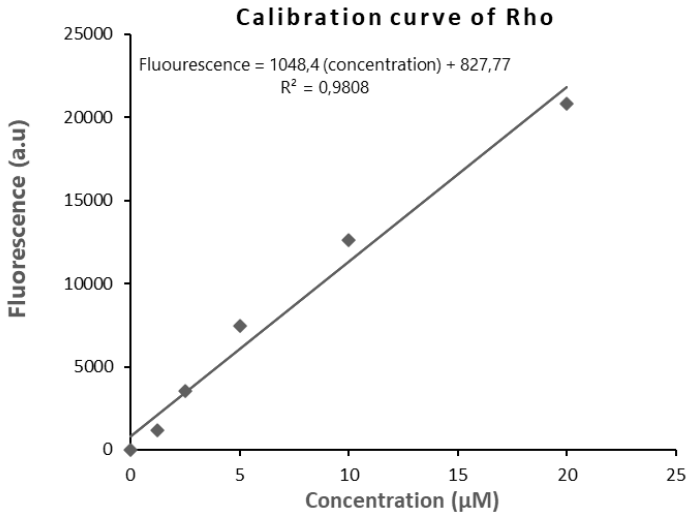


Figure 160. Calibration curve for rhodamine uncaged.

ICP/MS:

After performed the binding reactions, the samples were resolved in an SDS-PAGE. The bands observed in the gel, were cut and digested with 70% HNO_3. After that, the samples were separated in duplicates and analysed by ICP-MS.

	101 Ru	
Sample Name	**Conc. [ug/l]**	**Conc. RSD**
BLK	0,0045	34,5464
HALO-**RU1**	1,9331	0,8750
HALO-**RU4**	0,0327	1,8606
HALO-mutated(D106A)- **RU1**	0,0621	1,7424

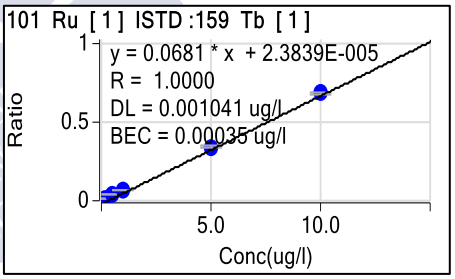


Figure 161. ICP-MS results. *Left*: Results of ICP-MS of samples digested from SDS-PAGE of one experiment, showing the ruthenium detected in each sample and its standard relative deviation (RSD). The results indicated in the chapter corresponding of mean of three independent experiments, after calculating the proportion of ruthenium per protein. *Right*: Calibration curve of ruthenium used in ICP-MS experiments.

Cell experiments

Intracellular assembly:

HeLa cells were seeded on glass-bottom plates (100.000 cells per well) 24 h before treatment. At 70% confluence, the different constructs were transfected using FuGENE® HD transfection reagent (*Promega*) for 48 h. After that, the cells were washed with PBS and fresh FBS-DMEM containing 10 µM of ruthenium ligands was added to the cells for 1 h. The cells were washed twice with PBS before being observed in a confocal microscope.

Reactivity:

After performed the intracellular assembly and washed twice with PBS, to ensure the removal of excess of ruthenium ligands no bounded, the Rho(alloc)$_2$ probe was added (50 μM) in a fresh FBS-DMEM for 4 h. The cells were washed twice with PBS before being observed in a confocal microscope.

List of Publications

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Title	Metal-Dependent DNA Recognition and Cell Internalization of Designed, Basic Peptides.
Authors	Soraya Learte Aymamí, Natalia Curado, Jéssica Rodríguez, M. Eugenio Vázquez, and José Luis Mascareñas.
Year	2017
Journal	Journal of the American Chemical Society (*JACS*)
Volume, pages…	139, 16188-16193
Impact Factor	14.357
PhD student contribution	Peptides synthesis and characterization, EMSA experiments, anisotropy and CD assays, and cells experiments.

Title	DNA-binding miniproteins based on zinc fingers. Assessment of the interaction using nanopores.
Authors	Jéssica Rodríguez, Soraya Learte Aymamí, Jesús Mosquera, Garbiñe Celaya, David Rodríguez Larrea, M. Eugenio Vázquez and José Luis Mascareñas.
Year	2018
Journal	Chemical Sicence
Volume, pages…	9, 4118-4123
Impact Factor	9.556
PhD student contribution	Synthesis of some peptides and characterization, some EMSA experiments and some anisotropies.

Title	Assembly of a Ternary Metallopeptide Complex at Specific DNA Sites Mediated by an AT-Hook Adaptor.
Authors	Soraya Learte Aymamí, Jéssica Rodríguez, M. Eugenio Vázquez, and José Luis Mascareñas.
Year	2020
Journal	Chemical European Journal
Volume, pages…	26, 8875-8878 (Hot Paper)
Impact Factor	5.439
PhD student contribution	Peptides synthesis and characterization, EMSA experiments, anisotropy and CD assays.

Title	Intracellular Reactions Promoted by Bis(histidine) Miniproteins Stapled Using Palladium(II) Complexes
Authors	Soraya Learte Aymamí, Cristian Vidal, Alejandro Gutiérrez González, and José Luis Mascareñas.
Year	2020
Journal	Angewandte Chemie International Edition
Volume, pages…	59, 9149-9154
Impact Factor	12.959
PhD student contribution	Peptides synthesis and characterization, *in vitro* assays, and cell experiments.

Title	Stimuli-Responsive DNA Binding by Synthetic Systems
Authors	Jessica Rodriguez, Jesuś Mosquera, Soraya Learte Aymamí, M. Eugenio Vaźquez, and José Luis Mascareñas
Year	2020
Journal	Accounts of chemical research
Volume, pages…	53, 2286-2298
Impact Factor	20.832
PhD student contribution	Review. Writing some parts of the manuscript.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA