Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Towards peptide-based tunable multistate memristive materials

Salvador Cardona-Serra*†, Lorena E. Rosaleny,*,† Luis Martínez-Gil‡, Silvia Giménez-Santamarina†, and Alejandro Gaita-Ariño*,†

† *Instituto de Ciencia Molecular, Universitat de València, Catedrático José Beltrán 2, E46980 Paterna, Spain*

‡ *Dept. de Bioquímica y Biología Molecular, ERI BioTecMed, Universitat de València, Dr. Moliner 50, 46100 Burjassot, Spain*

E-mail: salvador.cardona@uv.es, Lorena.Rosaleny@uv.es, alejandro.gaita@uv.es

S1 Distance and relative orientation of two lanthanide coordination centers in daLBT

For the prediction of the distance and relative orientation between two given lanthanide coordination centers in peptides coordinating two or more ions such as **daLBT**, it is useful to analyse the two existing crystal structures of double lanthanide binding tag proteins such as the ones belonging to proteins deposited in the Protein Data Bank (PDB)[1] repository with PDB ID 2OJR (dLBT-Ubiquitin) and 3VDZ (xq-dSE3-Ubiquitin). In the case of 2OJR, the lanthanide-binding part is at the N-terminal end of ubiquitin, while in 3VDZ it is inserted in a loop. Additionally, the two sequences have an important difference which in principle can affect the folding, namely the Trp residue in 2OJR is replaced with Ser in the case of 3VDZ. Other than that, in both cases, these are peptides consisting on the repetition of two identical lanthanide-binding sub-units with no spacer, like the **daLBT** we present in this work. Let us study their differences in terms of distance and relative orientation, since this is an important question for the organization of spins.

In terms of the metal-metal distance, analysis of the structures shows that it is 19.1 Å in the case of 2OJR and 21.1 Å in case of 3VDZ (see Table S1). This probably means that non-spaced LBTs, such as **daLBT**, can be consistently expected to situate lanthanides at minimum distances of about 20 ± 1 Å. Longer distances can be trivially achieved, either, by introducing the different lanthanide binding tag in loops at known distant positions in structurally characterized proteins, or by introducing bulky spacers, e.g. a poly-alanine segment which will tend to form a long α-helix.

Table S1. *Distances (metal-metal), and angles between first metal, α-carbon belonging to Tyr in the first LBT unit, and second metal, and between first metal, α-carbon belonging to Tyr in the second LBT unit, and second metal. Measurements on 2OJR and 3VDZ protein structures were performed using PyMOL[2]*

	r(Ln_1-Ln_2) (Å)	**α(Ln_1-C_α(Tyr1)-Ln_2)** (°)	**α(Ln_1-C_α(Tyr2)-Ln_2)** (°)
2OJR	19.1	133.8	119.7
3VDZ	21.1	124.7	138.5

A more challenging problem is the control of the relative orientation of two given lanthanide binding tags. In the case of 2OJR and 3VDZ, we took as reference points the position of the two metals and of the two α-carbons belonging to Tyr aminoacids. One of these Tyr is at the beginning of the first sub-unit and the other one is precisely between the two sub-units, so we measured the angle formed by Ln_1-C_α(Tyr)-Ln_2 as a rough estimate of the relative orientation of the two sub-units. As can be seen in table S1, these Ln_1-C_α(Tyr)-Ln_2 angles are in the same range, but with important variations. As can be seen in the Figure S1, local interactions (a small β-sheet can be appreciated) seem to determine the relative orientations of the two sub-units, rather than the bulkier protein they are attached to. This permits speculating that some sort of control of the orientation might be attained through careful proteic design.

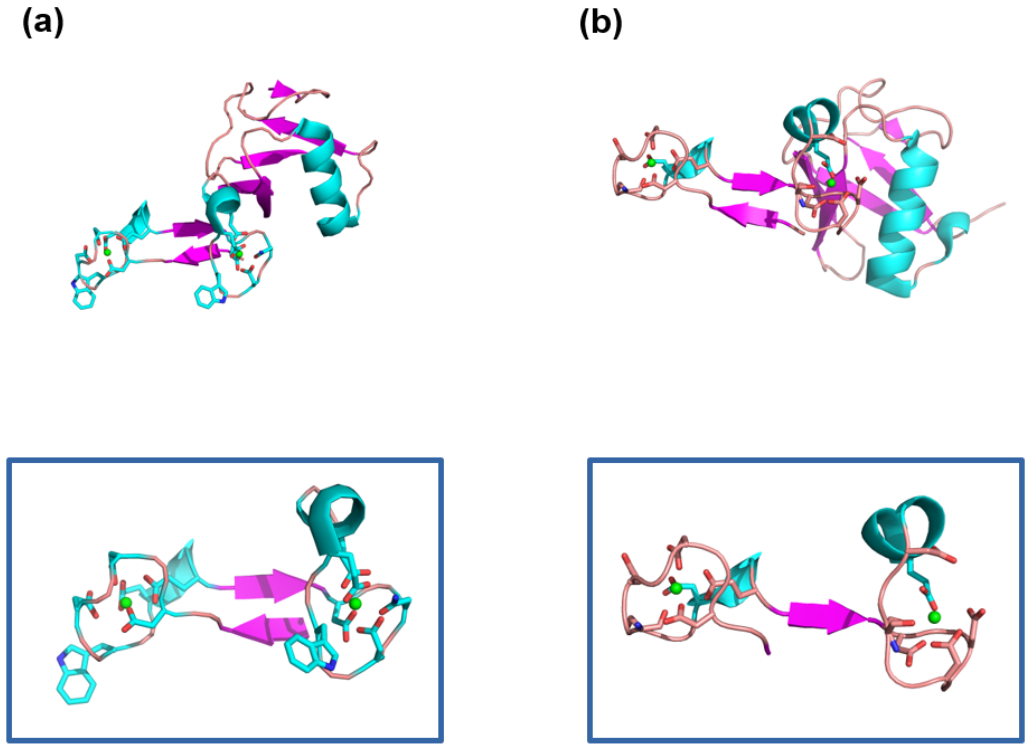

Figure S1. *Structures from dLBT-ubiquitin (PDB id: 2OJR)[3] (a) and xq-dSE3-ubiquitin (PDB id: 3VDZ)[4] (b) depicted using PyMOL are shown. The magenta arrow represents β-sheet secondary structure, whereas the light blue helix shows a 3_{10} helix. The structures shown enclosed in blue rectangles display only the dLBT fragment belonging to dLBT-ubiquitin and xq-dSE3-ubiquitin respectively. The aminoacidic sequences corresponding to the fragments shown are: PGYIDTNNDGWIEGDELYIDTNNDGWIEGDELLA (belonging to dLBT-ubiquitin), and YIDTDNDGSIDGDELYIDTDNDGSIDGDELLA (belonging to xq-dSE3-ubiquitin).*

S2 Bacterial expression of LBT peptides

Peptides LBT and LBTC were purchased from Genscript. Peptides were synthesized by the stepwise Solid Phase Peptide Synthesis (SPPS) chemical method and quality-checked by MS and HPLC by the manufacturer. Ln^{3+} stock solutions were prepared from the $LnCl_3$ hydrate salts from Sigma-Aldrich (as 50 mM solutions in 1 mM HCl) and were diluted as needed.

For the bacterial biosynthesis we employed a pGEX-2T plasmid with a GST tag in a culture of E. coli. After cell harvesting by centrifugation, sonication for cell lysis and another centrifugation step to get rid of cellular debris and other cellular structures, the supernatant underwent Affinity Chromatography using a Glutathion Sepharose 4B resin to retain the GST-tagged daLBT peptide. The purified peptide was analysed by Polyacrylamide Gel Electroforesis (SDS-PAGE).

For our purposes, this means that it is possible to make controlled changes in the coordination environment, altering the distribution of the charges but maintaining the locations of the ligands. Thus, as a double-spin peptidic prototype we employed YIDTDNDGWYEGDELYIDTNNDGWYEGDELLA (double assymetric LBT or in short: daLBT). This is a novel fusion of two known lanthanide binding subunits with the mentioned carboxamide-carboxylate modification in the coordination sphere. Therefore, the residue marked as N (asparagine) in Figure 4, is occupied by carboxylate in the first unit and carboxamide in the second unit, resulting in two different coordination environments. In this dinuclear system we expect a Ln-Ln distance of 20±1 Angstrom (see Supporting Information section S1).

For the preparation of this novel dimer we employed bacterial biosynthesis, or recombinant protein expression, a standard Molecular Biology technique to allow production of target proteins. It is an affordable alternative to chemical synthesis for long peptides/proteins (longer than a few decades of aminoacids). For our purposes, this technique will be essential when the production of longer peptides is required i.e. for peptides containing multiple Lanthanide Binding Tags. Typically, in recombinant protein expression, the DNA sequence coding the target protein is placed under the control of an inducible promotor, meaning that expression only occurs when a specific chemical compound is present in the culture media. This DNA sequence is then inserted into a commercial plasmid (an auto-replicative circle of DNA that propagates in bacteria). When the control compound is added to a culture of plasmid-containing bacteria, the protein is expressed. Standard peptidic tags, that are already in the plasmid, fused in N- or C-terminal position to the protein of interest, allow the purification of the target protein via Affinity Chromatography.

Bacterial biosynthesis, followed by purification by chromatography and analysis by electrophoresis resulted in the detection of a peptide with a molecular weight MW ≃ 30 kDa, compatible with the predicted mass of the recombinant protein (MW (GST-daLBT) = MW(GST) + MW (daLBT) = 26 KDa + 4 KDa). In a posterior Western-Blot analysis with an anti-GST antibody, the putative GST-daLBT band produced luminiscence, further supporting the hypothesis that GST-daLBT has been obtained.

The peptides were produced as gene fusions to the GST protein.[5] They were constructed in the vector pGEX-2T, and fusion proteins were expressed in *E. coli* according to the methods described in the GST gene fusion system handbook from GE Healthcare[6]. The original sequence for the **LBT** peptide (YIDTNNDGWYEGDELLA) was obtained from Nitz *et al*., 2003[7] and the double assymmetric LBT, from now on **daLBT**, (YIDTDNDGWYEGDELYIDTNNDGWYEGDELLA)

was designed so as the two **LBT** units were distinct and offer different coordination environments, thus different magnetic energy levels, to make them distinguishable by pulsed EPR.

To generate the GST-TEV-**LBT** and GST-TEV-**daLBT** fusion proteins, the DNA fragment encoding a TEV (Tobacco Etch Virus protease) cut site and the 17-aminoacid **LBT** (or 32-aminoacid **daLBT**) along with BamHI and EcoRI recognition sites at 5'and 3' respectively was ordered as a long oligonucleotide from IDT (Table S2). The TEV cut site (ENLYFQ aminoacidic sequence) was added to facilitate the removal of the GST moiety, which could be required for posterior purification steps. The Forward/Reverse pair of oligonucleotides were annealed as described previously.[8] Briefly, duplex annnealing for both constructions (**LBT** and **daLBT**) was achieved by preparing a mixture of 25 μM final concentration of both primers (F/R), heating at 95ºC for 5 min and letting it cool down slowly to room temperature for 45 min, left on ice for 2 minutes, and finally stored at -20ºC. The annealed duplexes (F/R-BamHI-TEV-LBT-EcoRI and F/R-TEV-daLBT-EcoRI) and vector pGEX-2T were doubly digested with BamHI and EcoRI following ThermoFisher Scientific recommendations. The DNA fragments containing **LBT** or **daLBT** sequences were then ligated into the BamHI-EcoRI pGEX-2T gel-purified digested plasmid, and the product was transformed into XL10-GOLD ultracompetent cells for storage.

Colonies were checked by Sanger sequencing with pGEX-5'-SP and pGEX-3'-SP primers. Plasmids from the verified constructions (both bearing **LBT** or **daLBT** sequences) were recovered and transformed into the *E.coli* expression strain BL21 DE3 pLys. The expression of the peptides was monitored as described in the GST gene fusion system handbook (GE Healthcare)[6] and can be seen in Figure S2.

Table S2. *Oligonucleotides used for cloning and verifying sequences.*

Name	Sequence (5' to 3')
F-BamHI-TEV-LBT-EcoRI	ACTGGGATCCGAAAACCTGTATTTTCAGGGCTATATTGATACCAACAACGATGGCTGGTATGAAGGCGATGAACTGCTGGCGTAAGAATTCACTGC
R-BamHI-TEV-LBT-EcoRI	GCAGTGAATTCTTACGCCAGCAGTTCATCGCCTTCATACCAGCCATCGTTGTTGGTATCAATATAGCCCTGAAAATACAGGTTTTCGGATCCCAGT
F-BamHI-TEV-daLBT-EcoRI	ACTGGGATCCGAAAACCTGTATTTTCAGGGCTATATTGATACCGATAACGATGGCTGGTATGAAGGCGATGAACTGTATATTGATACCAACAACGATGGCTGGTATGAAGGCGATGAACTGCTGGCGTAAGAATTCACTGC
R-BamHI-TEV-daLBT-EcoRI	GCAGTGAATTCTTACGCCAGCAGTTCATCGCCTTCATACCAGCCATCGTTGTTGGTATCAATATACAGTTCATCGCCTTCATACCAGCCATCGTTATCGGTATCAATATAGCCCTGAAAATACAGGTTTTCGGATCCCAGT
pGEX-5'-SP	GGGCTGGCAAGCCACGTTTGGTG
pGEX-3'-SP	CCGGGAGCTGCATGTGTCAGAGG

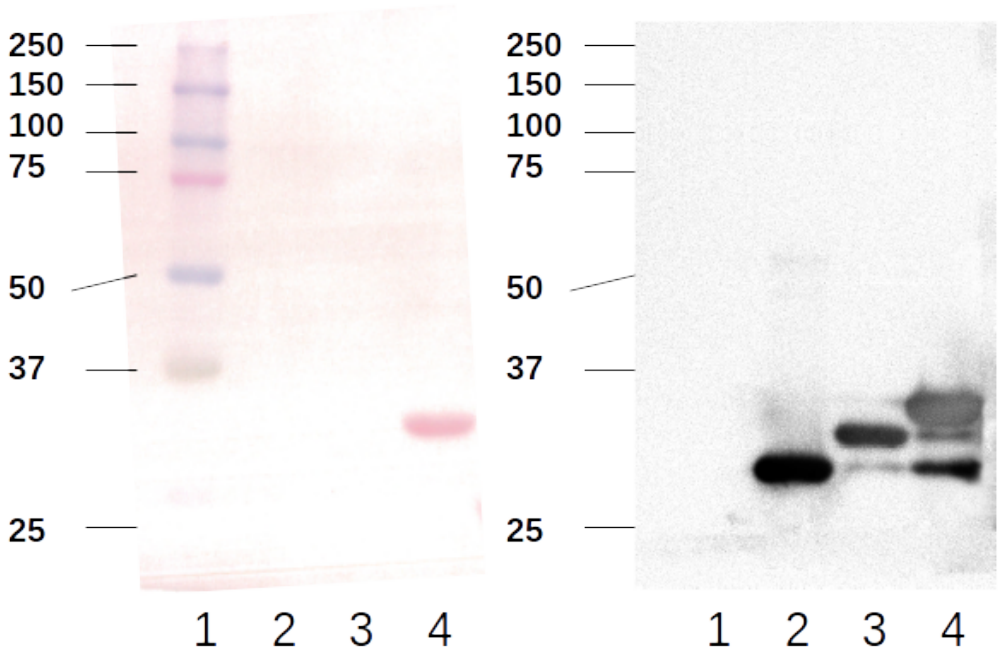

Figure S2. *Western blot of the GST-LBT and GST-daLBT fusion proteins expressed. Left panel shows a Ponceau S staining of the PVDF membrane, and right panel shows a western-blot developed using anti-GST Horseradish Peroxidase conjugated antibody (GE Healthcare) with SuperSignal West Pico Chemilluminescent substrate (Thermo Fisher) as recommended by manufacturers. Molecular weight markers are shown on the left. Lanes are as follow: 1) molecular weight marker (BioRad Precision Plus Protein Kaleidoscope Standards), 2) Elution of GST, 3) Elution of GST-TEV-LBT, 4) Elution of GST-TEV-daLBT.*

S3 Design of a nonanuclear metallopeptide

The DNA sequences and corresponding peptides defining what we denote as 0 and 1, both in the daLBT in the previous section and in the present nonanuclear sequence, are defined in Table S3.1. There are no fundamental differences between the design of sequences of different lengths. The different linkers suggested are listed in the table S3.2. We aimed for this in order to obtain a high affinity binding of lanthanide cations to our designed protein.

Table S3.1. *LBT units used for designing a lanthanide-bearing nonanuclear metallopeptide.*

LBT unit		Sequence (5' to 3') (N-t to C-t)
0	DNA	TATATTGATACCAACAACGATGGCTGGTATGAAGGCGATGAACTGCTGGCG
	protein	Y I D T N N D G W Y E G D E L L A
1	DNA	TATATTGATACCGATAACGATGGCTGGTATGAAGGCGATGAACTG
	protein	Y I D T D N D G W Y E G D E L

One could use standard techniques of recombinant protein expression for the preparation of this peptide with nine LBT units. Also, it should be easy to prepare polypeptides containing 3 to 9 LBT units due to the use of unique linkers for the spacing of the units. The unique spacers allow for the specific PCR amplification of DNA encompassing 1, 2, 3, 4, 5, 6, 7, 8 or 9 LBT units, which in turn could be sub-cloned and purified in *E. coli*. The linkers were all designed with a GASAG amino acid sequenced (unique DNA sequences) to facilitate independent folding of each LBT.
In short, the DNA sequence coding this protein of interest would be placed under the control of an inducible Lac promoter (expression only occurs when Isopropyl β-D-1-thiogalactopyranoside (IPTG) is present in the culture media)[9] and inserted into an *E. coli* commercial plasmid containing an Ampicillin resistance marker. The cloning procedure would be performed in a similar way as presented for the dimer in section S2. Then, the plasmid-containing bacteria could be grown in the presence of IPTG, the 's010011101' fusion protein expressed, and purified from the culture. The purificaction is usually enabled by peptidic tags. These tags can either be fused in N- or C-terminal position to the protein of interest in the context of the plasmid, allowing to perform Affinity Chromatography to purify the cloned protein.

Table S3.2. Linkers employed for the designed of the nonanuclear LBT peptide.

# Linker		Sequence (5' to 3') (C-t to N-t)
1	DNA	ggcagcggcgcgagcgcgggc
	protein	G S G A S A G
2	DNA	gcgggcagcggcgcg
	protein	A G S G A
3	DNA	ggcggcagcgcggcg
	protein	G G S A A
4	DNA	gcggcgggcgcggcg
	protein	A A G A A
5	DNA	ggcggcgcgggcggc
	protein	G G A G G
6	DNA	gcggcgagcggcggc
	protein	A A S G G
7	DNA	gcgagcggcagcgcg
	protein	A S G S A
8	DNA	ggcagcgcgagcggc
	protein	G S A S G
9	DNA	ggcgcgggcgcgagc
	protein	G A G A S
10	DNA	taatagtgaaattgg
	protein	stop codons

S4 Further approaches for the biochemical design of polynuclear metallopeptides

The nucleosome, which is key for the first approach, is the fundamental repeating element of chromatin (nuclear structure that contains the genetic material in the eucharyotic cell). It comprises between 157 and 240 base pairs (bp) of DNA, the four core histone proteins (H2A, H2B, H3 and H4) and the linker histone H1. The nucleosome core contains 147 bp of DNA supercoiled in 1.67 left-handed turns around a core histone octamer. The octamer consists of one tetramer of $(H3\text{-}H4)_2$, and two dimers of (H2A-H2B), and has a molecular weight of 110 kDa. The possibility of using the biomimetic self-organization of proteins that stably bind to DNA to form the nucleosome core superstructure, is something one can benefit from. Histones are small (11-16 kDa), basic proteins, with two differentiated domains: a flexible N-terminal tail with basic properties 15-30 aminoacids long, and a C-terminal globular domain (approx. 65 aminoacids) that contains a structural motive called histone-fold, important for molecular contacts within the histone octamer. The tagging of the core histones with LBTs is feasible using standard molecular biology tools.
Nucleosome assembly in vitro has been the object of many studies, so reliable protocols are readily available in the literature.[10] Core histone octamers that are repetitively spaced along a DNA molecule are known as nucleosomal arrays. Nucleosomal arrays have been reconstituted from recombinant histone octameres, using as template DNA well known nucleosome positioning sequences (NPS).[11]

These DNA sequences are able to strongly position a nucleosome in vitro, suggesting the possibility that favorable instrinsic signals might reproducibly structure chromatin fragments which allows for the organization of nanoobjects. These NPS would be vital for this approach as they provide specific docking sites for the histone octamer (our nanoobjects) on in vitro reconstitution assays (also called chromatin assembly assays). Nucleosome assembly can be achieved then, by combining core histone octamers that will bind to tandemly repeated nucleosome positioning sequences. Widom 601 sequence, 5s rDNA or MMTV-B sequence are NPS of about 170 bps that are currently used in nucleosome assembly in vitro.[12]

The reconstitution assays allow the assembly of a compositionally uniform and precisely positioned nucleosomal array that could act as the scaffold of our spin-LBT bearing histones. Histone binding in nucleosome reconstitution is not expected to be strongly affected by the addition of an LBT in its N-terminal end (or other positions within the protein), as this tag is only 15 aminoacids long. The second approach would be based on proteic tandem repeats as LBT scaffold. Up to 14% of all proteins contain repetitive regions.[13] Repeats vary considerably in size, order and complexity. In some cases, protein repeats can occur tandemly in sequence forming integrated assemblies units when viewed as three-dimensional structures (for example TALEN). Such repeats are essentially defined by their multiplicity and thus differ from both domains and motifs since these can occur in a variable number of repeats without altering the integrity, activity or stability of the overall protein.[14] The structure of many of these proteins has already been solved, giving an atomic (Å) resolution of the location of the repeats and the distance between them. Thus, these proteins represent a splendid scaffold for the insertion of Lanthanide Binding Tags (LBT) in an ordered manner.

References

[1] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. *Nucleic Acids Res.*, 2000, **28,** 235-242.

[2] Schrodinger. The PyMOL Molecular Graphics System, Version 1.8, 2015.

[3] Silvaggi, N. R.; Martin, L. J.; Schwalbe, H.; Imperiali, B.; Allen, K. N. *J. Am. Chem. Soc*. 2007, **129**, 7114–7120

[4] Daughtry, K. D.; Martin, L. J.; Sarraju, A.; Imperiali, B.; Allen, K. N. *Chembiochem Eur. J. Chem. Biol*. 2012, **13**, 2567–2574.

[5] Harper, S.; Speicher, D. W. *Methods Mol. Biol.* 2011, **681**, 259–280.

[6] GST gene fusion system handbook

http://www.gelifesciences.com/file_source/GELS/Service%20and%20Support/Documents%20and%20Downloads/Handbooks/pdfs/GST_gene_fusion_system_handbook.pdf

(accessed Jul. 3, 2017).

[7] Nitz, M.; Franz, K. J.; Maglathlin, R. L.; Imperiali, B. *Chembiochem Eur. J. Chem. Biol*. 2003, **4**, 272–276.

[8] Kennedy, S. R.; Schmitt, M. W.; Fox, E. J.; Kohrn, B. F.; Salk, J. J.; Ahn, E. H.; Prindle, M. J.;Kuong, K. J.; Shen, J.-C.; Risques, R.-A.; et al. *Nat. Protoc*. 2014, **9**, 2586–2606.

[9] Jacob, F.; Monod, J. *J. Mol. Biol.* 1961, 3, 318–356.

[10] Rogge, R. A., Kalashnikova, A. A., Muthurajan, U. M., Porter-Goff, M. E., Luger, K., & Hansen, J. C. *JoVE*. 2013, **79**, e50354.

[11] Nishikawa, J. I., & Ohyama, T. *Nucleic Acids Res*. 2013, **41**, 1544-1554.

[12] Tóth, K., Böhm, V., Sellmann, C., Danner, M., Hanne, J., Berg, M., Barz, I., Gansen, A., Langowski, J. *Cytometry Part A*, 2013, **83**, 839-846.

[13] Marcotte, E. M., Pellegrini, M., Yeates, T. O., Eisenberg, D. *J. Mol. Biol*. 1999, **293**, 151-160.

[14] Andrade, M. A., Perez-Iratxeta, C., Ponting, C. P. *J. Struct. Biol*. 2001, **134**, 117-131.