UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO

Scuola di Dottorato in Chimica del Farmaco

DIPARTIMENTO

DISFARM (Dipartimento di Scienze Farmacologiche)

CURRICULUM

Sintesi Speciali di Prodotti ad Attività Biologica

Ciclo XXVI

NEW DIMENSIONS INTO PROTEIN-NUCLEIC ACIDS INTERACTIONS

CHIM/03 e CHIM/06

CHIARA ANNONI

PROF.SSA MARIA LUISA GELMI

PROF. ERMANNO VALOTI

A.A.

2012/2013

Index

Acknowledgments

Before start talking about my research results, I would like to dedicate few words to introduce how my research path took root. My thesis work started as a collaboration between the Dipartment of Pharmaceutical Science (DISFARM) of Milano University and the Energy Science Institute of Kyoto University. I spent my first year of doctorate under the guide of Professor Gelmi, at Milano University, developing a structural model of the DNA binding domain of Maf proteins, a family of transcription factors involved in numerous biological events as well as showing a leading role in several pathologies. Since our interest focused on protein-nucleic acids interactions, Professor Gelmi allowed me to expand my knowledge in this field working in the research group of Professor Morii, at Kyoto University. The design of biomacromolecules and their assemblies for molecular recognition, catalysis and signal transduction are the main research topics of Morii's group. Thanks to this collaboration still ongoing, I spent the two following years of doctorate at Kyoto University dealing with protein/nucleic acids assemblies with the aim of realizing biomimetic function of biological systems. I must admit that these years were a school of life for me. I will always be grateful to both my mentors who gave me this big opportunity and supported me along the way: under their supervision I learnt a lot and I grew up not only as a scientist but also as a person.

"Learn from yesterday, live for today, hope for tomorrow.

The important thing is to not stop questioning."

– Albert Einstein, *Relativity: The Special and the General Theory*

General introduction

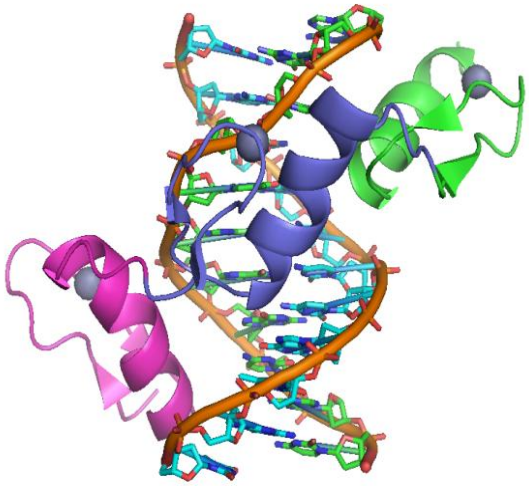


Figure 1: DNA-Zinc finger complex (Pavletich, N. P., *et al*., *Science* **1991**, 252, 809; Sera, T., *et al. Biochemistry* **2002**, 41, 7074).

Constructing Functional Molecules by Using Biopolymers

The production system employed by living organisms is the most basic system for supplying the materials that we utilize, yet the efficiency, the variety and the versatility of this system remain untapped. Attempts to harness these systems directly or to imitate, reproduce and improve them by means of artificial materials is one of the world's supreme propositions; various strategies for constructing the requisite functional molecules have been developed.[1-37] In particular, a number of unique methods with the capability to utilize the features of bio-systems have been already established in the construction of functional biopolymers, in which the "life phenomenon" is controlled by the function of the biopolymer. This system has been applied to the field of energy and material production as well as the medical and food production fields by exploiting the catalytic activity of enzymes and the chemical properties of nucleic acids that bear the preservation and the mediation of genetic codes. Natural biopolymers utilize the individual function of thousands of enzymes and behave as an organization by the cooperation of each supermolecule, even though the constituent molecules are constructed by very simple elements and monomers based on carbon, hydrogen, nitrogen, oxygen, phosphorus and sulfur. In other words, there is the potential for constructing various molecules with excellent functionalities by assembling very simple elements and components.

Methods for constructing functional biopolymers

The substrate selectivity and the specificity of enzymatic reactions are important parameters that are difficult to design and reproduce using artificial catalysts and attempts to obtain feedback based on the information garnered by clarification of the structure-activity relationship of the native enzymes have been applied to the design of the desired functional materials. Site-specific mutation of the native enzyme based on the insight provided by molecular chemistry and three-dimensional structural information is often utilized as a representative method for the construction of functional biopolymers (structure-based design). Moreover, construction of functional biopolymers or ligands by computational design, including *de novo* design, which utilizes a simulation based on molecular dynamics calculations and information from a database of the relationship between the amino acid sequence and the folding of the native protein has been performed. However, the applicability of these methods is still limited because of the database restrictions on the structure of biopolymers and the calculation method; consequently, the constructed molecule often does not have the originally designed function. On the other hand, there is a method in which the improvement and modification of the enzymatic activity of the native enzyme can be achieved by introducing a random mutation into the part of the enzyme that is difficult to design based on the structural information (directed evolution method).[4-9] This method does not require structural information for the part of the molecule where the random mutation is introduced, moreover, mutations that are burdensome to be estimated from the structural set of information may be obtained. However, in many cases, it is challenging modifying the activity of the original enzyme dynamically and, nowadays, the application of this method is limited. In addition, another approach has been developed, in which the target molecule is isolated based on some indexes, such as the affinity for the target molecule or the activity of the target reaction, from a library composed of various molecules constructed by introducing the randomized component (combinatorial chemistry). In vitro selection or SELEX (systematic evolution of ligands by exponential enrichment) that utilizes a nucleic acid library,[10-11] phage display,[12] ribosome display,[13] mRNA display,[15-18] as well as emulsion techniques (in vitro compartmentalization)[19-22] that operates using protein and peptide libraries, the split and mix method,[23-26] which employs solid phase synthesis by the use of resins, have been established as representative methods for realizing this concept. Nevertheless, it is difficult to decide in advance what setup of the selection biases enables construction of the molecule that has the desired function and restriction of the library size compared to the ideal diversity of molecules that can be estimated. Besides, en efficient method for the sequential construction of polymers by means of

organic synthesis is necessary when the artificial materials are utilized as components. It is very difficult selecting the relevant synthetic approaches and solving the issues related to the purity of products and the design of the template molecule or monomer that becomes the scaffold for the library construction. Biopolymers represent a potential means of solving these problems considering that they have the advantage of being more adaptable to library construction and selection compared to artificial molecules, thanks to the limited number of monomers in the biopolymer and the high purity level of the product as a result of the reaction specificity of enzymes. In addition, the characteristic of amplification that cannot be achieved in a conventional artificial molecule can be conveniently realized. Conventional methods mainly utilize the affinity for the target compound as an index, but when the activity is assumed as an index, it becomes necessary to edit other design parameter individually. These methods ultimately lead to the concept in which the entire envisioned molecule is constructed and the functions evaluated one by one. However, an exceptional technical breakthrough would be necessary for the achievement of this concept because the exhaustive examination of all the possibilities is laborious to be executed. Recently, the development of the single molecule and single substance measurement technique has emerged. For example, the use of the micro flow channel of the cell sorter, affords a novel selection method.[27-30] Though only just burgeoning, and although a number of evaluations will be necessary in order to achieve generality and practicality, the creation or the maturation of a novel method for constructing functional molecules by using measurements or detection and chemical synthesis technologies, that differ from the currently employed concepts, is anticipated.[31,32]

Objectives of this work of thesis

In the work of my thesis I attempted making a new step forward a better knowledge of protein-nucleic acid biopolymers, focusing on two complexes of strong interest. One of the important factors for determining the expression of the function is the structure of the biopolymer. Clarification of the relationship between the structure and the function of biopolymer is of crucial importance to the development of methods for constructing tailor-made functional molecules. The first goal of my thesis focalized on the obtainment of a structural model of a well-known protein-DNA complex: Maf DNA binding domain and it DNA target (T-MARE was used in this study) with the aim of outlining a strategy for the obtainment of activity modulators. Both *in silico* simulations and experimental studies were carried out leading to the definition of a *modus-operandi* based on a disorder-order transition of the complex which is commendatory for the protein activation. On the other hand, a number of

artificial enzymes have been recently constructed by using the molecular design based on structural information, screening methods that utilize libraries or by a combination of these two methods. However, the above mentioned approaches use single proteins or single nucleic acid as the structural unit, and the activity of the constructed artificial enzymes is remarkably lower in many cases, than that of the native enzymes. The construction of functional complexes (rather than a single biopolymer) as scaffold can be considered as one potential solution to these drawbacks. Further in my studies, using the HIV-Rev peptide and RRE (Rev Responsive Element) RNA complex as a scaffold, for which the tridimensional structure was fully characterized, the assemble of ribonucleopeptidic fluorescent sensors was accomplished in a stepwise manner. In this method, a randomized nucleotide sequence was introduced into the RNA subunit of RNP to construct a RNP library on which the in vitro selection method was applied.[33-38] In the second step, the Rev peptide was modified with the fluorophore without altering the affinity and specificity of the RNP receptor.[39] In the absence of a ligand for RNP, fluorescence emission was effectively quenched in the RNP complex, but recovered upon ligand binding. RNP sensors were thus created, as the ligand-binding event can be monitored by measurement of the fluorescence signals.

References

1) Damborsky, J., and Brezovsky, J., *Curr. Opin. Chem. Biol.* **2009**, 13, 26-34.
2) Chica, R. A., Doucet, N., and Pelletier, J. N., *Curr. Opin. Biotech.* **2005**, 16, 378.384.
3) Bolon, D. N., Voigt, C. A., and Mayo, S. L., *Curr. Opin. Chem. Biol.* **2002**, 6, 125-129.
4) May, O., Nguyen, P.T., and Arnold, F. H., *Nat. Biotech.* **2000**, 18, 317-320.
5) Moore, J. C., and Arnold, F. H., *Nat. Biotech.* **1996**, 14, 458-467.
6) Altamirano, M. M., Balckburn, J. M., Aguayo, C., and Fersht, A. R., *Nature* **2000**, 403, 417-468.
7) Xia, D., Chen, L., Sera, T., Fa, M., Schults, P.G., and Romesberg, F. E., *Proc. Natl. Acad. Sci. USA* **2002**, 99, 6597-6602.
8) Bittler, J. A., Le, B. V., Liu, J. M., and Liu, D. R., *Proc. Natl. Acad. Sci. USA* **2004**, 101, 7011-7016.
9) Schmidt-Dannert, C., Umeno, D., and Arnold, F. H., *Nat. Biotech.* **2000**, 18, 750-753.
10) Elligton, A. D., and Szostak, J. W., *Nature* **1990**, 346, 818-822.
11) Tuerk, C., and Gold, L., *Science* **1990**, 246, 505-510.
12) Smith, G. P., *Science* **1985**, 228, 1315-1317.
13) Mattheakis, L. C., Bhatt, R. R., and Dower, W. J., *Proc. Natl. Acad. Sci. USA* **1994**, 91, 9022-9026.
14) Hanes, J., and Plückthun, A., *Proc. Natl. Acad. Sci. USA* **1997**, 94, 4937-4942.
15) Nemoto, N., Miyamoto-Sato, E., Husimi, Y., and Yanagawa, H., *FEBS Lett.* **1997,** 414, 405-408.
16) Roberts, R. W., and Szostak, J. W., *Proc. Natl. Acad. Sci. USA* **1997**, 94, 12297-12302.
17) Keefe, A. D., and Szostak, J. W., *Nature* **2001**, 410, 715-718.
18) Tabuchi, I., Soramoto, S., Nemoto, N., and Husimi, Y., *FEBS Lett.* **2001**, 508, 309-312.
19) Tawfic, D. S., and Griffiths, A. D., *Nat. Biotechnol.* **1998**, 7, 652-656.
20) Agresti, J. J., Kelly, B. T., Jaschke, A., and Griffiths, A. D., *Proc. Natl. Acad. Sci. USA* **2005**, 102, 16170-16175.
21) Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A. D., Magdassi, S., and Tawfik, D. S., *Anal. Biochem.* **2004**, 325, 151-157.
22) Doi, N, and Yanagawa, H., *FEBS Lett.* **1999**, 457, 227-230.
23) Furuka, A., Sebestyen, F., Asgedon, M., and Dibo, G. J., *Pept. Protein Res.*, **1991**, 37, 487-493.
24) Lam, K. S., Salmon, S.E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., Knapp, R. J., *Nature* **1991**, 354, 82-84.

25) Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., Cuervo, J. H., *Nature* **1991**, 354, 84-86.

26) Tabuchi, I., Soramoto, S., Ueno, S., and Husimi, Y., *BMC Biotech.***2004**, 4, 19-26.

27) Sepp, A., Tawfik, D. S., and Griffiths, A.D., *FEBS Lett.* **2002**, 532, 455-458.

28) Ohuchi, S., Nakano, H., and Yamane, T., *Nucleic Acids Res.* **1998**, 26, 4339-4346.

29) Nakano, H., Koyabashi, K., Ohuchi, S., Sekiguchi, S., and Yamane, T. J., *Biosci. Bioeng.* **2000**, 90, 456-458.

30) Granieri, L., Barnet, J. C., Griffiths, A. D., and Merten, C. A., *Chem. Biol.* **2010**,17, 229-235.

31) Mendousa, S. D., and Bowser, M. T., *J. Am. Chem. Soc.* **2004**, 126, 20-21.

32) Wassenaar, J., Jansen, E., Zeist, W. J., Bickelhaupt, F. M., Siegler, M. A., Spek, A. L., Reek, J. N. H., *Nature Chemistry*, **2010**, 2, 417-421.

33) Morii, T., Hagihara, M., Sato, S., and Makino, K., *J. Am. Chem. Soc.* **2002**, 124, 4617-4622.

34) Sato, S., Fukuda, M., Hagihara, M., Tanabe, Y., Ohkubo, K., and Morii, T., *J. Am. Chem. Soc.* **2005**, 127, 30-31.

35) Hagihara, M., Fukuda, M., Hasegawa, T., and Morii, T., *J. Am. Chem. Soc.* **2006**, 128, 12932-12940.

36) Hasegawa, T., Hagihara, M., Fukuda, M., Nakano, S., Fujieda, N., and Morii, T., *Nucleosides Nucleotides Nucleic Acids* **2007**, 26, 1277-1281.

37) Hasegawa, T., Hagihara, M., Fukuda, M., Nakano, S., Fujieda, N., and Morii, T., *J. Am. Chem. Soc.* **2008**, 130, 8804-8812.

38) Hasegawa, T., Ohkubo, K., Yoshikawa, S., and Morii, T., *e-J. Surf. Sci. Nanotech.* **2005**, 3, 33-37.

39) Battiste, J.L., Mao, H., Rao, N. S., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R., *Science* **1996**, 273, 1547-1551.

Chapter 1:

Maf proteins: an insight into their DNA binding

1.1. Introduction

1.1.1. Disorder-order transitions in transcription factors

A deeply understanding of how transcription factors interact with genomic target sites is crucial for reconstructing transcriptional networks *in vitro*. In particular, in recent years, great attention has been paid to structural and functional intrinsic disorder inside the cell.[1] The occurrence of unstructured regions of significant size (>50 residues) is surprisingly common in functional proteins.[2,3] In addition, the existence of functional unstructured proteins, for example, polypeptide hormones,[4] has been recognized for many years, and unstructured proteins were observed in intact cells in early proton NMR experiments.[5] However, the functional role of intrinsically disordered proteins in crucial areas such as transcriptional regulation, translation and cellular signal transduction[6-9] has only recently been recognized, as a consequence of the use of new paradigms in biochemical methodology.[10] Intrinsic disorder (ID) regions promote molecular recognition primarily through four features.[11] ID proteins of ID regions are characterized by the unique combination of high specificity and low affinity in their interactions with functional partners, which is very important for transient protein-protein and protein-nucleic acids interactions, such as those that frequently occur during the signal transduction, recognition, and regulation events. Intrinsic plasticity enables a single ID protein or region to recognize and bind many biological targets with high specificity.[12-15] The ID protein or region have the propensity to form large interaction surfaces allowing them to wrap-up or surround their binding partners.[12, 16,17] The rapid turnover and reduced lifetime of ID proteins in cells might represent an important regulatory mechanism.[18] These observations have particular importance to the basic leucine zipper (b-LZ) class of transcription factors, whose activity relies on disorder–order transitions of their DNA binding domains (DBD) for specific dimerization and DNA binding.[19] The basic leucine zipper (bZIP) proteins form one of the largest families of eukaryotic transcription factors and play roles in a wide variety of biological phenomena, from responding to endoplasmic reticulum (ER) dysfunction to regulating immune responses and oncogenesis.[19] Members of this superfamily contain a positively charged DNA binding region composed of basic residues linked to a leucine zipper sequence and homo- or hetero-dimerize via this leucine zipper. Although considered to be the simplest known protein-DNA recognition motif, crystal structures of bZIP domains bound to DNA have revealed functional variability in how these conserved residues contact DNA, and no universal code linking basic region sequence with target DNA preferences has been developed. Within b-LZs, Maf proteins represent a new paradigm for DNA binding specificity as they recognize a long palindromic DNA sequence (known as Maf recognition

element, MARE) which is composed by a core and a flanking region.[1,20] In order to acquire a deeper knowledge regarding b-LZ proteins disorder-order transitions, in this chapter we will focus on the *in silico* conformational studies performed on c-Maf, obtained generating first an homology model and subsequently, carrying out molecular dynamics simulations.

1.1.2. References

1) Pellegrino, S., Annoni, C., Contini, A., Clerici, F., Gelmi, M. L., *Amino Acids* **2012**, 43, 1995-2003.
2) Dunker, A: K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., Obradovic, Z., *Biochemistry* **2002**, 41, 6573-6582.
3) Uversky, V. N., *Protein Sci.* **2002**, 11, 739-756.
4) Boesch, C., Bundi, A., Opplinger, M., Wüthrich, K., *Eur. J. Biochem.* **1978**, 91, 209-214.
5) Daniels, A. J., Williams, R. J., Wright, P. E., *Neuroscience* **1978**, 3, 575-585.
6) Wright, P. E., Dyson, H. J., *J. Mol. Biol.* **1999**, 293, 321-331.
7) Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., *J. Mol. Biol.* **2002**, 323, 573-584.
8) Iakoucheva, L. M., *Nucleic Acids Res.* **2004**, 32, 1037-1049.
9) Tompa, P., *Trends Biochem. Sci.* **2002**, 27, 527-533.
10) Dyson, H. J., Wright, P.E., *Nature Review* **2005**, 6, 197-208.
11) Uversky, V. N., Oldfield, C. J., Dunker, A. K., *J. Mol. Recognit.* **2005**, 18, 343-384.
12) Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Garner, E. C.; Obradovic, Z., *J. Mol. Graphics Modell.* **2001**, 19, 26-59.
13) Dyson, H. J., Wright, P. E., *Curr. Opin. Struct. Biol.* **2002**, 12, 54-60.
14) Wright, P. E., Dyson, H. J., *J. Mol. Biol.* **1999**, 293, 321-331.
15) Dyson, H. J., Wright, P. E., *Nat. Rev. Mol. Cell Biol.* **2005**, 6, 197-208.
16) Meador, W. E., Means, A. R., Quiocho, F. A., *Science* **1992**, 257, 1251-1255.
17) Gunasekaran, K., Tsai, C. J., Kumar, S, Zanuy, D., Nussinov, R., *Trends Biochem. Sci.* **2003**, 28, 81-85.
18) Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., Dunker, A. K., *Biochemistry* **2006**, 45, 6873-6888.
19) Miller, M., *Curr. Prot. Pep. Sci.* **2009**, 10, 244-269.
20) Blank, V., Andrews, N. C., *TIBS* **1997**, 22, 437-441.

1.2.Maf proteins: discovery, functions and structural motifs

1.2.1. Physiological roles of Maf proteins

The Maf transcription factors belong to the AP1 super-family of basic leucine zipper (bZIP) proteins, which also includes Fos, Jun, CREB and ATF families (**Figure 2**).

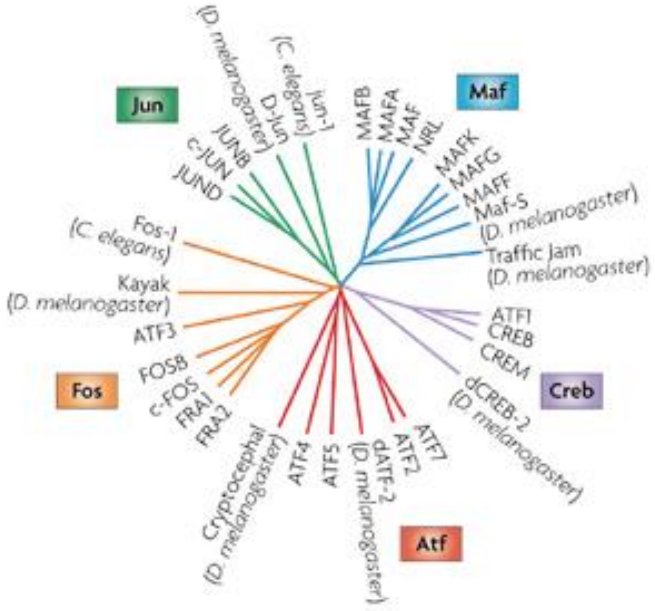


Figure 2: An unrooted phylogenic tree of the AP1 family of proteins. The tree was drawn using a CLUSTALW alignment algorithm at the National Laboratory for Computational Science and Engineering at the University of California, San Diego. Where available, *Caenorhabditis elegans* and *Drosophila melanogaster* orthologues were included.

The Maf family derives its name from the founding member, v-maf, that was originally discovered as the transduced transforming component of avian musculoaponeurotic fibrosarcoma virus, AS42.[1] Subsequent studies identified the cellular homolog of this gene, c-maf, from which the v-maf oncogene was originally transduced[2] but in addition, C-maf was found to be but one member of an extended multigene family.[3] Nowadays, the codifying genes for Maf family members has been isolated both in vertebrates (mammals, birds, frogs, fishes) and in invertebrates (*drosophila*)[3,4] (**Figure 3**).

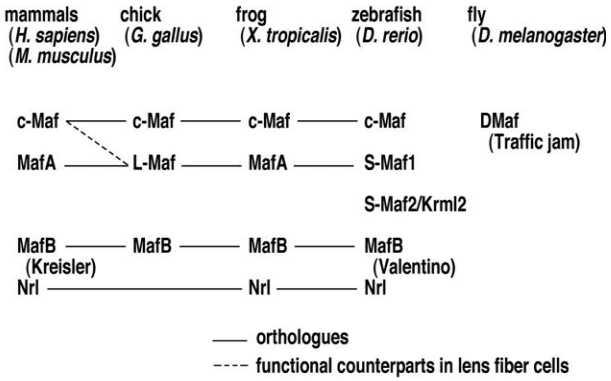


Figure 3: Maf family members identified in different species

Products of the maf proto-oncogene and related family members (the Maf family proteins) share a common, relatively well-conserved basic region and leucine zipper (b-Zip) motif which mediate DNA binding and dimer formation (**Figure 4**).[5]

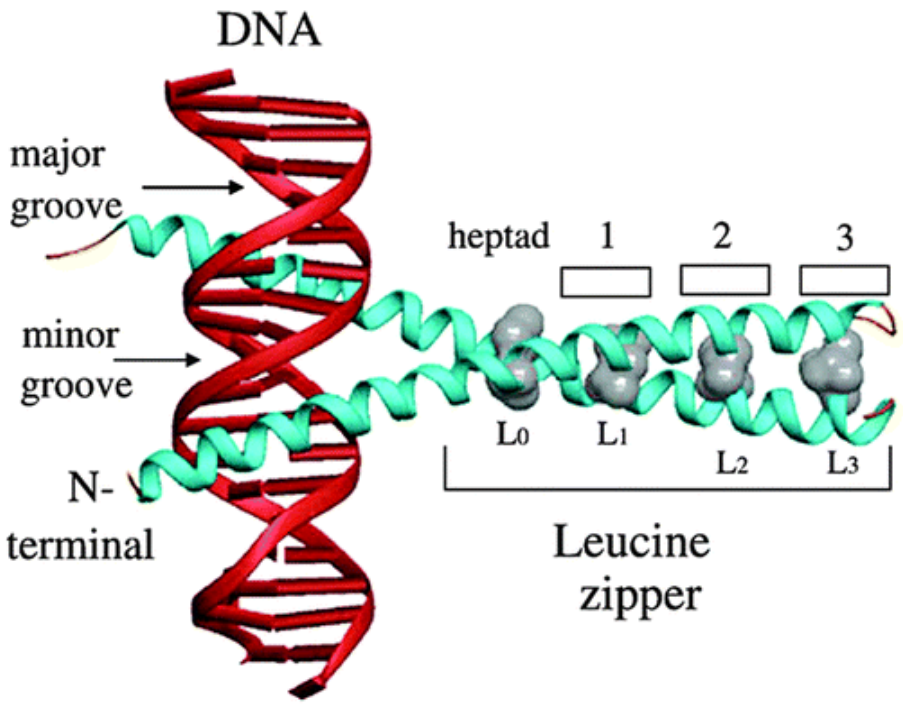


Figure 4: b-ZIP motif of Maf family members.

Seven members of the Maf family have been reported and were subdivided into two groups depending on their molecular weight: large Mafs (MafA, MafB, C-Maf and NRL, 26-39kDa) which have a putative transactivation domain at their N-termini, and small Mafs (MafF, MafG and MafK, 17-18kDa) which lack of such a domain (**Figure 5**).[5-8]

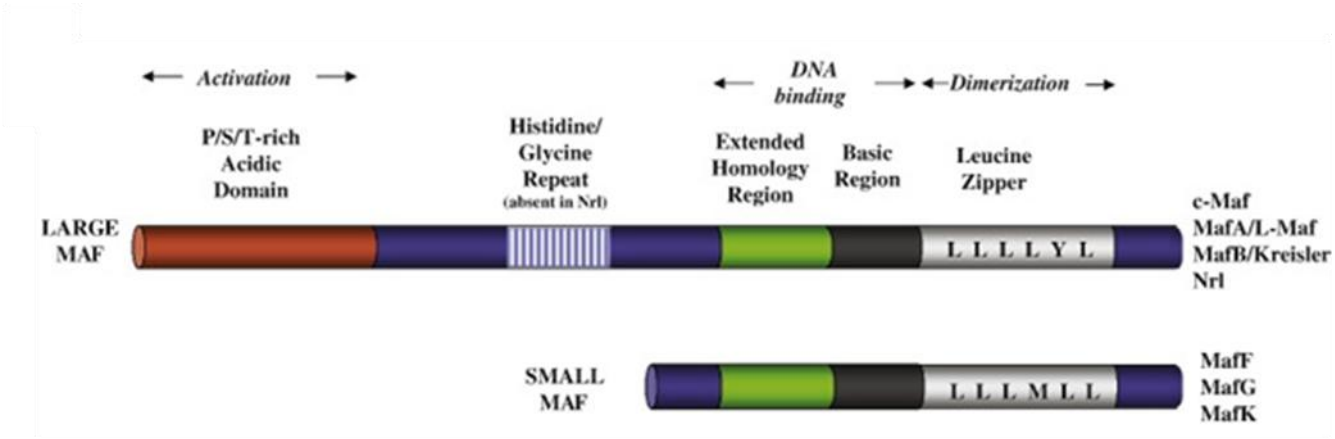


Figure 5: A comparison of large and small Maf transcription factors. The modular structure of the large and small Maf bZIP transcription factors with their respective functional domains is shown.

Mafs family proteins play important roles in cellular differentiation and morphogenesis, and their unique functions seem to be determined by their selective dimerization and DNA-binding specificity.[9] Large Maf proteins can activate transcription by recruiting the co-activators p300 (EP300), CRE binding protein (CBP, also known as CREBBP)[10] and P/CAF (also known as KAT2B)[11] or the TATA binding protein TBP.[12] Lacking the transactivation domain, small Mafs homodimers are not able to activate transcription, nevertheless, they are known to repress transcription[13,14] by competing, notably with large Maf proteins, for binding to target gene promoters.[15] Therefore, the ratio between small Maf homodimers and large Maf-containing complexes might have important biological consequences.[9,16] On the other hand, small Mafs can form heterodimers with other LZP transcription factors. For example, they are known to dimerize with p45 (the large subunit of the nuclear factor-erythroid 2, NF-E2) to activate erythroid cell-specific transcription. In addition, as dimeric partners of the bZIP NF-E2/Nrf family, small Maf proteins might contribute to oncogenic processes by participating in antioxidant response (**Figure 6**).[5-8]

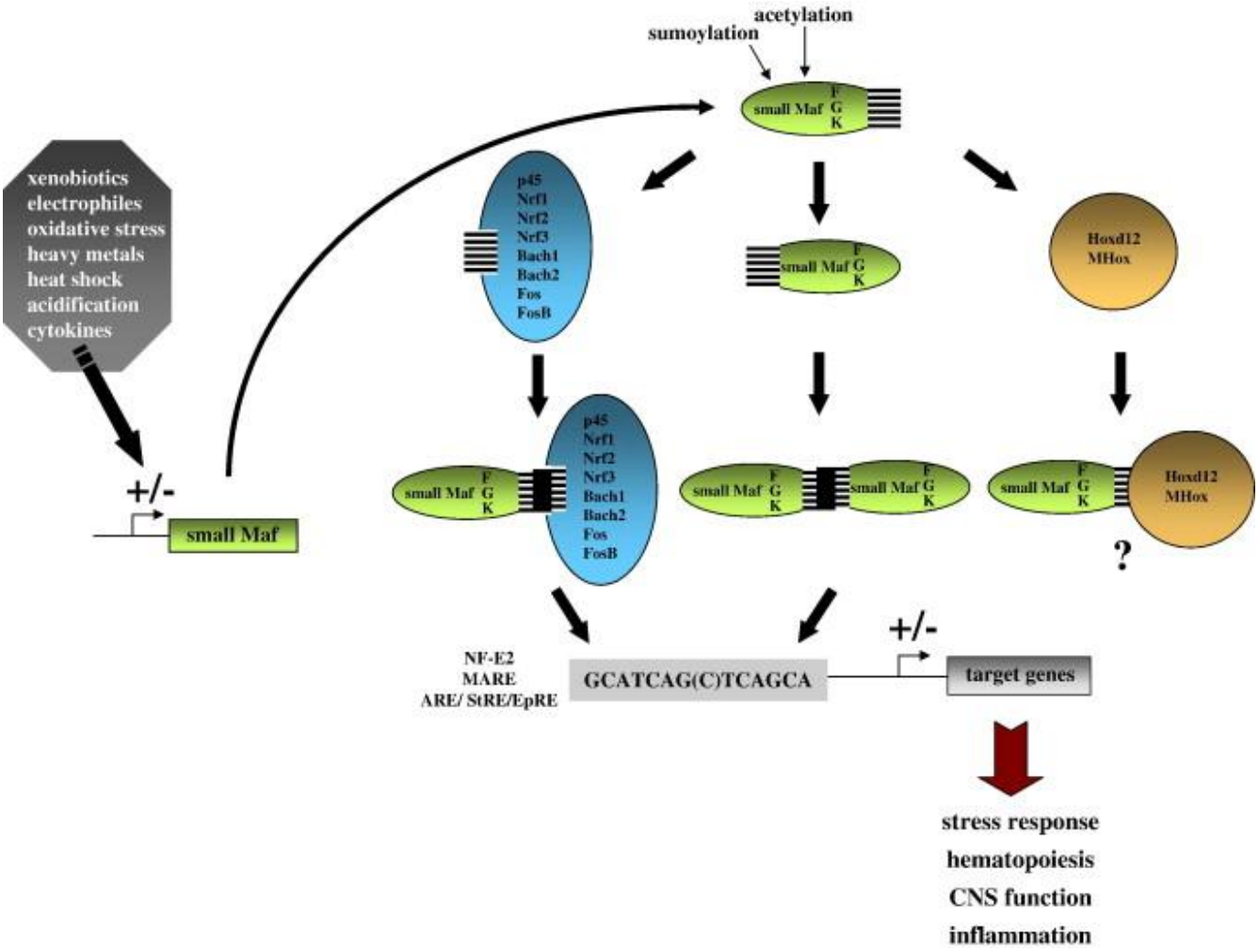


Figure 6: Small Maf protein network. The small Maf proteins are embedded in a regulatory network that is essential for various cellular functions, including the response to stress. The expression of small Maf genes can be modified by diverse reagents or conditions. The proteins may be modified post-translationally through acetylation or sumoylation (shown for MafG only). The small Mafs are able to dimerize with CNC or other bZIP factors *via* their leucine zipper domain. The resulting hetero- and homodimers bind to MARE sequences or similar motifs regulating, for example, stress response genes. The small Mafs have been shown to interact with Hox proteins, but the functional significance of these interactions is not known.

The large Maf genes display a restricted expression pattern[15,16] and, in contrast to JUN, Fos and other nuclear oncogenes, Maf genes do not behave as immediate early genes.[17] The expression of large Maf genes is tightly regulated in a spatio-temporal manner during development,[18,19] and, although these genes display some overlap in their expression profile, different members can be expressed in a specific window of time in a developing tissue.[9] This is exemplified in the lens and pancreas, where a cascade of expression of large Maf genes is observed. Therefore, the expression of the preceding gene could be

directly involved in the expression of the next gene in the cascade. Loss of function mutations in mice showed that large Maf genes are involved early in tissue specification,[20] and later in terminal differentiation (**Table 1**). within a given tissue, their roles appear to be highly specific. Although in some instances they can negatively regulate cell proliferation, their major role is not to control essential cellular functions such as cell cycle, apoptosis or migration, but rather to regulate the expression of tissue-specific genes involved in terminal differentiation (**Table 1**). In agreement with this, MAF and NRL are mutated in diseases such as *pulverulent cataract*, clumped pigmentary retinal degeneration and *retinitis pigmentosa*.[18,19] Interestingly, in virtually every case their ability to regulate gene expression depends on their synergy with other transcription factors, some of which are direct binding partners (**Table 1**), and they often control the expression of a set of genes that are expressed at high levels, such as insulin in β-cells or crystallins in lens fibre cells.[9]

Maf	Organ, tissue, cell type	Biological response(s)	Target gene(s)	Cooperating factor(s)	KO Phenotype(s)
MafA	Pancreas β-cells	Insulin transcription and production in β-cells	Insulin (*Ins*)	PDX1, NeuroD	Diabetes mellitus
MafB	Rhomben-cephalon	Rhomboemeres 5 and 6 identity	Hoxa3 and Hoxb3	KROX20, Ets	Hindbrain segmentation problems in the Kreisler (Kr^{ENU}) mutant; fatal apnea at birth to lack of neurons from the PreBötC complex in the knockout mutant.
	Myeloid progenitors	Macrophage differentiation	Unknown	Unknown	Altered actin-dependent macrophage morphology; F4/80 expression suppressed in non-adherent macrophages
	Kidney podocytes	Podocyte differentiation	Unknown	Unknown	Aberrant podocyte foot process formation and proteinuria in the kr^{ENU} mutant; tubular apoptosis and renal dysgenesis in the knockout mutant
	Pancreas	Pancreatic endocrine cell commitment and maturation	Glucagon (*Gcg*)	PAX6	Reduced number of cells expressing insulin and glucagon in both kr^{ENU} and knockout mutants
Maf	T cells	T helper 2 cell differentiation	*Il*4	NFAT	Lack of IL4 production (embryonic and perinatal lethality)
	Lens	Lens differentiation	Crystallins, Cryaa, Cryab, Crybb1 and Crygf	PAX6, CREB, SOX1, SOX2, PROX1	Microphtalmia due to defect in lens fibre elongation
	Endochondral bone	Chondrocyte differentiation	Collagen type II A1 (Col2a1)	SOX9	Decreased fetal bone length, abnormal terminal differentiation of hypertrophic

					chondrocytes
Nrl	Neuroretina	Rod differentiation	Rhodopsin (*Rho*)	CRX	Lack of rods

Table 1: CREB, cAMP responsive element binding protein; CRX, cone-rod homeobox containing gene; Il, interleukin; KO, knockout; NFAT, nuclear factor of activated T cells; PAX, paired Hox; PDX1, pancreatic and duodenal homeobox factor 1; PROX1, prospero homeobox 1; SOX, sex-determining region y (SRY) box.

1.2.2. DNA binding domain

The Maf transcription factors are a unique family among b-ZIP factors, characterized by the presence of three specific regions located on the C-terminal side: the Maf extended homology region (EHR), the basic region and the Leucin Zipper region (**Figure 7**).

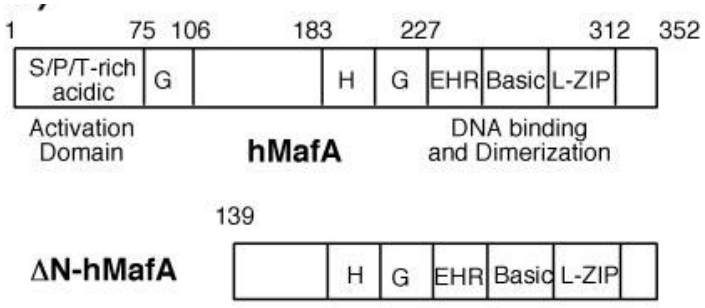


Figure 7: Schematic of deduced hMafA amino acid sequence shows various structural [serine/proline/threonine (S/P/T)-rich acidic domain, glycine (G)- and histidine (H)-rich regions, extended homology region (EHR); basic leucine zipper (L-LZP)] and functional (transcriptional actrivation, DNA-binding, and protein dimerization) domains. Schematic of ΔN-hMafA shows lack of the putative transcriptional activation domain.

The Leucin Zipper region consists of a repetition of hydrophobic residues in particular leucine. The leucine zipper domain allows the formation of homodimers or heterodimers, a prerequisite for DNA binding, and it is completely conserved within Maf family members (**Figure 8**).

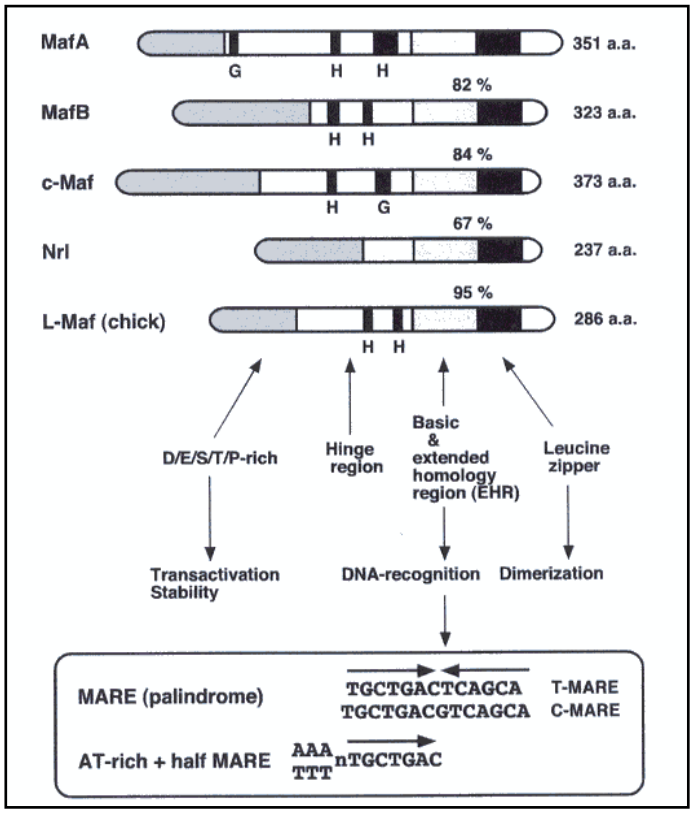


Figure 8: Structures and DNA recognition sequences of Maf family transcription factors. Schematic representation of the structure of human (MafA, MafB, c-Maf and Nrl) and chicken (L-Maf) large Maf factors and their amino acids identities (%) with MafA are shown.

The Maf family is defined by the presence of an additional homologous domain called the extended homology region (EHR) or ancillary domain, which also contacts DNA. The EHR region is rich in hydrophobic residues which, intramolecularly, interact determining the amphipathic α-helix conformation on the N-terminal side. The Maf extended homology region is conserved only within the Maf family and, together with the basic region, defines the DNA binding specific to Mafs. Recent NMR studies on MafG,[9] a small Maf, clarified the solution structure of the well conserved Mafs DNA-binding domain (residues 1-76). The analysis revealed the presence of three α-helices, H1 (residues 26 to 30), H2 (residues 35 to 39), and H3 (residues 46 to 99), in MafG. The Maf-specific structural element EHR consists of H1, H2, and the N-terminal region of H3 (**Figure 9**). H3 (residues 46 to 99) corresponds to the bZip structure.[21]

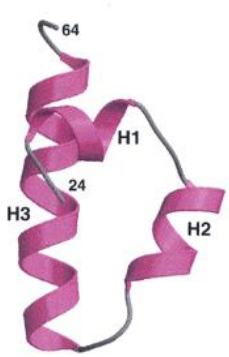


Figure 9: Ribbon diagram of the energy-minimized average structure of MafG (1-76). The α-helices are shown in purple and labelled

H1 and H2 within the EHR interact with the basic region (residues 46 to 62 in H3 of MafG) through both hydrophobic and hydrophilic interactions. The helices are stabilized by hydrophobic interactions among H1, H2 and the basic region. In particular these interactions involve the following residues: Leu 24 from the N-terminal region; Leu 29 and Val 30 from H1; Met 32 from a turn between H1 and H2; Leu 37 and Leu 41 from H2; Leu 44 from a turn between H2 and H3; Ile 49, Leu 52 and Leu 59 from H3. Conservation of these hydrophobic residues and the N-terminal capping boxes among Mafs indicates that there is a similarity in the 3D structures of their DNA-binding domains. Most of the basic amino acid residues reside on one side of the protein surface: Arg 35 from H2 and Lys 46, Lys 53, Arg 56, Arg 57 and Lys 60 from H3 (**Figure 10**).[9] In addition, loss of DNA binding was reported for the R57E mutant of MafG and the corresponding mutants of MAfK, MafF and MafB.[23,24]

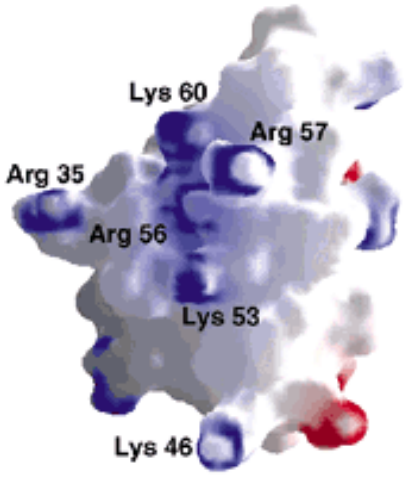


Figure 10: Elrctrostatic potential surfaces of MafG (1-76). Positive and negative potentials are in blue and red, respectively.

Therefore the basic cluster on the protein surface is probably involved in the interaction with target DNA sequences. MafG contains a helix-turn-helix (HTH) motif between H2 and H3. The HTH motif is one of the most common DNA-binding motifs and is found in many eukaryotic and prokaryotic transcriptional regulatory proteins. The second helix of the motif usually binds to the major groove of DNA and, therefore, is mainly involved in recognition and interaction with specific DNA sequences.[22] In MafG the interhelical angle between the H2 and H3 (149°) is considerably larger than the typical value for the common HTH motif (~110°). Thus, MafG may have a protein-DNA interaction mode distinct from that of the ordinary HTH-type DNA-binding domains.[8] All large Maf proteins have an N-terminal serine/proline/threonine-rich acidic activation domain that is absent in the "small Maf" proteins. However, both large and small Maf proteins have a high degree of similarity in the C-terminal basic and leucine zipper domains (**Figure 11**).[23]

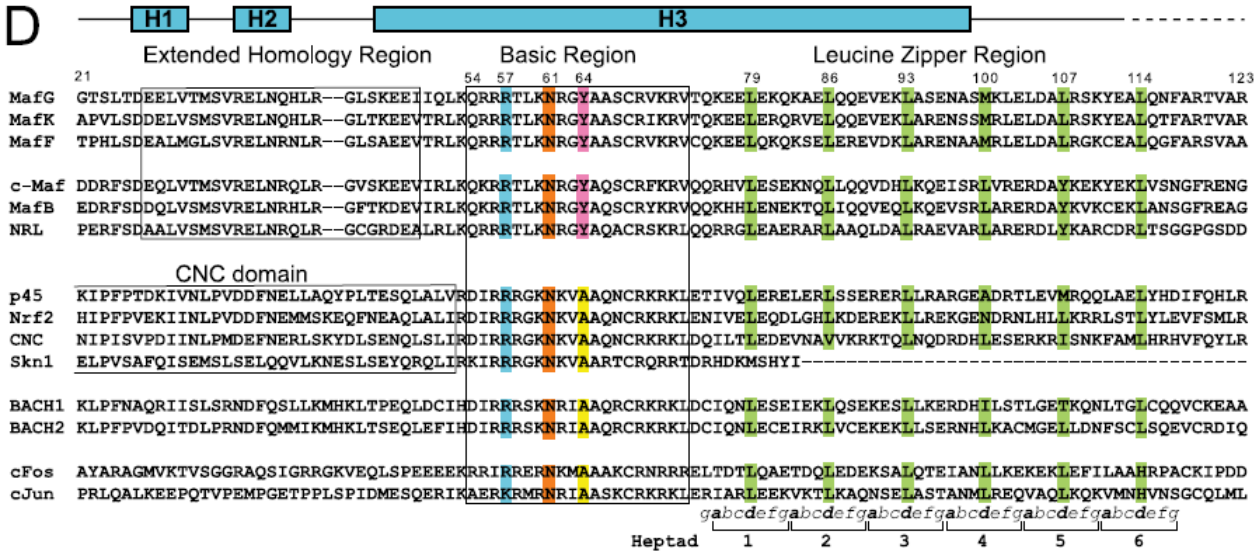


Figure 11: The EHR of Maf proteins, the basic regions, and the partial CNC domains of the CNC family are boxed. H1, H2, and H3 along the top indicate the α-helical regions. The amino acid sequences of the EHR and CNC domains are aligned to each other according to structural correspondence. The amino acid positions in the lucine heptad repeats are indicated at the bottom, and amino acids at position “d” are highlighted in green.

1.2.3. DNA target recognition mode

The AP1 super-family is characterized by the ability to bind TRE (12-O-tetradecanoyl phorbol 13-acetate (TPA)-responsive element) or CRE (cAMP- responsive element) DNA sequences (**Figure 12**) through their bZIP domain.

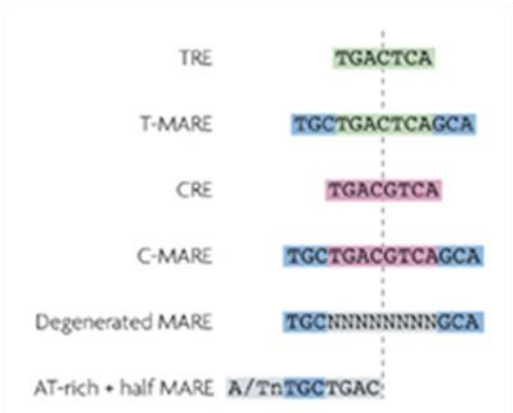


Figure 12: Members of the AP1 family have different DNA recognition sequences: TRE (12-*O*-tetradecanoyl phorbol 13-acetate (TPA)-responsive element), CRE (cAMP-responsive element) and MAREs (Maf recognition elements). Maf proteins can bind to a T-MARE (containing a TRE core), a C-MARE (containing a CRE core), a degenerated MARE or a half MARE flanked by AT-rich sequences.

However, of the various bZip transcription factors, only the Maf family members recognize the exceptionally long *cis*-acting DNA sequence of the MARE. A TRE or CRE core flanked on each side by GC sequences constitutes the MARE (TGCTGACTCAGCA or TGCTGACGTCAGCA) (**Figure 13**). No other member of the bZip family requires the GC flanking sequence adjoining the core.[21]

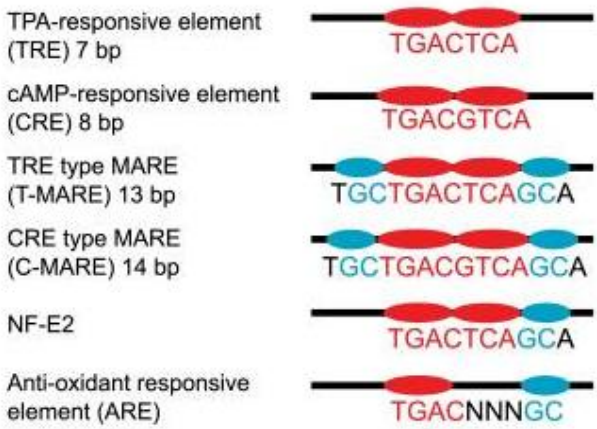


Figure 13: Consensus DNA sequences recognized by bZip transcription factors. The TRE and the CRE lie within the T-MARE and the C-MARE, respectively, and act as the established core sequences of these MAREs. Three bases on each side of the core form the flanking

sequence, which is critical for the recognition of the MARE by the Maf protein family. The NF-E2 binding site and ARE are symmetrical sequences composed of half of the TRE site and half of the MARE site with one flanking sequence. The core and flanking regions are in red and blue, respectively, cAMP, cyclic AMP.

Although the TGC motif is crucial for Maf binding, the TRE or CRE core can be more degenerate. It is noteworthy that many natural Maf target genes contain only half of a MARE palindromic site. However, as homodimers, Maf proteins can also bind half MARE sites if they are flanked by 5-AT-rich sequences. Alternatively, but not exclusively, some of these Maf target genes might be controlled by Maf- containing heterodimers, with the Maf protein binding to the half MARE site.[9]

Previous reports suggested that at least two structural elements are responsible for the unique DNA recognition property of the Maf proteins. One is a tyrosine residue in the basic region, and the other is the EHR domain localized on the N-terminal side of the basic region, both of which are highly conserved in Maf proteins. To explore the structural basis by which Maf proteins recognize the unique DNA sequence of the MARE, Kurokawa et all. performed X-ray crystallography of a MafG homodimer in complex with its cognate DNA as previously mentioned. They obtained the crystal of MafG bound to a 15-bp MARE sequence termed MARE25. Kurokawa et all. also compared the crystal structure of the MafG homodimer with the crystal structures of Fos-Jun and Skn1, both of which do not demand the GC flanking region for DNA binding, and found that a single α-helix of one monomer grips the half-site of the DNA in the major groove in all three cases. Surprisingly, Maf-specific structural elements, including Tyr64 and Thr58 residues in the basic region and EHR en bloc, were not involved in the base recognition of the extended region. However, in sharp contrast to the case of Fos-Jun, invariant Asn61 in the basic region of MafG has a distinct side-chain orientation enabling the recognition of the extended MARE sequence, and another invariant residue, Arg57, contacts with the flanking guanine base. Thus, the Maf-specific elements appear to stabilize the unique side-chain orientations of conserved Asn61 and Arg57 toward the extended GC bases through indirect interactions involving DNA backbone phosphates.[21]

The MARE25 contains the NF-E2 binding site, which is critical for erythroid-specific gene regulation and was previously shown to bind the Maf homodimer tightly. Since MARE25 contains one nucleotide deviation from the GC-(TRE core)-GC consensus sequence, it is possible to assess the singular DNA recognition of each MafG monomer. The asymmetric unit includes two MafG fragments and one

double-stranded 15-bp DNA. Each MafG monomer in the dimeric MafG-DNA complex recognizes a slightly different DNA sequence but displays a similar overall structure. The asymmetry of the DNA sequence used in the crystallization allows a distinction to be made between the two MafG monomers (called subunit A and subunit B). Subunit A is in contact with the GA flanking region and subunit B is in contact with the GC flanking region. The structural differences between subunits A and B of MafG appear to be due simply to the different crystal packing. Like GCN4 and Fos-Jun, the MafG homodimer embraces the major groove of the DNA via its two long H3 helices akin to a pair of chopsticks (**Figure 14**).[21]

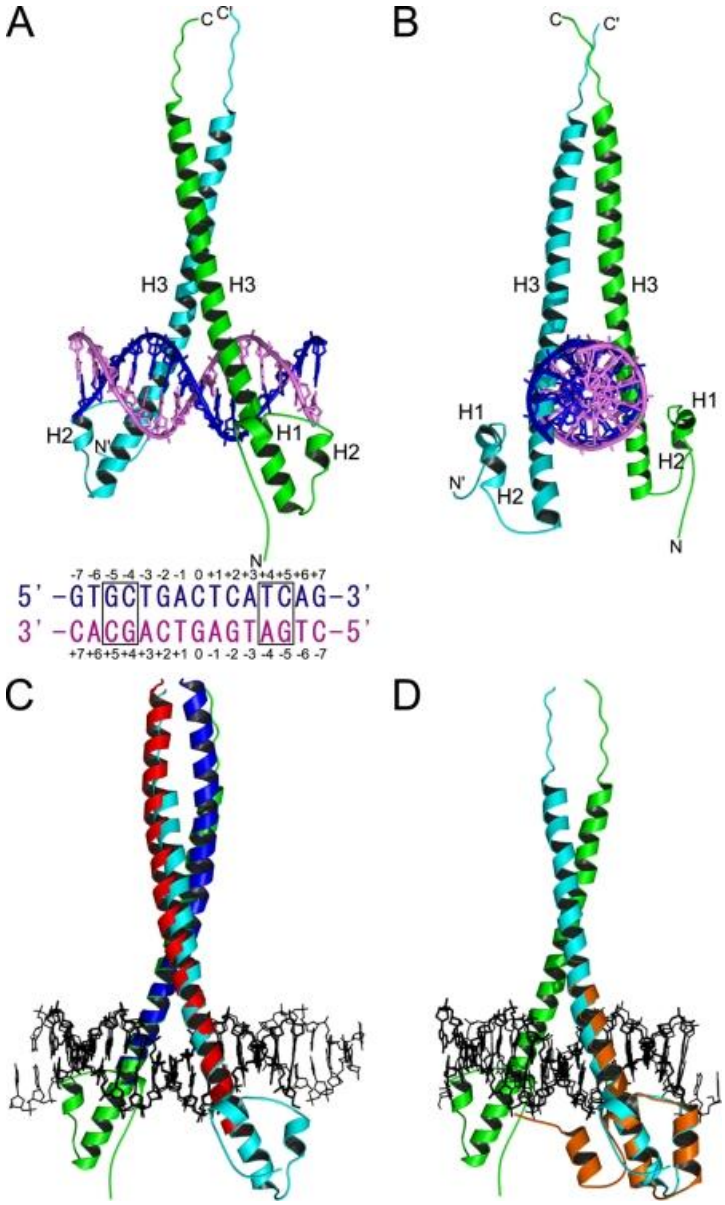


Figure 14: Main-chain structure of MafG in complex with MARE DNA. (A and B) Orthogonal views of MafG in complex with its cognate DNA. MafG consists of three helices: H1 and H2 in the EHR and H3 in both the basic region that recognizes DNA and the leucine zipper region that forms a dimer with the other subunit. The sequence of the duplex oligonucleotide MARE25 that was used for cocrystallization is shown (A). Subunit A (in green) is in contact with the GA flanking region, and subunit B (in cyan) is in contact with the GC flanking region. (C) Comparison with Fos (in red)-Jun (in purple) in complex with DNA. (D) Comparison of MafG with Skn1 (in orange). The protein regions are used for superimposition. Structural alignment yields $rmsd_{57C\alpha}$ of 2.4 Å when the AP1 half-cores are superimposed and 1.8 Å when only the protein regions are superimposed.

The DNA in the complex is in a straight B form and has an average helical rise and twist of 3.2 Å and 34.6°, respectively. Furthermore the MafG H1 and H2 region is essentially similar to Skn1 H1 and H2 region. Showing very good agreement with the case for AP1-type bZip factors, conserved Ala65 residue of MafG create Van der Waals contacts with the T base at positions +1 and -3. Arg69 of subunit B forms a hydrogen bond with the G base at position 0 through a water molecule. A series of salt bridges is formed between Arg62 and the phosphate backbone of thymine at position +1, between Arg69 and the phosphate backbone of adenine at position -1, and between Lys71 and the phosphate oxygen of thymine at position -3. Arg56 of subunit B interacts with the phosphate backbone of thymine at position -7. Asn61 of MafG subunit A forms a direct hydrogen bond with adenine (position -4) in the GA flanking sequence (**Figure 15B**) and an indirect hydrogen bond with cytosine (position +2) of the opposite sequence (**Figure 15A**) in the core region of MARE25. Asn61 of MafG subunit B creates van der Waals forces with cytosine (position -4) in the GC flanking region of MARE25 (**Figure 15C**). Thus, MafG Asn61 recognizes mainly the flanking region, although it forms weak indirect hydrogen bonding with the TRE core region. In stark contrast to MafG, the invariant Asn511 of Skn1 interacted only with cytosine (position +2) in the TRE region (**Figure 15D**).[21]

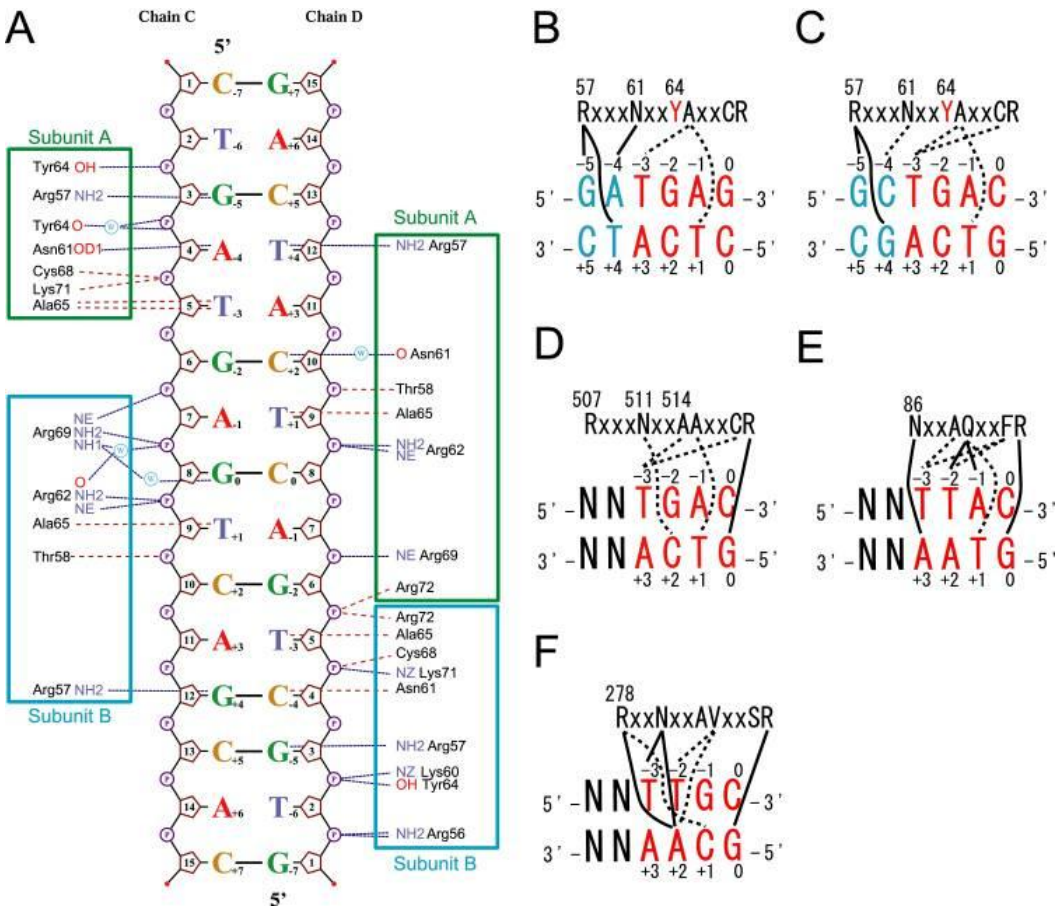


Figure 15: MARE recognition by the basic region of MafG and comparison with AP1-type bZip factors. (A) The phosphate groups of MARE25 engaged by the MafG protein are shown. Chain C and chain D are the bottom and top strands of MARE25, respectively. The figure was drawn by NUCPLOT. (B to F) The conserved amino acid side chains of the motifs that make direct contacts with the DNA bases of the consensus sequence are shown for MafG subunit A (B), MafG subunit B (C), Skn1 (D), Pap1 (E), and C/EBP (F). Hydrogen bonds and van der Waals contacts are shown as solid and broken lines, respectively. PDB accession numbers 1SKN for Skn1 (D), 1GD2 for Pap1 (E), and 1GU4 for C/EBP (F) were used to estimate the DNA base recognition mode.

Arg57 in MafG subunit A forms critical hydrogen bonds with the O6 of guanine at position -5 and the O4 of thymine at position +4 (**Figure 15A** and **B**), while Arg57 in MafG subunit B forms hydrogen bonds to the O6 of guanine at position -5 and the O6 of guanine at position +4 (**Figure 15A** and **C**). It should be noted that Arg57 of MafG participates in DNA base recognition by changing its side-chain orientation; Arg57 in MafG shows a distinct side-chain orientation compared to the corresponding arginine residues of AP1-type and CNC-type factors. In fact, Arg507 and Arg143 residues are the residues corresponding to Arg57 of MafG in Skn1 and Fos, respectively, but the side chain of Arg507 in Skn1 interacts with a backbone phosphate of DNA, and that of Arg143 in Fos is exposed to solvent,[24] and these side chains are not involved in DNA base recognition (**Figure 16**).

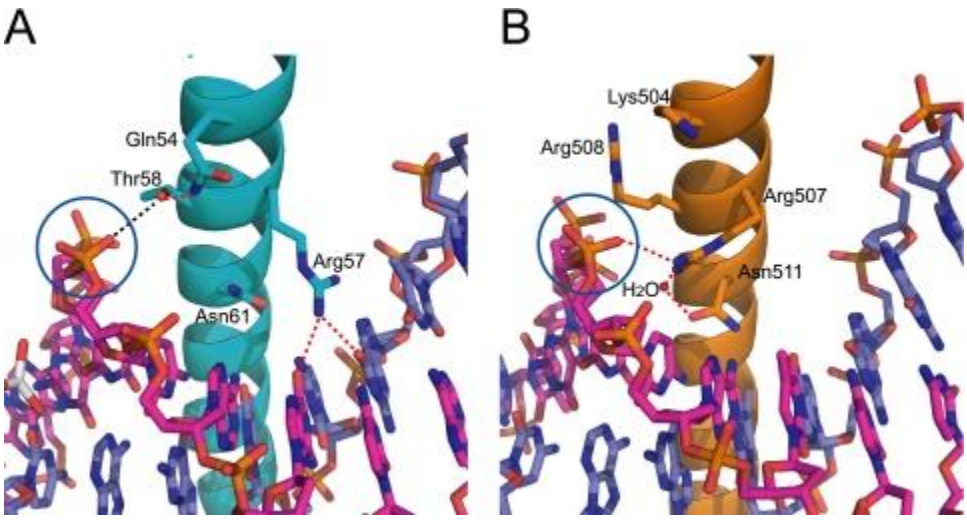


Figure 16: Interactions of the DNA binding domain of MafG with DNA. Van Deer Waals forces between Maf-specific Thr58 and the DNA backbone phosphate. The hydrogen bond between Gln54 and Thr58 fixes the Thr58 side chain in an optimal orientation against the backbone phosphate. **B**) The DNA backbone phosphate referred to above interacts with Arg507 of Skn1. Arg507 corresponds to ARG57 of MafG and forms indirect contacts with the invariant Asn511 via water molecule, thereby stabilizing the orientation of the Asn511 side chain for recognition of the code TRE half-region.

Utilization of Arg57 for DNA recognition makes the Maf subfamily possess a unique and exceptionally long consensus protein sequence for DNA base recognition (13 residues, Rxxx NxxYAxxCR) (**Figure 15B** and **C**). In contrast, those of AP1-type/ CNC-type factors (NxxAAxxC/SR) and the yeast AP1-type factor Pap1 (NxxAQxxFR) contain a 9-residue sequence, and the arginine residue corresponding to Arg57 of MafG is not included in the interaction with DNA (**Figure 15D** and **E**). Another characteristic protein sequence is RxxNxxAVxxSR for the C/EBP subfamily, where the arginine residue involving DNA recognition is 1 amino acid closer to the asparagine residue than in MafG. The arginine residue of C/EBP at the position corresponding to Arg57 of MafG in the consensus sequence does not participate in DNA base recognition (**Figure 15F**). Thus, it is likely that the distinct side-chain orientation of Arg57 and Asn61 explains why MafG requires the flanking region for binding and recognizes the TRE core with specificity relatively broader than that of AP1-type and CNC-type transcription factors.[21] An interesting achievement in Maf folding knowledge was obtained examining the effect of DNA binding on the conformation of the Maf bZIP domain and ancillary DNA-binding region using circular dichroism (CD). Comparison of the CD spectra of free Maf and Maf in the presence of a consensus binding site oligonucleotide reveals a large change in ellipticity at 208 and 222 nm (**Figure 17**).

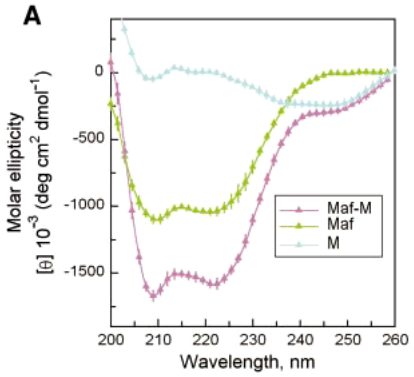


Figure 17: Comparison of the CD spectra of free Maf (green), the oligonucleotide containing a consensus Maf-binding site M (blue) and the Maf-M complex (purple).

This change is consistent with an increase in the α-helicity of Maf upon binding the optimal recognition sequence. In the absence of DNA, Maf homodimers have a α-helix content that is consistent with folding of the leucine zipper prior to DNA binding. It is therefore likely that the ancillary DNA-binding region adopts a conformation of high α-helix content upon Maf binding to the consensus recognition sequence.

Maf family members recognize a DNA sequence element that is almost twice as long as elements recognized by other bZIP proteins. Substitution of a single base pair in one of the extended recognition elements (position -6) reduces the change in Maf α-helicity induced by DNA binding by approximately half (**Figure 18**).

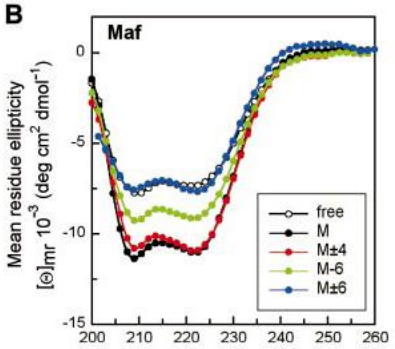


Figure 18: Comparison of Maf conformations at different recognition sites. The mean residue ellipticity of maf241-369 bound to different recognition sites (solid symbols) was calculated by subtraction of the ellipticity of each oligonucleotide from the ellipticity of the complex formed by Maf at each site, and was compared with that of free Maf241-369 (open symbols). There were no significant difference between the spectra of the oligonuclotides alone.

Substitution of the same base pair in both half-sites (positions ±6) abolishes the induced α-helical conformation of Maf (**Figure 18**, blue line). This base pair is therefore essential for Maf to undergo a conformational change upon binding to DNA. Other base substitutions within the extended recognition elements (positions ±5 and positions ±7) have smaller effects on Maf conformation. Base substitutions within the CRE (position ±4) do not influence the α-helix content of Maf (**Figure 18**, red line). In contrast, these base substitutions eliminates the coil-to-helix transition of Fos-Jun heterodimers (**Figure 19**).

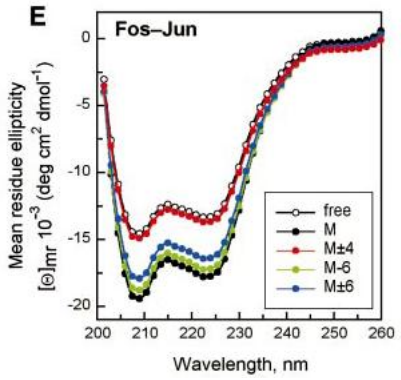


Figure 19: Comparison of Fos-Jun conformations at different recognition sites. The mean residue ellipticity of Fos 139-200--Jun257-318 free and bound to different recognition sites was plotted as in B.

No increase in the α-helix content of either Maf or Fos-Jun is observed in the presence of a GC co-polymer oligonucleotide. Consequently, base substitutions in the core and extended recognition elements have reciprocal effects on the conformational changes of Maf and Fos-Jun.

1.2.4. References

1) Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K., Kawai, S., *Proc. Natl. Acad. Sci USA* **1989**, 86, 7711-7715.
2) Kataoka, K., Nishizawa, M., Kawai, S., *J. Virol.* **1993**, 67, 2133-2141.
3) Kataoka, K., *J. Biochem.* **2007**, 141, 775-781.
4) Coolen, M., Sii-Felice, K., Bronchain, O, Mazabraud, A., Bourrat, F., Retaux, S., Felder-Schmittbuhl, M.,P., Mazan, S., Plouhinec J., L., *Dev. Genes Evol.* **2005**, 215, 327-339.
5) Motohashi, H., Shavit, J. A., Igarashi, K., Yamamoto, M., Engel, J. D., *Nucleic Acids Research* **1997**, 15, 2953-2959.
6) Blank, V., Andrews, N. C., *Trends Biochem. Sci.* **1997**, 22, 437-441.
7) Motohashi, H., Katsuoka, F., Shavit, J. A., Engel, J. D., Yamamoto, M., *Cell* **2000**, 103, 865-875.
8) Kusunoki, H., Motohashi, H., Katsuoka, F., Morohashi, A., Yamamoto, M., Tanaka, T., *Nature structural biology* **2002**, 9, 252-256.
9) Eychène, A., Rocques, N., Pouponnot, C., *Nature Reviews* **2008**, 8, 683-693.
10) Chen, Q., Dowhan, D. H., Liang, D., Moore, D. D., Overbeek, P. A., *J. Biol. Chem.* **2002**, 277, 24081-24089.
11) Rocques, N., *Mol. Cell.* **2007**, 28, 584-597.
12) Friedman, J. S., *J. Biol. Chem.* **2004**, 279, 47233-47241.
13) Blank, V., *J. Mol. Biol.* **2008**, 376, 913-925.
14) Motohashi, H., O'Connor, T., Katsuoka, F., Engel, J. D., Yamamoto, M., *Gene* **2002**, 294, 1-12.
15) Kataoka, K., Noda, M., Nishizawa, M., *Oncogene* **1996**, 12, 53-62.
16) Coolen, M., *Dev. Genes Evol.* **2005**, 215, 327-339.
17) Lecoln, L., Sii-Felice, K., Pouponnot, C., Eychène, A., Felder-Schmittbuhi, M., *Gene Expr. Patterns* **2004**, 4, 35-46.
18) Kataoka, K., *J. Biochem.* **2007**, 141, 775-781.
19) Yang, Y., Cvekl, A., *J. Biol. Med.* **2007**, 23, 2-11.
20) Cordes, S. P., Barsh, G. S., *Cell* **1994**, 79, 1025-1034.
21) Kurokawa, H., Motohashi, H., Sueno, S., Kimura, M., Takagawa, H., Kanno, Y., Yamamoto, M., Tanaka, T., *Mol. Cell. Biol.* **2009**, 29, 6232-6244.
22) Witntjens, R., Rooman, M., *J. Mol. Biol.* **1996**, 262, 294-313.
23) Kataoka, K., *Mol. Cell. Biol.* **1995**, 15, 2180-2190.

24) Kataoka, K., Fujiwara, K. T., Noda, M., Nishizawa, M., *Mol. Cell. Biol.* **1994**, 14, 7581-7591.

1.3.Mafs and multiple myeloma

1.3.1. Introduction

Large Maf genes are bona fide oncogenes as highlighted from their initial isolation to their involvement in human cancer. Maf proteins can transform primary cells[1,2-4] and their oncogenic activity has been validated in vivo both in birds and mammals: the As42 avian retrovirus, expressing v-Maf, induces tumours in chickens,[5,6] and transgenic mice expressing Maf in the lymphoid compartment develop T cell lymphomas.[7] In humans, Maf genes are overexpressed in 60% of human angioimmunoblastic T-cell lymphomas (AITLs)[7,8] and in 50% of cases of multiple myeloma (MM), and contribute directly to cancer progression.[9] The oncogenic activities of the large Maf proteins have all been tested in vitro except for that of NRL, the only member that has not yet been found to be deregulated in human cancer. In contrast to small Maf proteins, MAF, MafB and MafA can efficiently transform primary fibroblasts *in vitro*,[1,2-4] and, interestingly, this does not require an activating mutation (no such mutations have been found within large Maf genes in human cancers), although it can be enhanced by point mutation.[2,10] MafA and MAF display the strongest oncogenic activity, whereas MafB is less effective at transforming cells.[3,4] This weaker activity of MafB correlates with its reduced expression level compared with the other large Maf proteins, and possibly reflects the fact that MafB is a more unstable protein.[4] Indeed, a threshold level of Maf expression might be required for transformation as only mice carrying a high copy number of the MAF transgene develop T-cell lymphoma.[7] Moreover, Maf genes are also overexpressed in human cancers, suggesting that transformation requires a high level of Maf expression. The transforming activity of the Maf proteins depends on their ability to act as transcription factors.[2-4,10,11] Accordingly, some of their target genes, such as those that encode cyclin D2, integrin β7 and ARK5 (also known as NUAK1), have deregulated expression in MM,[9,12-15] AITL and in Maf transgenic mice.[7] However, the capacity of the large Maf genes to induce transformation does not strictly correlate with their transactivation strength, as measured using a reporter luciferase gene under the control of synthetic MARE sequences.[3,10] In summary, when expressed at high levels, unmutated large Maf genes display oncogenic activity in cell culture, animal models and in human cancer.[16]

1.3.2. Multiple myeloma

Multiple myeloma (MM) is a malignant disorder characterized by the clonal proliferation of plasma cells (PCs) in the bone marrow (BM). It constitutes about 10% of all hematologic cancers and its annual incidence, age-adjusted to the 2000 United States population, is 4.3 per 100,000.[17] Occurrence of the disease is more common in men than women and twice as high in African-Americans compared to Caucasian.[18] The frequent invasion of the adjacent bone by malignant PCs causes osteolytic bone lesions and fractures; anemia occurs in 70% of patients at diagnosis, and renal dysfunction and hypercalcemia in 50% and 25% of patients, respectively.[18] The clinical course of MM is quite heterogeneous: frequently the tumor can be preceded by a condition known as monoclonal gammopathy of undetermined significance (MGUS), a pre-malignus disorder of the plasma cells characterized by a decrease in the number of plasma cells (<10%) and monoclonal proteins in the bone marrow, in the absence of osteolytic lesions, anemia, hypercalcaemia or depletion of kidney functions. Multiple myeloma is usually confined in the bone marrow (intramedullary MM) where the myeloma cells strictly depend on the marrow microenvironment that assures them growth, survival, and resistance to conventional drug therapies. The final and more aggressive stage of the myeloma is called plasmacellular leukemia (PCL) and it is characterized by the presence of cancerogenic plasma cells in extramedullary sites such as blood vessels and skin (**Figure 20**).

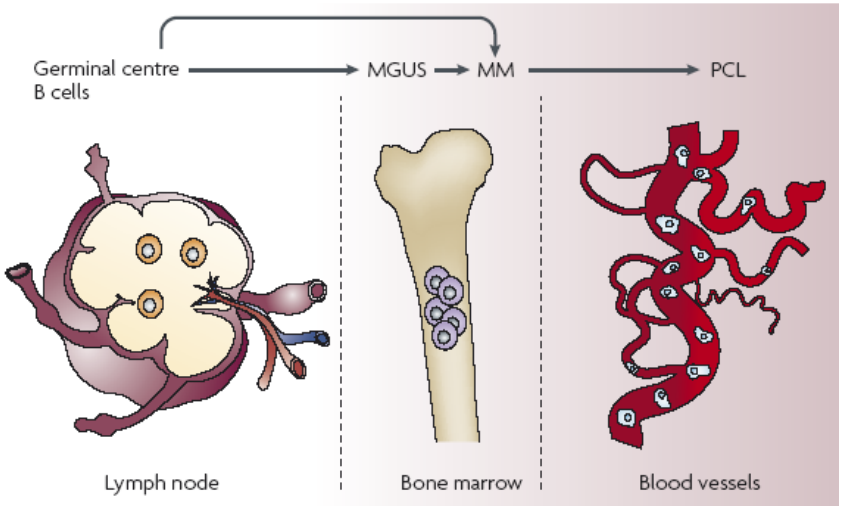


Figure 20: Progression of multiple myeloma. MM lacks pathognomonic genetic lesion(s) that could account for all MM cases, and is instead characterized by a marked interpatient heterogeneity in its genetic background.[19]

1.3.3. Large Mafs and MM

Genes encoding "large" Maf proteins, c-Maf, MafB and MafA, are involved in three chromosomal translocations representing primary oncogenic events in MM. These translocations involve the immunoglobulin heavy chain locus (IGH@), at chromosome 14q32, and are mostly mediated by errors in one of the three B-cell specific DNA modification mechanisms: VDJ recombination, IGH switch recombination, and somatic hypermutation. The effect of these translocations is the juxtaposition of strong Ig enhancers to non-random oncogene-harboring chromosomal loci, with the consequent deregulation of their expression; one of the three groups of recurrent IGH translocations in myeloma is represented by MAF (16q23), MAFB (21q12) and MAFA (8q24.3) genes, that are partners of IGH locus respectively in 5, 2 and less than 1% of MM cases. The chromosomal translocation t(14;16) was associated with a poor prognosis,[20,21] and its presence is considered sufficient to place patients in the high risk category based on stratification models proposed by the Mayo Clinic[22] and the International myeloma Working Group,[23] which recommends the molecular cytogenetics analysis in routine clinical practice for an adequate prognostic characterization.

Although a recent retrospective analysis of a large number of patients does not confirm the adverse prognostic impact of t(14;16),[24] there is no doubt that the deregulation of MAF genes plays an important pathogenetic role in the context of the genetic heterogeneity characterizing myeloma. Transcriptional activation of CCND2 is probably the mechanism through which oncogenic Maf proteins induce cell proliferation; it was in fact observed that MM Maf group is characterized by high proliferative index. XBP is a bZIP family member involved in the differentiation of plasma cells. It is abundantly expressed in multiple myeloma and it could contribute to the expression increase of c-Maf. Indeed, knock-out mice show that an increasing in XBP expression in the lymphoid compartment is correlated with altered expression levels both of MafB and c-Maf. The mechanism that subtends the alteration of the expression profile of large Mafs remains unclear as the evidence that XBP is able to bind Maf proteins promoter. Nevertheless, these observations support the hypothesis for which other events, independently from chromosomal translocations, could lead to an increment of large Mafs gene expression. In the multiple myeloma, the principal c-Maf gene targets are: cylinD2, integrinβ7, CCR1 (Chemokine receptor C-1) and ARK5 (protein kinase AMP-activated kinase-related 5) (**Figure 21**).

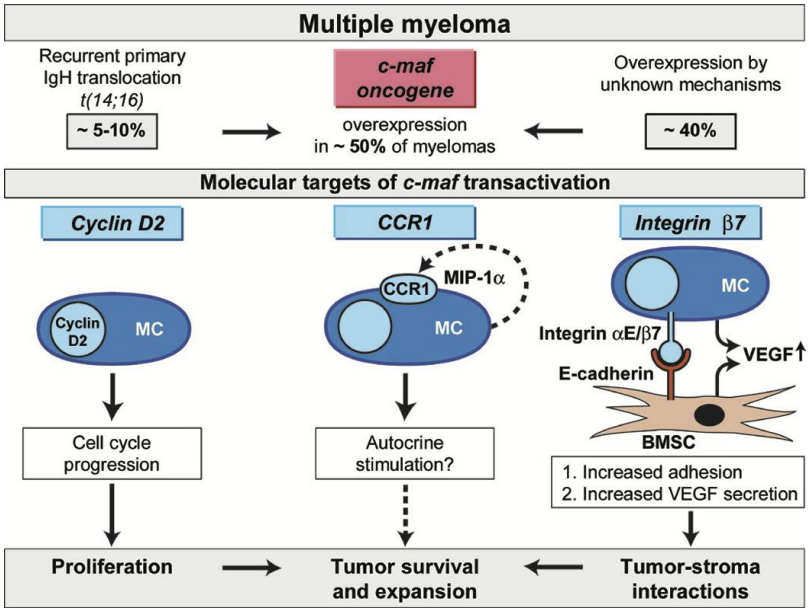


Figure 21: c-Maf oncogene targets in multiple myeloma.

The over-expression of other genes belonging to Maf signature gives to myeloma cells a greater ability of invasion. Another fundamental pathogenetic aspect on which Maf deregulation acts is the interaction between tumor cells and stroma, which is favored by over-expression of integrin B7. This molecule allows myeloma cells to adhere more effectively to bone marrow stromal cells expressing on cell surface the integrin-binding protein E-cadherin. These cell-cell interactions induce the expression of pro-angiogenic cytokine VEGF, which helps to create a more favorable environment to the tumor (**Figure 22**).

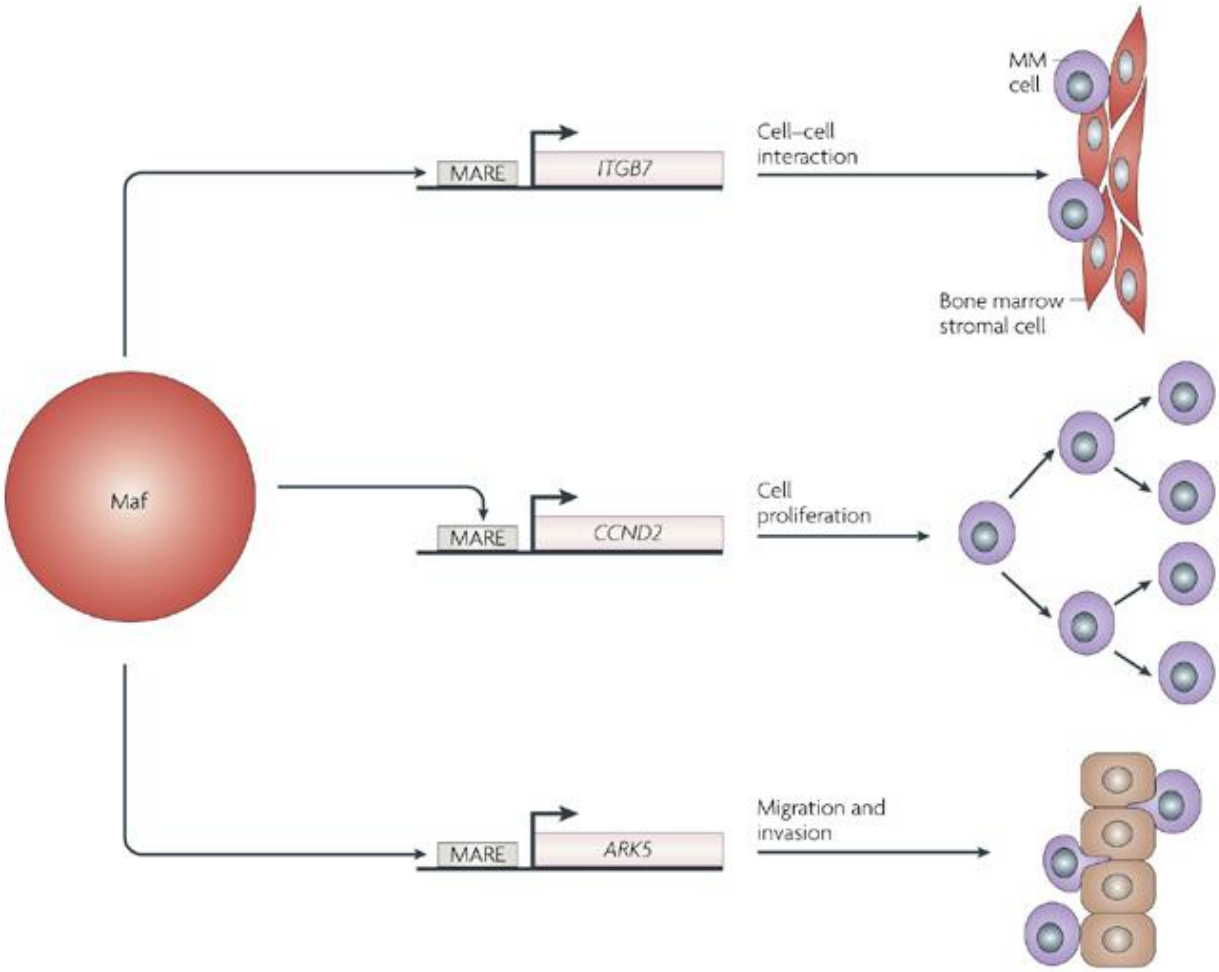


Figure 22: Overexpression of large Maf proteins is involved in multiple myeloma (MM) progression by deregulating several cellular processes through the regulation of distinct gene expression. Large Maf proteins promote MM cell proliferation through inducing expression of *CCND2*, pathological interactions between bone marrow stromal and tumor cells by inducing expression of *ITGB7* and cell migration and invasion by inducing expression of *ARK5*. MAREs (Maf-recognition elements) are present in the promoter of these genes.

1.3.4. References

1) Kataoka, K., Fujiwara, K. T., Noda, M., Nishizawa, M., *Moll. Cell. Biol.* **1994**, 14, 7581-7591.
2) Kataoka, K., Nishizawa, M., Kawai, S., *J. Virol.* **1993**, 67, 2133-2141.
3) Nishizawa, M., Kataoka, K., Vogt, P. K., *Oncogene* **2003**, 22, 7882-7890.
4) Pouponnot, C, *Oncogene* **2006**, 25, 1299-1310.
5) Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K. T., Kawai, S., *Proc. Natl. Acad. Sci. USA* **1989**, 86, 7711-7715.
6) Kawai, S., *et al*., *Virology* **1992**, 188, 778-784.
7) Morito, N., *et al*., *Cancer Res.* **2006**, 66, 812-819.
8) Murakami, Y., *et al*., *J. Surg. Pathol.* **2007**, 31, 1695-1702.
9) Hurt, E. M., *et al., Cancer Cell* **2004**, 5, 191-199.
10) Kataoka, K., Noda, M., Nishizawa, M., *Oncogene* **1996**, 12, 53-62.
11) Kataoka, K., Shioda, S., Yoshitomo-Nakagawa, K., Handa, H., Nishizawa, M., *J. Biol. Chem.* **2001**, 276, 36849-36856.
12) Mattioli, M., *et al., Oncogene* **2005**, 24, 2461-2473.
13) Suzuki, A., *et al., Oncogene* **2005**, 24, 6936-6944.
14) Zhan, F., *et al., Blood* **2006**, 108, 2020-2028.
15) Monteiro, P., *et al., Biochem. Biophys. Res. Commun.* **2007**, 358, 442-448.
16) Eychène, A., Rocques, N., Pouponnot, C., *Nature Reviews* **2008**, 8, 683-693.
17) Kyle, R. A., Rajkumar, S. V., *The New England Journal of Medicine* **2004**, 351, 1860-1873.
18) Rajkumar, S. V., *Curr. Probl. Cancer* **2009**, 33, 7-64.
19) Mitsiades, C. S., Mitsiades, N., Munshi, N. C., Anderson, K., *Cancer Cell J.* **2004**, 6, 439-444.
20) Fonseca, R., Blood, E., Rue, M., Harrington, D., Oken, M. M., Kyle, R. A., Gordon, W., Dewald. W., Van Ness, B., Van Wier, S. A., Henderson, K. J., Bailey, R. J., Greipp, P. R., *Blood* **2003,** 101, 4569-4575.
21) Zhang, F., Huang, Y., Colla, S., Steward, J. P., Hanamura, I., Gupta, S., Epstein, J., Yaccoby, S., Sawyer, J., Burington, B., Anaissie, E., Hollmig, K., Pineda-Roman, M., Tricot, G., Van Rhee, F., Walker, R., Zangari, M., Crowley, J., Barlogie, B., Shaughnessy, J., *Blood* **2006**, 15, 2020-2028.
22) Dispenzieri, A., Rajkumar, S. V., Gertz, M. A., Lacy, M. Q., *Mayo Clinic Procc.* **2007**, 82, 323-341.
23) Fonseca, R., Bergsagel, P. L., Drach, L., Shaughnessy, J., Gutierrez, N., Steward, A. K., Morgan, G., Van Ness, B., Chesi, M., Minvielle, S., Neri, A., Barlogie, B., Kuehl, W. M., Liebisch, P.,

Davies, F., Chen-Kiang, S., Durie, BGM., Carrasco, R., Sezer, O., Reiman, T., Pilarski, L., Avet-Loiseau, H., *Leukemia* **2009**, 23, 2210-2221.

24) Munshi, N. C., Anderson, K. C., Bergsagel, P. L., Shaughnessy, J., Palumbo, A., Durie, B., Fonseca, R., Stewart, A. K., Harousseau, J. L., Dimopoulos, M., Jagannath, S., Hajek, R., Sezer, O., Kyle, R., Sonneveld, P., Cavo, M., Rajkumar, S. V., San Miguel, J., Crowley, J., Avet-Loisseau, H., *Blood* **2011**, 117, 4696-4700.

1.4. Construction of a DNA binding domain (DBD) structural model for Mafs

Large Mafs, as described in chapter 1.2, are proteins of considerable dimension (300 amino acidic residues), we focused on the DNA binding domain (DBD) as potential biological target for the obtainment of modulators. This domain corresponds to a 75 mer sequence and harbors the LZP region combined to a unique and conserved N-terminal motif called extended homology region (EHR) as shown in **Figure 23.**

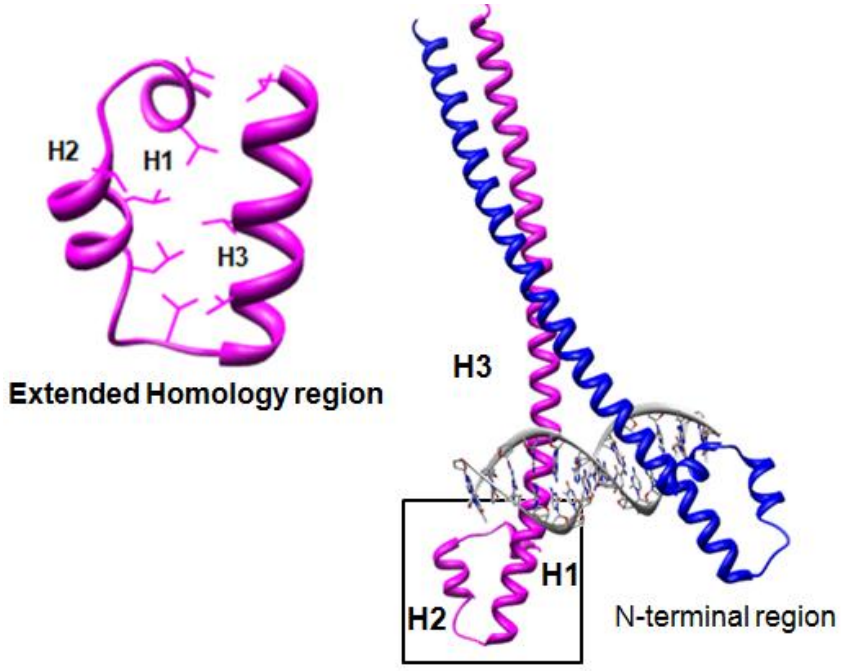


Figure 23: Structure of the dimeric DBD of Mafs. From crystallographic data of Maf-G it is evicted that the EHR region is characterized by a H-T-H motif which comprehends three helices (H1, H2, H3).

The comprehension of the DBD/ciclyn D2 promoter interaction (**Figure 24**) is an important task for the development of a structural model, as in the literature it is available only the model between MafG (a small Maf which is not involved in the pathogenesis of MM) and a TRE containing MARE. will be of importance in order to define the key amino acids (AA) and sequences responsible for the interaction with DNA.

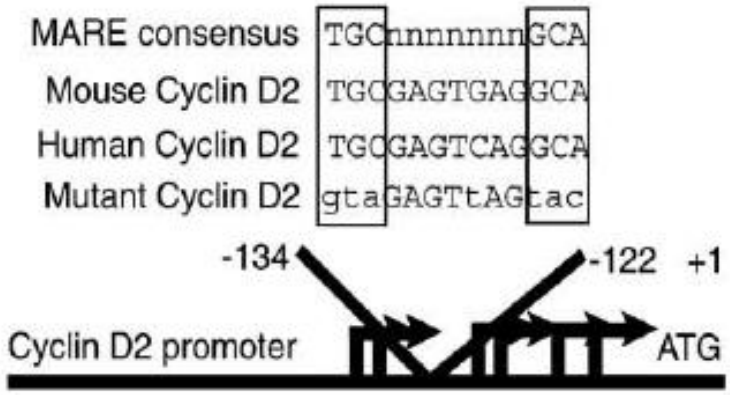


Figure 24: DBD/Cyclin D2 promoter schematic illustration.

In this work we focused on the study of two large Mafs, c-Maf and MafA, approaching the issue from two complementary directions. From an experimental point of view, carrying out the total chemical synthesis of MafA and performing folding and DNA target interactions studies using circular dicroism (CD); this work is material of an article published in 2012 (*Amino Acids* **2012**, 43,1995–2003). A theoretical study was also performed with the aim of creating an homology model of c-Maf. In a second time the design of DBD peptidic modulators was accomplished. C-Maf was appointed as biological target for the realization of modulators considering that its over-expression is strictly related the malignant progression of multiple myeloma, currently untreatable. In particular peptides were designed to interact with the LZP moiety in order to inhibit dimerization which is required for Mafs activation and DNA binding. LZP aminoacidic sequence is the most variable portion of DBD, that could guarantee a fine target modulation. Combining Molecular Mechanics Poisson Boltzmann Surface Area strategy with Computational Alanine Scanning method a small library of peptides (**Figure 25**) as LZP dimerization inhibitors was outlined. Starting from the LZP wild type aminoacidic sequence, two different classes of peptides were studied which bound respectively, or the N-termini region or the C-termini region of LZP.

NAME	SEQUENCE	LENGTH
Peptide wild type 1	RFKRVQQRHVLESEKNQLLQQV	22 aa
Peptide 1	EFKEVQERHVRLSEKNQLIQQV	22 aa
Peptide 1_variante	EFKDVQERHVKLSEKNQLIQQV	22 aa
Peptide 2	EFKEVQERHVRLSEKNQLI	19 aa
Peptide 2_variante	EFKDVQERHVRLSEKNQLLQQI	22 aa
Peptide 4	EFKDVQERHVLESEKNQLL	19 aa

NAME	SEQUENCE	LENGTH
Peptide wild type 2	ISRLVRERDAYKEKYEKL	18 aa
Peptide 5	LSRLLRERDAYKEKYEKI	18 aa

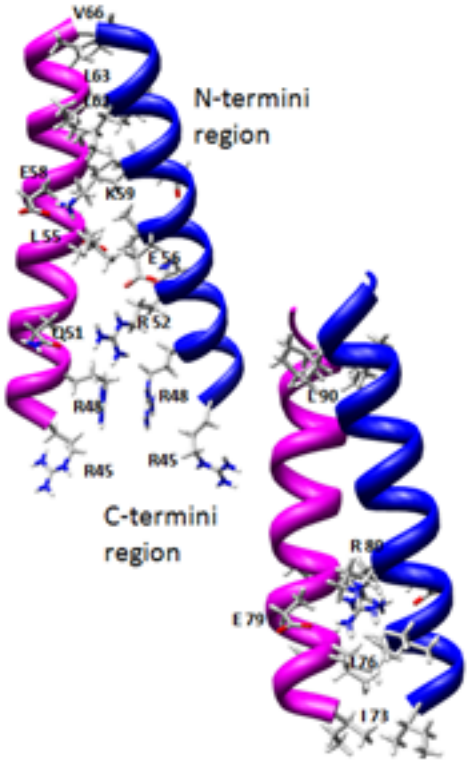


Figure 25: c-Maf LZP inhibitors outlined.

1.5. Chemical synthesis of MafA 75mer DNA binding domain

1.5.1. Microwave assisted solid phase synthesis strategies

The chemical synthesis of the 75mer peptide from MafA DBD represented a tough challenge, not being available previous reports on the preparation of such domain using conventional SPPS protocols. Furthermore, this sequence is particularly rich in Arg residues, which are somewhat difficult to couple. As the usefulness of microwave irradiation was recently demonstrated for the preparation of long peptides and, particularly, for difficult sequences,[1] we planned to use automated MW-SPPS for obtaining 75mer MafA DBD and its different domains (i.e., the EHR, the basic region BR, and the dimerization domain; **Figure 26**).[2]

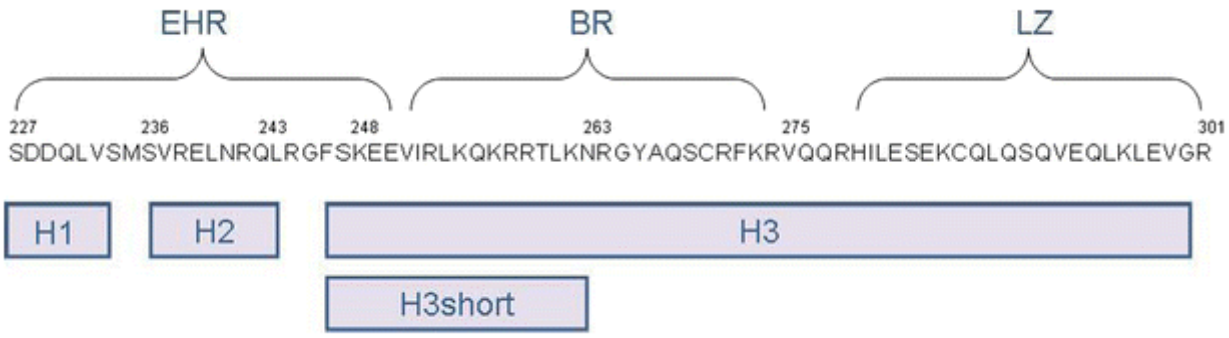


Figure 26: MafA DBD sequence.

We also planned the synthesis of shorter fragments corresponding to H1–H2, to the N-terminus of H3 (named H3 short) and to the sequence H2-H3 short in order to perform circular dichroism (**Table 2**).

Peptide	Length	Protocol[a]	MW_{calc}(Da)	MW_{found}(Da)[b]	t_r (min)[c]	Total yield (%)[d]
248–263 (H3short)[e]	16	A	2,039.49	2,038.18	14.08	60
227–243 (H1H2)	17	A	2,032.28	2,031.20[f]	18.19	50
277–301 (LZ)	25	A	3,020.44	3,019.65	16.91	20
248–275 (BR)	28	A	3,531.19	3,532.30	15.27	18
236–263 (H2H3short)	28	A	3,537.16	3,536.12	16.54	17
227–263 (EHR-H3short)	37	B	4,500.18	4,497.00	18.80	15
227–278 (EHR–BR)	52	B	6,365.33	6,360.50	19.23	10
248–301 H3	54	B	6,622.71	6,621.30	16.39	10
227–301 DBD	75	B	9,041.48	9,034.70	17.25	6.5

Table 2: A)Standard MW assisted single coupling; B) standard MW-assisted single coupling up to 35 residue, then double coupling; MALDI-TOF mass analyses; [c]see text for analytical HPLC gradient; [d]purified peptides.

Rink amide PEG MBHA resin (0.35 mmol/g loading) was chosen for its higher swelling efficiency with respect to Rink amide polystyrene. All the peptides were prepared using Fmoc/tBu methodology and entirely assembled using automatic MW-SPPS, applying different microwave protocols to the coupling steps depending on the amino acid type and the chain length (**Table 2**). In the standard protocol, which was applied to peptides up to 28 residues, couplings were performed using a fivefold excess of Fmoc-amino acids, HOBT/HBTU (5 equiv., 0.45 M in DMF) as activators and DIPEA (10 equiv., 1 M in NMP) as base and with a 5 min microwave irradiation at 75ºC. Fmoc deprotection was achieved using piperidine (20 % in DMF) with a 3 min microwave irradiation at 75ºC. In order to avoid cysteine and hystidine racemization, their couplings were performed at 50ºC. For arginine, which possesses a low reactivity toward peptide couplings, we adopted a double coupling approach (first coupling: 30 min at 75ºC without microwave energy; second coupling: 5 min at 75ºC with microwave energy). This last protocol was used in the elongation of longer peptides starting from the 35th residue, regardless of the amino acid type. All the peptides were N-terminally acetylated (tenfold excess Ac2O and DIPEA in DCM) and cleaved from the resin using TFA and scavengers. The crude HPLC traces indicated a good homogeneity for all compounds (Fig. 2), which were then purified with preparative HPLC and identified using mass spectrometry. It should be underlined that in a relatively short time (70 h) it was possible to obtain 50mg of pure 75mer DBD, thus demonstrating the effectiveness of stepwise MW-SPPS technique also for peptides with a chain length higher than 50 residues.[2]

1.5.2. Circular dicroism experiments

1.5.2.1. Folding properties studies

The CD technique is particularly useful for examining macromolecular structures which in solution lack a well-ordered tertiary structure, as well as for determining dynamic protein/DNA interactions.[3,4] In phosphate buffer (0.1M, pH7.5), the entire 75mer peptide showed a negative band close to 200 nm (amide $\pi \longrightarrow \pi^*$ transition) and a negative shoulder near 222 nm (amide $n \longrightarrow \pi^*$ transition) (**Figure 27**).

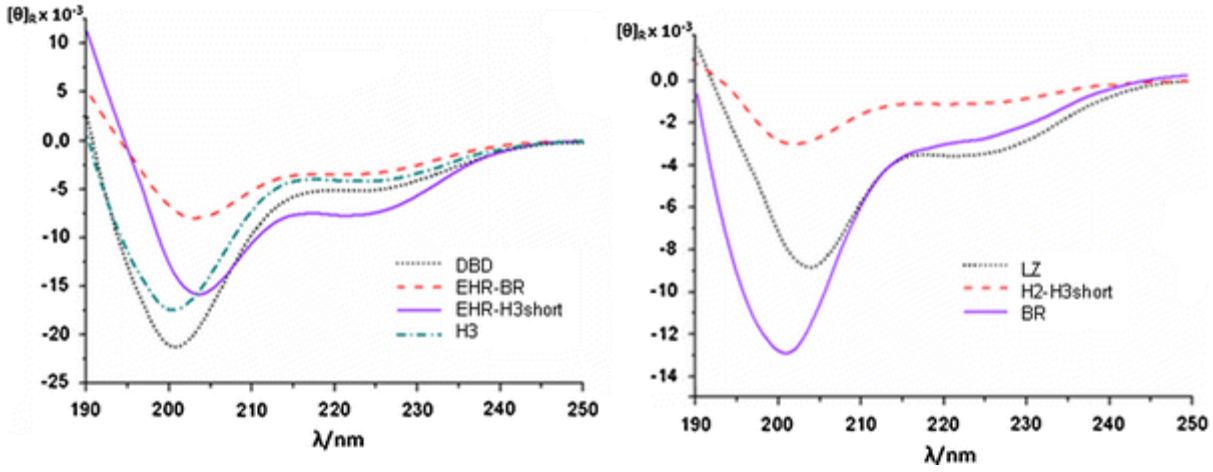


Figure 27: CD spectra of synthesized peptides in phosphate buffer (20μM solution).

The absence of a positive band below 200 nm and the blue shift of the $\pi \longrightarrow \pi^*$ one indicated the presence of different conformations. This behavior was independent from sample concentration (data recorded from 5 to 50M, while at higher concentration, the DBD was not soluble). Accordingly, secondary structure analysis of the CD spectrum with the CONTIN algorithm[5,6] indicated the presence of 46% unordered fraction together with a helix and, surprisingly, β sheet structure (**Table 3**).

Peptide	Buffer				50 % TFE			
	% Helix	% Strand	% Turns	% Unordered	% Helix	% Strand	% Turns	% Unordered
DBD	20	20	14	46	100	–	–	–
H3	15	29	15	41	100	–	–	–
EHR–BR	14	31	15	40	100	–	–	–
EHR-H3short	25	21	14	40	85	–	7	8
H2H3short	6	30	24	40	67	–	14	19
BR	11	30	15	44	70	–	10	20
LZ	20	12	22	46	71	1	13	15
H1H2	9	36	14	41	55	5	11	29
H3short	12	30	8	50	73	–	12	15

Table 3: CD spectra analyzed using CONTIN algorithm.

By analyzing the CD spectra of different DBD fragments (**Figure 28**), it was found that the LZ fragment contained lesserβsheet contribution (α/β ratio, 2:1) than the other peptides. The entire DBD and the EHR-H3short (227–263) sequence possessed an α/β ratio of 1:1, while for other fragments this ratio was in favor of β sheet. These findings suggest that the presence of both the EHR and LZ is crucial for the stabilization of a helix structure. Nevertheless, in all cases the addition of 50% TFE led to the increase of the helix conformation indicating that β sheet structure is favored by a hydrophilic environment (**Table 3**). EHR domain is a peculiarity of Maf proteins and, from the literature data on MafG, hydrophobic interactions between residues from H1 and N-terminus of H3 (H3short) are evicted.[7] In order to ascertain if they are fundamental for the folding, we performed interaction studies between EHR-H3short (227–263) peptide and H1H2 (227–243) and H3short (248–263) fragments. In order to minimize the contribution of β sheet conformation, the experiments were carried out in 20% TFE. Under these conditions, all the three peptides were found almost helical (EHR-H3short: 61%, H1H2: 48 %, H3short: 76 %; helix percentage was estimated using the procedure of Chen et al.[8]; **Figure 29**).

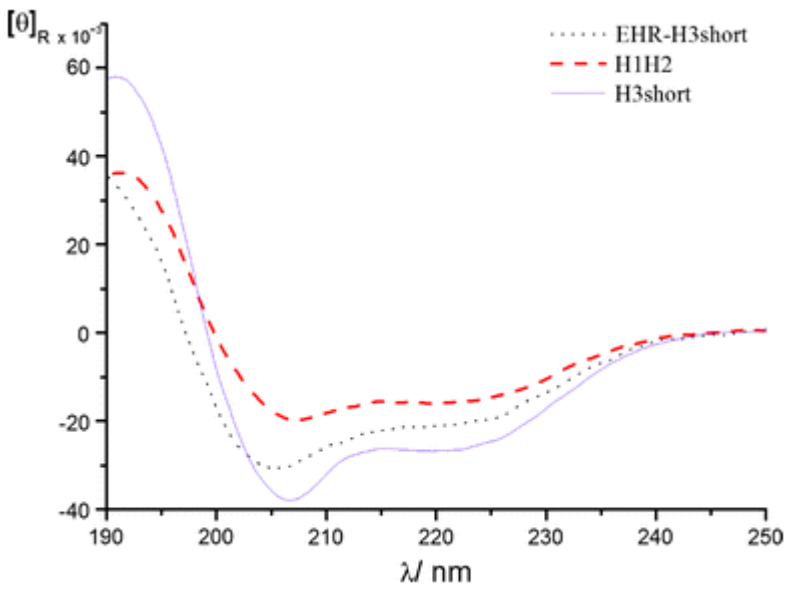


Figure 28: CD spectra in 20% TFE (20µM solution).

As shown in **Figure 24A**, the CD spectrum of the mixture containing EHR-H3short and H3short peptide was not superimposable to the arithmetic sum of the CD spectra of the single components. The same behavior was observed also for the mixture of EHR-H3short and H1H2 (**Figure 29B**), indicating that in both cases an interaction was effective.

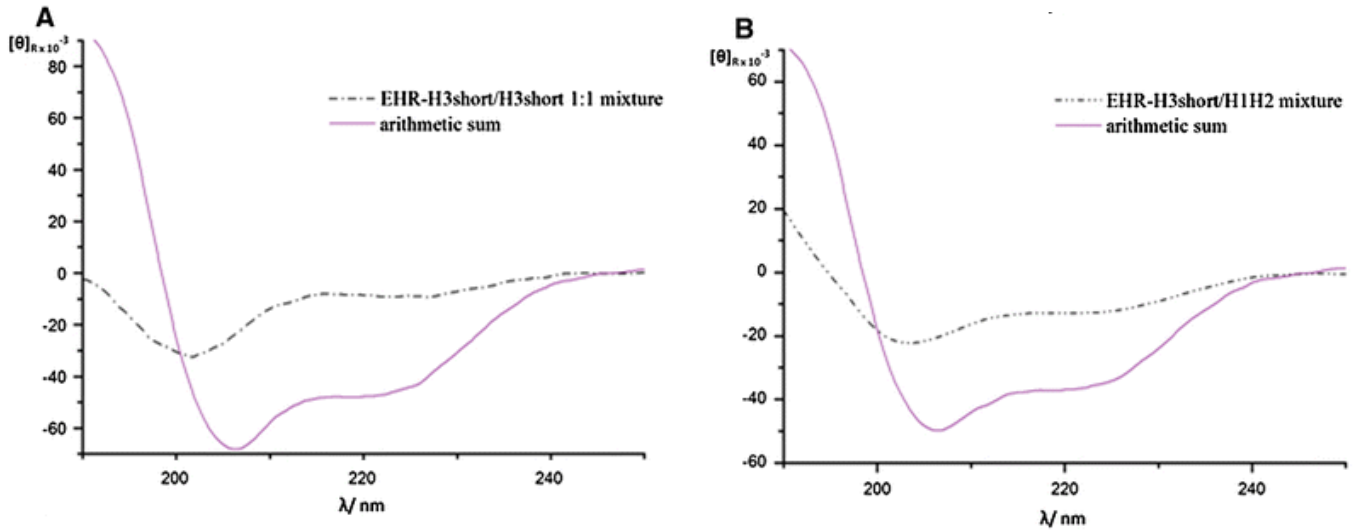


Figure 29: Interaction between EHR-H3short with H3short (**A**) and H1H2 (**B**).

In particular, a blue shift of the $\pi \longrightarrow \pi^*$ band was observed, suggesting that an increase of unordered fraction, probably, occurs for all peptides. From these findings, we postulated that, when single H1H2 or H3short fragments were added to EHR-H3 short, they could act as folding disrupter interfering in the formation of intramolecular interactions within EHR-H3short.

1.5.2.2. DNA target interaction studies

The interaction of MafA DBD with a DNA sequence of 26 bp from human insulin enhancer containing MafA MARE (CGGAAATTGCAGCCTCAGCCCCCAGC)[9] was studied. First, the CD spectrum of the DNA duplex was recorded in phosphate buffer (0.1M, pH7.5; **Figure 30**), showing a B-DNA conformation characterized by the presence of a positive peak at 270 nm and two negative ones, at 250 and 210 nm, respectively.[10] The CD spectrum of 75mer DBD alone was compared with the spectrum derived from a mixture with human insulin MARE (DBD/DNA, 2:1) by subtracting DNA contribution: an increase of α-helical content was observed (62% helix content found), indicating that an interaction was effective (**Figure 30**).

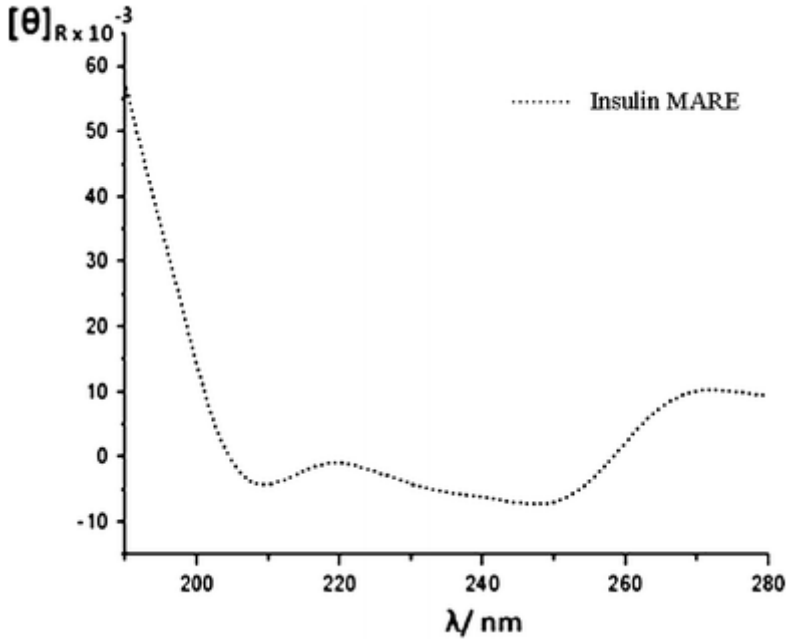


Figure 30: CD spectrum of 26bp DNA sequence from human insulin enhancer.

In order to investigate if the binding to DNA was

Sequence specific, we also studied the interaction of DBD with a scrambled DNA sequence (GATCAGTACCACGCGCTC CACGCCGA). As expected, we obtained no specific con- formational changes in the CD spectrum of DBD (**Figure 31**).

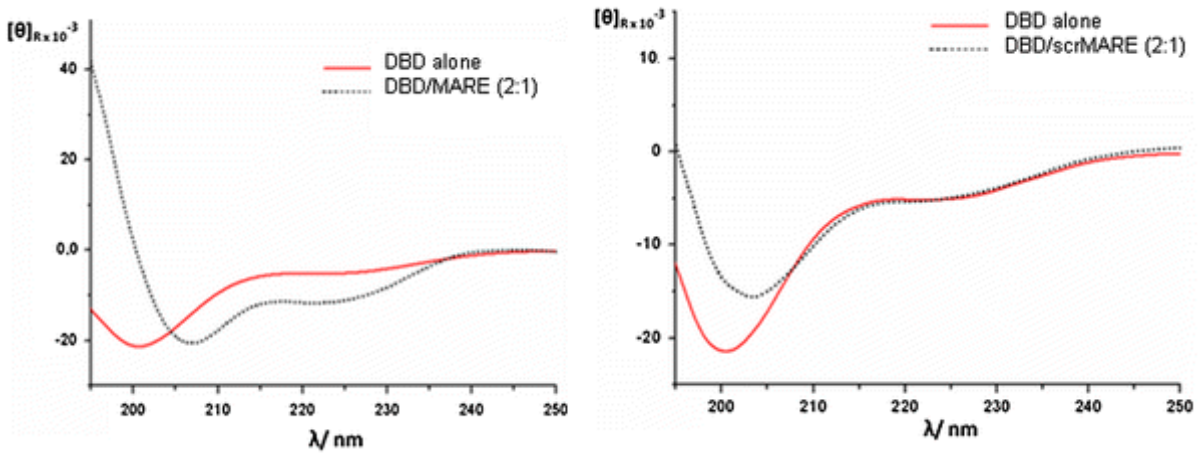


Figure 31: Comparison between DBD spectrum alone and then spectrum derived from the 2:1 mixture with insulin MARE (on the left) and scrambled DNA (on the right).

In order to gain an insight in the binding of MafA to insulin MARE which, as reported above, is uncommon with respect to standard ones,[9] CD studies were also performed using shorter sequences from DBD. In the presence of DNA, a structural change (20% helix structure in CD spectrum) was only observed for the peptide corresponding to EHR–BR sequence (227–278) (**Figure 32**).

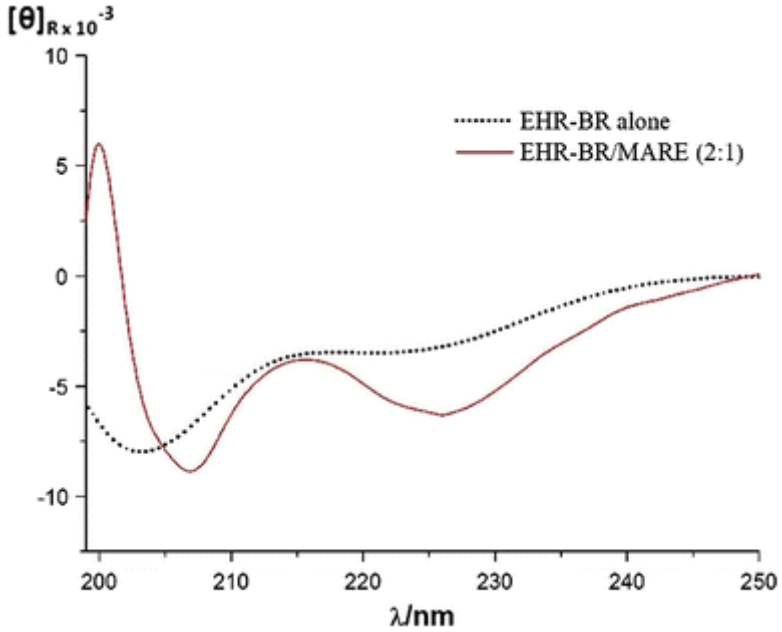


Figure 32: Comparison between EHR-BR spectrum alone and the one derived from the 2:1 mixture with insulin MARE.

On the contrary, we did not observe any interaction with either the EHR-H3short fragment (227–263) or, interestingly, the entire BR (248–275) and the entire H3 (248–301). From these results, we can postulate that under physiological conditions and in the absence of specific DNA, MafA is in a dynamic equilibrium between folded and unfolded states. In particular, important β sheet and unordered contributions were found for 75mer DBD and its fragments in an hydrophilic environment, while an hydrophobic medium stabilized the α helix structure. This conformation is also dependent on two domains of the DBD: the LZ and EHR. This last region is characteristic and highly conserved within the Maf family. This region is rich in hydrophobic residues forming a hydrophobic cluster which is surrounded by hydrophilic amino acids.[7] Our CD results demonstrated that, in the absence of DNA, these intramolecular hydrophobic interactions are effective in partially stabilizing the conformation of DBD. H1H2 (227–243) and H3short (248–263) fragments, in particular, could act as folding disrupters (**Figure 29**). The interaction with insulin MARE induces conformational changes by stabilizing the three-dimensional conformation of MafA domain in a similar manner as reported for other family members. Early CD studies on c-Maf and palindromic consensus MARE reported DNA sequence-dependent folding. In particular, the flanking region of MARE was found to be dramatically important for the interaction with DBD, thus stabilizing the a helix structure, while the core seemed to have a scarcer relevance.[10] The binding of MafA to insulin enhancer then represents a case in this paradigm, the MARE recognition sequence being not palindromic, and possessing degenerated core and flanking regions.[9] Our CD results, anyway, indicate that also in the case of a degenerated MARE, the interaction with DNA induces an increase of the helical content in MafA DBD (62% helical content found; **Figure 31**). Further conclusions can be drawn starting from the interaction studies between MafA shorter fragments and insulin MARE. From the crystal structure of the MafG/MARE25 complex (PDB ID 3A5T),[7] it was evicted that conserved residues from BR are involved in DNA binding. In particular, Arg57 and Asn61 (corresponding to Arg259 and Asn263 in MafA) were shown to directly interact with DNA, while residues Arg62, Tyr64, Ala65, Arg69, and Lys71 (corresponding to Arg264, Tyr266, Ala267, Arg272, and Lys274) are fundamental for the correct spatial orientation of Arg57 and Asn61 side chains.[7] Our CD studies on BR, anyway, showed that both peptides H3short (248–263) and EHR-H3short (227–263), containing the key residues Arg 259 and Asn 263, were found unable to interact with DNA, as well as the entire BR (248–275). Curiously, also the H3 peptide (248–301, corresponding to BR and LZ regions) was also unable to bind insulin MARE, whereas the EHR–BR (227–278) peptide interacted (20% helical content found; **Figure 32**). From these data, we can assume that the highly conserved EHR is not only important for DBD folding but also for binding to insulin

enhancer. Our hypothesis is that, since from the crystal data on MafG/MARE25 this part of the DBD seemed not to be directly involved in DNA interaction, its role probably lies in a correct folding inducer. In this context, the hydrophobic cluster within the EHR (in particular Leu 231 and Met 234) could then act as a local structural element that triggers the folding of the entire DBD in the presence of DNA. Considering that, from CD data, the LZ region seems not to be necessary for DNA binding, the dimerization of the DBD is an event that, probably, occurs after DNA recognition. Indeed, we can postulate that MafA undergoes a disorder-to-order transition through a coupled binding and folding mechanism similarly to that reported for other b-LZ domains.[11]

1.5.3. References:

1) Friligou, I., Papadimitriou, E., Gatos, D., Matsuokas, J., Tselios, T., *Amino Acids* **2011**, 40, 1431-1440.
2) Pellerino, S., Annoni, C., Contini, A., Clerici, F., Gelmi, M. L., *Amino Acids* **2012**, 43, 1995-2003.
3) Greenfield, N. J., *Nat. Protoc.* **2006**, 1, 2876-2890.
4) Garbett, N. C., Ragazzon, P. A., Chaires, J. B., *Nat. Protoc.* **2007**, 2, 3166-3172.
5) Provencher, S. W., Glockner, J., *Biochemistry* **1981**, 20, 33-37.
6) Whitmore, L., Fallace, B. A., *Biopolymers* **2008**, 89, 392-400.
7) Kurokawa, H., Motohashi, H., Sueno, S., Rimura, M., Takagawa, H., Kanno, Y., Yamamoto, M., Tanaka, T., *Mol. Cel. Bio.* **2009**, 29, 6232-6244.
8) Chen, Y. H., Yang, J. T., Chau, K. H., *Biochemistry* **1974**, 13, 3350-3359.
9) Matsuoka, T., Artner, I., Henderson, E., Means, A., Sander, M., Stein, R., *PNAS* **2004**, 101, 2930-2933.
10) Dlakic, M., Grinberg, A. V., Leonard, D. A., Kerppola, T. K., *EMBO J.* **2001**, 20, 828-840.
11) Seldeen, K. L., McDonald, C. B., Deegan, B. J., Farooq, A., *Arch. Biochem. Biophys.* **2008**, 473, 448-60.

1.6. Creation of a c-Maf DBD homology model

1.6.1. Theory

1.6.1.1. Molecular dynamics

Although the advantages of the availability of tridimensional structures of receptor-ligand complexes are evident for projecting new ligands, the crystallized structure represents a static image of a complex dynamic process. Molecular dynamics (MD) is a simulation computational method that constitutes an important instrument both for studying the temporal evolution of proteins and other biological macromolecules and for obtaining kinetics and thermodynamic information.[1,2] Molecular dynamics is based on Newton's second law:

$$\mathbf{F} = m\mathbf{a} \tag{1.1}$$

Where **F** is the net force acting on the object, *m* is the mass of the object and **a** is the acceleration of the object. Knowing the force exercised on every single atom of a complex system it is possible to determine the respective accelerations. For a complex system, the movement of every single atom is determined by the equation:

$$\mathbf{F}_1 = m_i\mathbf{a}_i, \quad m_i = \frac{d^2 ri}{dt^2} \quad \Longrightarrow \quad i = 1,2\ldots, N \tag{1.2}$$

The total force **F**i, exercised on each atom i can be expressed as a negative gradient of a potential energy:

$$\mathbf{F}i = -\frac{dV}{dr_i} \tag{1.3}$$

That combined with the previous one gives:

$$-\frac{dV}{dr_i} = m_i\frac{d^2 ri}{dt^2} \tag{1.4}$$

Knowing the initial coordinates (obtained from experimental data or theoretical studies taken from X rays crystallography, NMR of homology modeling), the initial speed attributed basing on the strating temperature and the forces applied on the system, it is possible to calculate the new coordinates and the atoms speed ata determine time (trajectory). The aim of this simulation is thus solving the Newton motion equation for small intervals of time (*time steps*) using different algorithms (i.e. Verlet algorithm) and obtaining a trajectory that corresponds to the position and speed in time of each atom involved in the system. From a molecular dynamics trajectory it is possible to extrapolate macroscopic properties calculating the average of the values that these properties acquire during the simulation. The choice of the suitable energetic function describing the intermolecular and intramolecular interactions is critical for the molecular dynamics simulation success. The potential energy **E**, as a function of the position of all the atoms of a system, is expressed by the sum of two terms:

$$\mathbf{E}\{\mathbf{r}\} = E_{\text{bonded}} + E_{\text{non-bonded}} \qquad (1.5)$$

Where E_{bonded} comprehends the interactions between covalently linked atoms which depends from the bound length, angles and rotations of the molecules bounds; $E_{\text{non-bonded}}$ is referred to the interactions developed between non chemically bonded atoms or atoms separated by three or more covalent bounds. The use of the simple nuclear coordinates as a representation of the atoms ignoring the electrons is justified by Born-Oppenheimer approximation.[3] The energetic function usually consists of a consisting number of parameterized terms. These parameters are extrapolated from empiric data obtained by model compounds with several experimental techniques (X rays, infrared spectroscopy, Raman spectroscopy, microwave spectroscopy) and/or quanto-mechanical studies on a small group of molecules or fragments assuming that these parameters could fit also a much larger group of molecules. The totality of the potential energetic functions and of the parameters consists in the so-called force field. Almost all force fields are characterized by a similar mathematic expression, however they are distinguished by the different values of the parameters associated with different energetic terms that vary according to the molecular family considered. One of the most used force field is AMBER,[4] which an analytic form is:

$$U = \sum_{bonds} K_r \left((r - r_0)\right)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\Phi \left[1 + cos(n\Phi + \Phi_0)\right] + \sum_{atom\ i} \sum_{j \neq i} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon_0 r_{i,j}} \quad (1.6)$$

The force field terms are briefly illustrated as follows:

1. *Stretching energy*

$$E_{\text{bond-stretch}} = \sum K_r (r - r_0)^2$$

In this expression, based on Hook law, the bond energy is written as a function of the divergence of the r bond length from the ideal r_0. K_r constant represents the bond force. Both r_0 and K_r are specific for each couple of atoms bonded , thus they depend on the type of atoms involved in the bond.

2. *Bending energy*

$$E_{\text{bond-bend}} = \sum K_\theta (\theta - \theta_0)^2$$

It is associated with the deviation of the bond angles from the ideal value θ_0, based on the Hook law. Also in this case K_θ and θ_0 depend on the atoms involved in the bond.

3. *Torsion energy*

$$E_{\text{rotate-along-bond}} = \sum K_\varphi \left[1 + cos(n\varphi + \varphi_0)\right]$$

It represents the potential energy deriving from the position of the torsion angle, it shows the same tendency of a period function, in which K_φ is the torsion energy (Kcal/mol), *n* is the torsion periodicity, φ and φ_0 respectively are the torsion angle and the reference torsion angle (usually 0 or 180º). These elements represent the interactions involving covalently bonded atoms, however, commonly the force field takes in consideration also the interactions between atoms not directly bonded, or rather Van der Waals forces and electrostatic forces. There are several force fields according to the family of molecules base on which they are parameterized. The most popular all-atom force fields usedn for biological macromolecules are AMBER, CHARMM, OPLS.[5] A noteworthy variety of experimental conditions could be set up during the first molecular dynamics simulation. In the past, the first simulation were carried out considering the molecules as separated entities in the void, nowadays the inclusion of water is and obliged choice in order to make the simulation the more realistic possible. One of the most important solvent effect is the ability to act as a shield against electrostatic interactions. The solvent inclusion into molecular dynamics could occur in an explicit or implicit way. The first method, that is also the simplest one, consists in including a dielectric constant ε in the electrostatic term of the potential energy, using, if necessary, an effective dielectric constant subjected to the distance from the solute. Common models of implicit solvent are: Poisson-Boltzmann (PB); Generalized Born (GB), Generalized Born Surface Area (GBSA). The principal advantage of these models is a reduction of the computational calculation time against the impossibility of identifying water molecules "functionally" relevant in receptor-ligand complexes. The explicit method consists in the explicit use of solvent molecules in the solvent energy calculation; those water molecules are treated with a specific force field, such as TIP3P, TIP4P, TIP5P. The explicit use of solvent molecules despise of implying the use of more power and calculating time, nowadays results to be the preferential approach. In the case of explicit solvent model, the so-called *period boundary conditions* are frequently chosen in order to prevent water molecule diffusion in areas too far from the solute and to guarantee that the simulation proceeds using the minimum number of solvent molecules. The solute is generally neutralized adding suitable ions and introduced in a box of solvent molecules of different shapes (cubic or poliedric). The molecular dynamics simulations in explicit solvent mode are usually carried out in "thermostatic" conditions, in which a system keeps the temperature stable. Besides isothermal conditions, a simulation can be also conducted at constant volume (NVT) or isobaric conditions (NPT). The molecular dynamics are carried out on a time scale that goes from few nanoseconds to microseconds. The principal limitations regarding the molecular dynamics methods regard the approximations introduced to handle the massive amount of data and calculations necessary for

generating the trajectory of a system composed of several thousands of atoms. The first approximation is introduced by the exclusive use of classical mechanic: in a macroscopic system characterized by atoms belonging to the fist three groups and standard temperatures the margin of error is risible, however, classical mechanic is not able to describe events caused by the migration of electrons from one atom to another one (such as the breaking/formation of a covalent bond) for which a quanto-mechanical component should be included. Moreover, the force fields are approximated since they don't consider all the typology of atoms interactions but just the principal ones. For what concerns the Lennard-Jones interactions and the columbian interactions, the introduced approximation consists in not considering the interaction between one atom and all the rest of them but just the interactions between atoms that are at close distance (within the *cut-off* ray). Before proceeding with a molecular dynamic simulation, usually a preliminary phase with the aim of minimizing the energy and getting closer to a local energetic minimum is performed; this step is usually carried out by the same software used for the dynamics and it is based on two complementary algorithms: *steepest descent* and *conjugate gradient*, the first one leads rapidly in proximity to the closest minimum while the second one get more closer to it but it cannot work efficiently if the starting point is far from the minimum selected.

1.6.1.2. Molecular Mechanics Poison Boltzmann Surface Area (MM-PB/GBSA)

The aim of conducting a molecular dynamics simulation is to obtain kinetic or thermodynamic data about the system used as a model of study. Many thermodynamic properties could be extrapolated by a sufficient number sampling pool of the system conformations, for instance the bond free energy of a protein-protein complex or a receptor-ligand complex. The bond free energy represents a measure of the complex stability, as a matter of fact only if it shows a negative value, the complex formation will be possible.[1] Protein-protein interactions are critical for many biological processes, such as in the signal transduction pathways that regulates many aspects of cellular life and constitute an important potential therapeutic target. Several approaches presenting different characteristics in terms of accuracy and computational cost, are available which allow to estimate the bond free energy. Methods like the *Free Energy Perturbation* (FEP) or *Thermodynamic Integration* (TI) are computationally costly, however they are applied successfully to estimate the bond energy of several complexes.[6,7] Many method were successively developed for estimating the free energy more rapidly, among which there are the *Linear Interaction Energy* (LIE),[8] the *Molecular Mechanics Poisson Boltzmann Surface Area* (MM/PBSA)[9,10] method, the *Chemical MonteCarlo Molecular Dynamics* (CMC/MD)[11] method and the λ-*Dynamics* method.[12] MM-PBSA[13,14] consists in a method for both evaluating the bond free energy and for calculating the absolute free energy of molecules in a solution, analyzed one the molecular dynamics simulation is completed.[4] Generally, from the molecular dynamic trajectory some *snapshots* are extracted for calculating variation of the free energy between two states, typically, the receptor-ligand complex state or the protein-protein state and the state in which the two systems are isolated. Only the initial and the final state of the system are evaluated for giving an estimation of the free energy change.[1] The aim of this method is calculating the bond energy of a complex as shown in **Figure 33**.

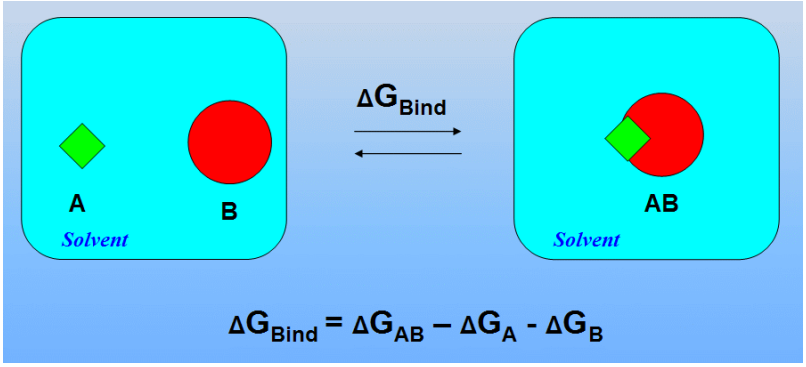


Figure 33: Formation of the AB complex in solution

The bond free energy of the complex, ΔG_{Bind} is calculated basing on the following equation:

$$\bar{G}=\bar{E}_{MM} + \bar{G}_{PBSA} - TS_{MM} \quad (1.7)$$

Where G is the average free energy of the complex of interest, G_{PBSA} in the solvation free energy, $\bar{E}_{MM}$ is the average molecular mechanic energy and $- TS_{MM}$ is the solute entropy. The average molecular mechanic energy is typically defined by:

$$\bar{E}_{MM} = \bar{E}_{bond} + \bar{E}_{angle} + \bar{E}_{tors} + \bar{E}_{vdw} + \bar{E}_{ellec} \quad (1.8)$$

Where E_{bond}, E_{angle}, $E_{torsion}$, E_{vdW} and E_{ellec} correspond to bonds, angles, torsions, Van der Waals interactions and electrostatic interactions of the force field. A *Cut-off* that tends to the infinite is used for all the interactions. The salvation free energy could be fatherly parted into an electrostatic and a hydrophobic component:

$$G_{PBSA}= G_{PB} + \gamma A \quad (1.9)$$

G_{PBSA} is the average solvation electrostatic component and it is calculated solving the Poisson-Boltzmann (PB) equation, γ is the solvent superficial tension while A represents the solvent-accessible surface area. The average molecular mechanics energy ($\bar{E}_{MM}$) is calculated using SANDER, the program that constitutes the core of the suite software AMBER, according to the force field selected. The electrostatic component could be evaluated by an alternative approach that consists of using an approximation of the PB equation, that is the Generalized Born (GB). In other words, the two methods, MM/PBSA and the MM/GBSA are distinguished by the method used for calculating the salvation free energy electrostatic component while both of them determine the hydrophobic component evaluating the solvent accessibility to the solute. For a complete evaluation of the free energy of the complex it is possible also considering an entropic component that could be obtained, for instance, analyzing the normal modes. This component is much smaller compared to the average molecular mechanic energy ($\bar{E}_{MM}$) or to the salvation free energy (G_{PBSA}) therefore frequently it is not considered in the free energy estimation. MM-PB(GB)SA method has the considerably advantage of being a retrospective method of analysis of the molecular dynamics trajectory and it results less expensive in term of calculation time-consuming.

1.6.1.3. Computational alanine scanning

Within the molecular modeling techniques, frequently used in protein-protein interaction studies, there is a method called *Computational Alanine Scanning* (CAS), which evaluates, in a protein-protein complex, the impact of an alanine substitution in the interfacial residues in terms of free energy.[15] The alanine is used because it presents a small and inert methyl group in its side chain. The methyl group, once substituted to one or more amino acidic residues, permits to evaluate the functional contribute of the side chain of the substituted residue and, lastly, which residues provide strength and specificity of the bond. In the CAS method, as in the molecular mechanics/Poisson Boltzmann (Generalized Born)[16] surface area method, a retrospective approach based on *snapshots* extrapolated from a molecular dynamics trajectory, is used for calculating the free energy. The molecular dynamics can be carried out both in explicit[17] and implicit solvent.[18] The free energy variation between the wild type complex and the mutated one is defined as:

$$\Delta G_{binding} = \Delta G_{binding\_mutant} - \Delta G_{binding\_wildtype} \quad (2)$$

Keeping in mind that the bond free energy between two molecules id the difference between the free energy of the complex and the one of the singular monomers (receptor and ligand). The computational alternative to the *alanine scanning mutagenesis* has the consistent advantage of being much faster and less laborious, but, above all, it is able to reproduce experimental data of mutagenesis.[16]

1.7.Design of c-Maf LZ modulators

1.7.1. Preparation of a c-Maf/T-MARE model

Since nowadays structural X rays or NMR studies on c-Maf are not available in literature, the first part of this work focused on the analyses and description of the crystallographic complexes of Maf proteins with DNA, available on Protein Data Bank (PDB),[4] that are 3A5T.pdb and 2WTY.pbd. The first one comprehends the MafG homodimer complexed with MARE25; the second one, the MafB homodimer creating a complex with a DNA sequence containing T-MARE site. Successively, using homology model strategies a template for c-Maf/T-MARE was obtained applying MafB as a template.

1.7.2. Analysis of the crystallographic complexes 3A5T.pdb and 2WTY.pdb

As previously described, recently the MafG crystal structure complexed with a double strand DNA sequence called MARE25 (15pb) was obtained (GTGCTGACTCATCAG).[19] The MRE25 sequence includes the NF-E2 binding site, that binds the Maf homodimer.[20] In MARE25 sequence was introduce a singular nucleotidic variation in comparison with the consensus sequence T-MARE (GCTGACTCAGCA) in order to evaluate the behavior of each monomer recognizing the DNA target. The crystallographic complex obtained by X ray diffraction, which includes the MafG homodimer bonded to MARE25, shows a resolution of 2.80Å and comprehends:

1- Chain A: residues 19-111 (subunit A MafG);
2- Chain B: residues 20-112 (subunit B MafG);
3- Chain C: 5'-CTGATGAGTCAGCAC-3';
4- Chain D: 5'-GTGCTGACTCATCAG-3';
5- Crystallization water;
6- Mg^{2+} ion.

1.7.3. Homology modeling and analyses of the complex c-Maf/T-MARE

The homology model is a method that allows the construction of a computational model of a protein of interest with atomic resolution starting from its amino acidic sequence and available tridimentional structures obtained by X rays crystallographic techniques or NMR analyses (templates). The homology modeling could be a very useful tool for generating structural models, formulating hypothesis concerning the protein functions and carrying on preliminary studies of *virtual screening* to individuate molecules that could interfere with the protein functions. The model obtained will be as much accurate as higher is the sequence identity between the protein in exam and the template, thus the choice of the most suitable template results determinant. Lacking of a crystallographic data on c-Maf we choose to construct a homology model. We used, as a template, MafB, another large maf protein. The homology levels within the Maf family members are elevated (70-80%) especially in the LZ region. For the model preparation SWISS MODEL[21] server was adoperated, which permits to generate homology models automatically, starting from the amino acidic sequence of human c-Maf. Once obtained the homodimer structure it was decided to introduce the T.MARE sequence derived from the crystallographic complex 2WTY.pdb. The homodimer complex of c-Maf/T-MARE is shown in **Figure 34**.

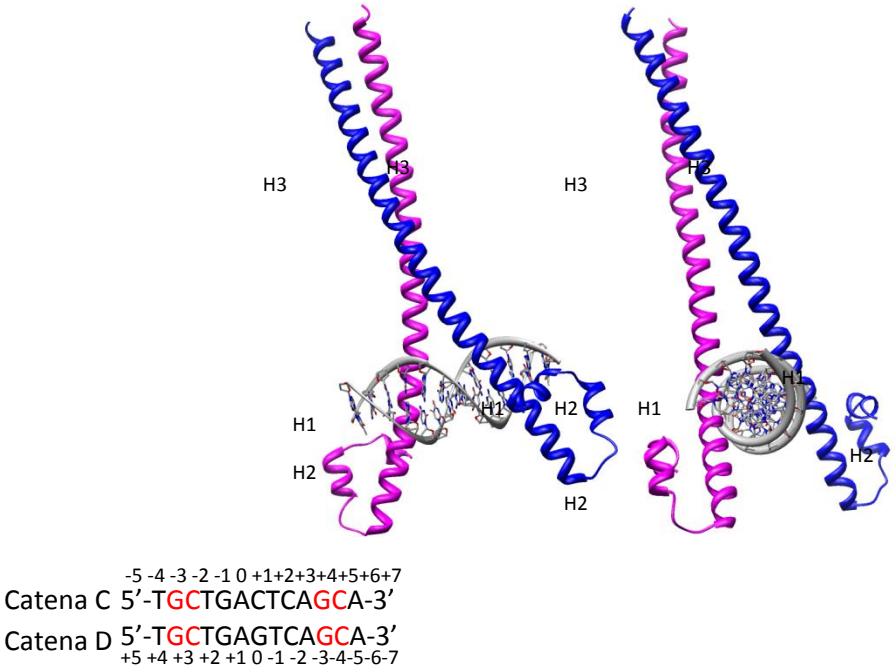


Figure 34: c-Maf homodimer complexed with T-MARE. The subunit A (purple) contacts the GC flanking region of the D chain, while the subunit B (in blue) binds the GC flanking region on chain C.

The ancillary region is stabilized by the following hydrophobic interactions: Leu35, Leu28, Val25 of helix H3, Val20 from the connecting region between H2 and He helices, Leu17, Leu13 and Val10 od helix H2, Leu5 and Val6 of H1 helix (**Figure 35**).

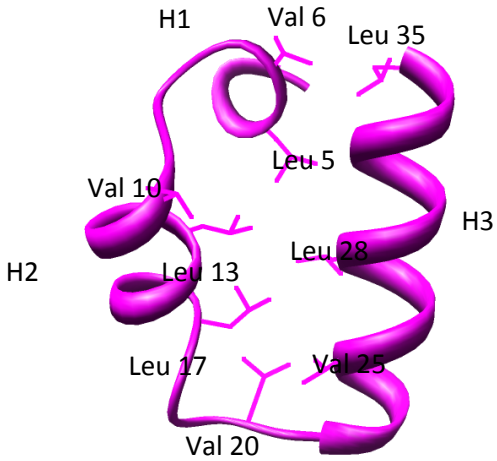


Figure 35: Hydrophobic cluster in the HER region of subunit A.

As in the case of MafG, the charged residues are exposed on the surface (**Figure 36**):

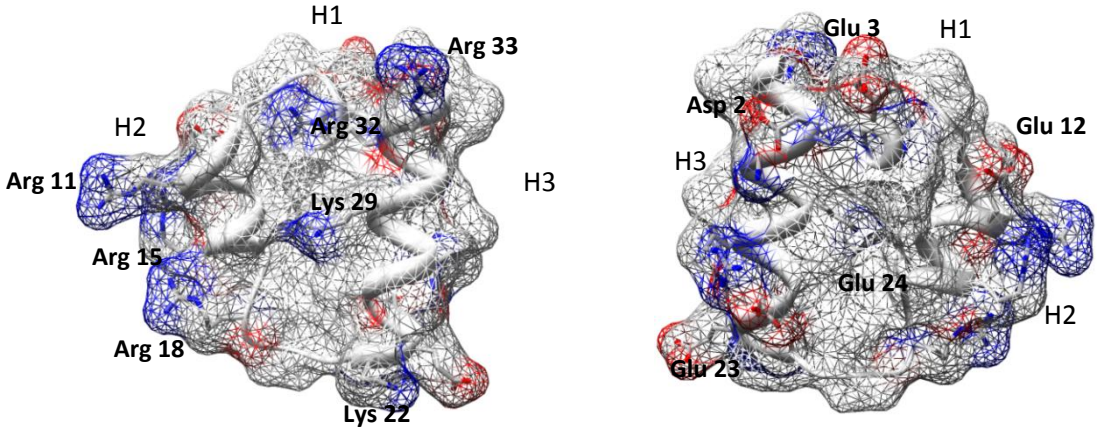


Figure 36: Charged residues exposed on the surface of EHR region.

In the EHR region we could identify:

Hydrophobic amino acid residues: 11/35 ⟶ 31%

Polar amino acidic residues: 10/35 ⟶ 29%

Negatively charged amino acidic residues: 5/35 ⟶ 14%

Positively charged amino acidic residues: 9/35 ⟶ 26%

The contacts between monomers in the leucin zipper region of c-Maf are mediated both by hydrophobic residues (Leu55, Leu62, Leu63, Val66, Leu69, Ile73, Val77, Leu90) and electrostatic interactions (Gln51 with Arg52, Glu58 with Lys59, Glu79 witharg80) (**Figure 37**).

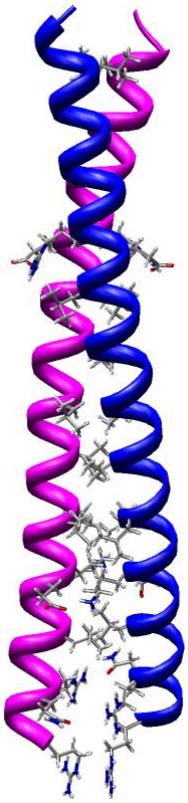


Figure 37: Interactions in the LZ region of c-Maf. Subunit A is in purple, subunit B in blue.

The interaction between the c-Maf homodimer and the T-MARE can be reassumed as follows:

Subunit A: Arg33 forms two hydrogen bonds with the carbonylic oxygen of chain D guanine in position +4. Thr34 creates a Van der Waals interaction with the sugar moiety of the thymine in position +2 on chain C and an hydrogen bridge with the phosphate group of the same base. Lys36

forms a Van der Waals interaction with the sugar moiety of thymine in position +5 of chain D. Asn37 is connected to thymine +2 on chain D through a Van der Waals interaction and it forms an hydrogen bond in the NH moiety of the cytosine +3 on chain D. Arg38 constructs an hydrogen bond with the phosphate group of thymine +2 on chain C. Tyr40 creates a van der Waals contact with the thymine in position + 5 and with the thymine in position +2 on chain D, an hydrogen bond with the phosphate group of the guanine in position +4 on chain D. Lys47 and the thymine residues in position +2 on chain D are connected through an hydrogen bond with the phosphate group on the base.

Subunit B: Thr34 forms a vdW interaction with the sugar moiety of the thymine -2 on chain D and an hydrogen bridge with the phosphate group on the same base. Lys36 creates a vdW interaction with the thymine in position -5 and an hydrogen bond with NH_2 of the cytosine in position -2 on chain C. Arg38 is connected to the phosphate group on thymine -2 on chain D through an hydrogen bond. Tyr40 forms a vdW interaction with the sugar moiety of the thymine in position -5 and an hydrogen bond with the phosphate group of the guanine in position -4 on chain C. Arg45 creates an hydrogen bridge with the phosphate group of adenine in position 0 of chain D. Also in this case, Asn37 and Arg33 (equivalent to Asn61 and Arg57 in MafG and Asn248 and Arg244 in MafB) establish interactions principally with the flanking regions.

1.7.4. Molecular dynamic simulation

1.7.4.1. Definition of a MD protocol

A MD protocol is generally subdivided into four steps:

1. Energetic minimization;
2. System equilibration;
3. Heating;
4. Production run.

In this case, the same MD protocol was carried out using two different force field (ff03.rl and ff99SB) both on the homodimer MafG in the absence and in the presence of MARE25 and on c-Maf homodimer in the absence and in the presence of T-MARE, with the aim of extrapolating information from the trajectory analyses about the behavior of small and large Mafs. The crystallographic structures derived from PDB data bank or obtained by homology modeling are not immediately suitable for a computational analyses. Frequently they don't posses information about the position of hydrogen atoms and they lack of precise data regarding the sequence and nature of the atoms bonds, therefore it is necessary to set information about the protonation state of residues and their tautomery. The starting sequences of MafG and c-Maf were protonated using H++ server,[22] calculating the electrostatic component through Poisson Boltzmann algorithm in ionic conditions of 0.15M pH 7.40 trying to reproduce the physiological conditions. The solvent molecules and Mg^{2+} ions were removed from MafG homodimer and both the C and the N- termini of each residue were protected using, respectively, ACE (acetamide) and NME (N-methyl). Once ultimate this preliminary steps, both the accurate atom type assignation and the correct protonation of the positively or negatively charged residues were visually evaluated. The following MD protocol was applied to both the models:

8. Minimization of hydrogens, water and ions: 5000steps of which the first 1000 carried out with *steepest descent* algorithm and the subsequently 4000 steps with *conjugate gradient* algorithm;
9. Volume constant equilibration (NVT) and, successively, pressure constant (NPT); 90ps, from 0m to 300K, with an harmonic restraint of 5Kcal/mol on the backbone;
10. Side chains, water and ions minimization: 5000steps of which the first 2500 carried out with *steepest descent* algorithm and the remaining 2500steps with *conjugate gradient* algorithm;
11. Gradual heating from 0 to 300K in 6 consecutive steps of 5ps each with a ΔT of 50K;

12. Volume constant equilibration (NVT), 100ps at 300K with an harmonic restraint on the backbone of 5kcal/mol and successively at constant pressure (NPT) in 4 consecutive steps of 100ps at 300K where the harmonic restraint on the backbone is progressively reduced with a Δof 1kcal/mol;
13. *Production run* at constant pressure (NPT) without restraint at 300k: 8 steps of 1000ps each.

The same protocol was performed using two different force fields: ff03.rl and ff99SB. The simulation time reached complexively 9620 ps for each MD. The minimization step is essential for relaxing the system and bringing the starting structure closely to the energetic minimum. The equilibration step that follows the gradual heating phase, allowed the convergence of the temperature, density, pressure and energy parameters, keeping them constant during time. The analyzed trajectory is obtained from the production run step. All the dynamics were conducted considering the covalent bonds comprehensive of the hydrogen atoms blocked using SHAKE[23] algorithm (periodic conditions with a NPT model, 8Å cut-off ray, explicit solvent H_2O TIP3P with a parallelepipeds-shape solvent box). All the charges were neutralized by Na^+ and Cl^- ions. In order to keep stable the temperature at 300K the Berendsen[24] thermostat was used. The program utilized for the energetic minimizations was SANDER belonging to the Amber software.[4]

7.4.1.2. Analysis of the trajectories

The trajectories post-processing step was performed with PTRAJ program, belonging to AMBER11 and with VDM[25] which has been used for the visual analyses of the trajectories and for realizing some pictures reported in this chapter. The visual analyses of MafG and c-Maf trajectories without the respective DNA target sequences, reveals that the leucine zipper region remain firmly coupled during all the production run time, while the DNA binding site, in both cases, shows a bending outside tendency that causes an increasing destructuring process of the α-helices regions EHR. This behavior is more evident using ff99SB force field (**Figure 38** and **39**).

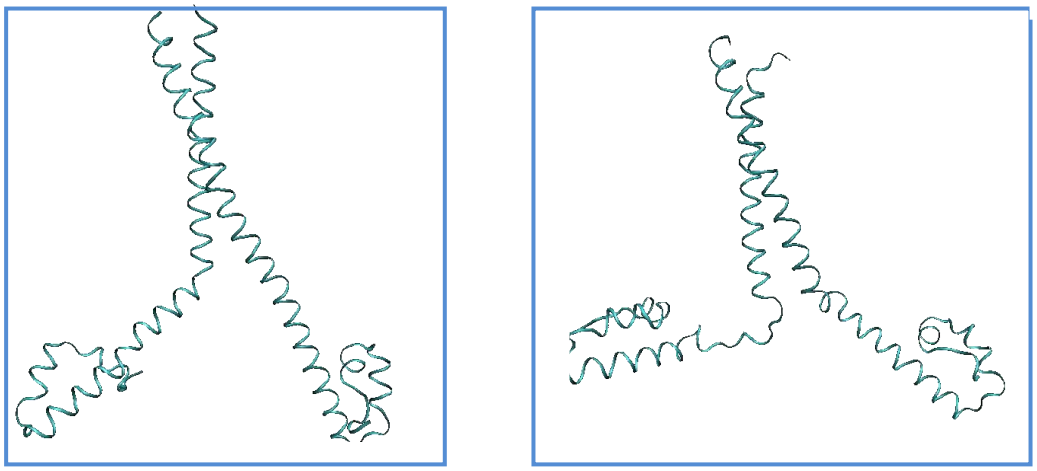


Figure 38: MafG homodimer is the absence of its DNA target MARE25 after the product run step (8ns); respectively left side ff03.r1 force field and right side ff99SB force field.

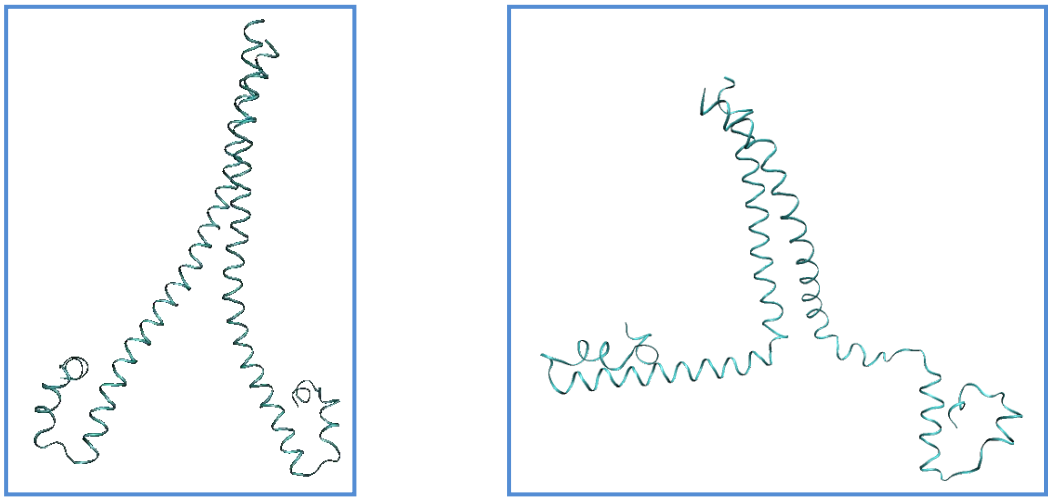


Figure 39: c-Maf homodimer is the absence of its DNA target T-MARE after the product run step (8ns); respectively left side ff03.r1 force field and right side ff99SB force field

The RMSD (Root-Mean-Square-Deviation) tendency during time compared with the initial structure, confirmed the results observed with the visual analyses obtained with ff99SB (**Figure 40**).

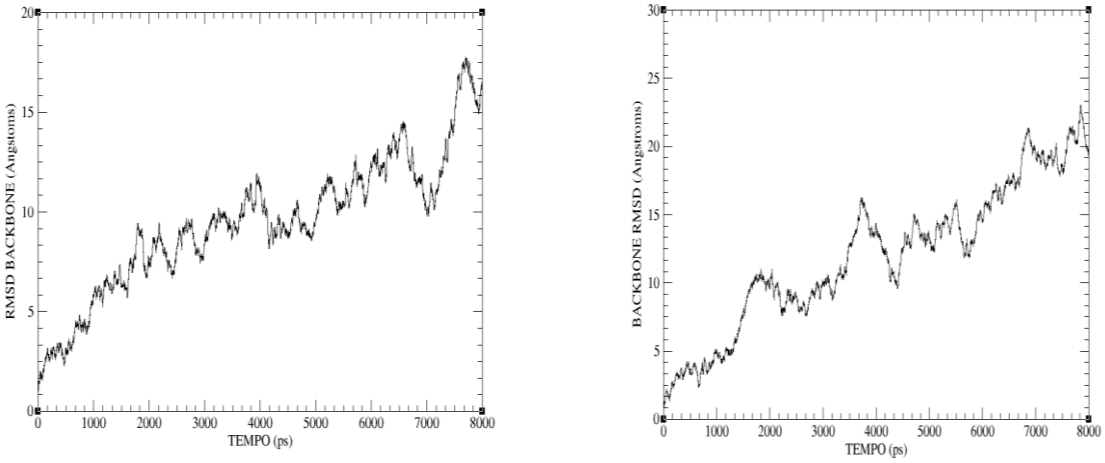


Figure 40: RMSD tendency of the backbone during time (left side c-Maf homodimer; right side MafG homodimer) conducted by ff99SB in the absence of the DNA target.

The backbone fluctuations reach 19Å in the case of c-Maf and 24Å in the case of MafG. At the contrary, the visual analyses of the trajectories of the MafG-MARE25 and c-Maf-T-MARE complexes revealed, using both force fields, an opposite behaviour: the DNA binding site reagion mantains a stable α-helical structure remaining positioned at the DNA major groove locus (**Figure 41** and **42**).

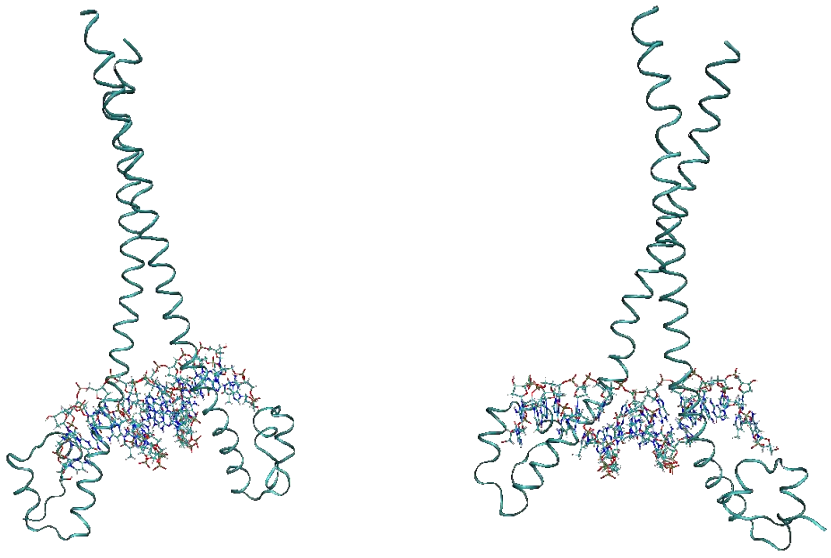


Figure 41: MafG homodimer in the presence of MARE25 DNA sequence at the end of the production run time. (left side ff99Sb force field; right side ff03.r1)

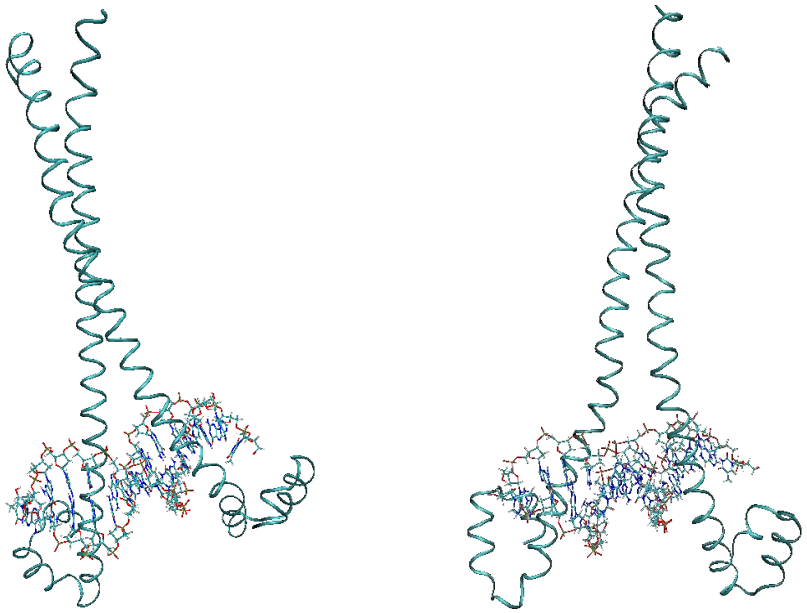


Figure 42: c.Maf homodimer in the presence of T-MARE DNA sequence at the end of the production run time. (left side ff99Sb force field; right side ff03.r1).

The comparison between the RMSD tendency of the complexes in time with the strarting structure (**Figure 43**) showed that the backbone fluctuations are much less significant than in the absence of the DNA target (around 7Å in both cases).

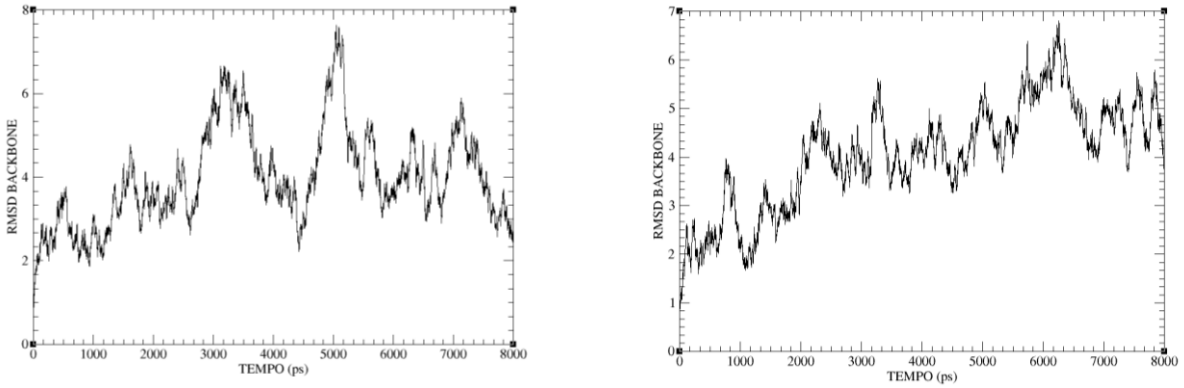


Figure 43: RMSD tendency of backbones in both models studied as a function of the production run time (on the left side c-Maf homodimer; on the right side MafG; simulations carrie out with ff99SB).

As a confirmation of this theory, the distance measured from the C carbonilic of the ACE group to the N-terminal extremity of a c-Maf monomer, passes from 73.27Å in the absence of DNA to 40.13Å in the presence of the DNA target (in the case of MafG, from 87.23Å to 60.31Å). Circular discroism experiments coducted by Dlakić et all., showed that in the presence of the DNA target containing the flanking regions (MARE), the DNA binding domain undergoes a conformational change increasing its α-helical contents. On the contrary, an unfolding tendency could be observed in the absence of the DNA target.[26] Therefore, the results obtained from the simulations showed that two different members of Maf family, a small maf, MafG, and a large Maf, c-Maf, have a similar folding behaviour in the absence/presence of the DNA target sequence. Moreover, these computational results are consistent with the circular dicroism experiments.

1.7.5. Design of peptidic modulators using computational alanine scanning

The free energy for the c-Maf homodimer formation was calculated using MMGBSA method (solvent type igb=5; salts concentration 100nM) on 50 snapshots derived from the homodimer trajectory in the absence of DNA target. The homodimer formation results thermodinamically favoured with a total free energy of -103.09 kcal/mol (st. dv. ±4.79) using ff03.r1 protocol and of -97.06 kcal/mol (st. dv. ± 4.78) with ff99SB force field. The protein-protein interface involves several residues: Leu55, Leu62, Val66, Leu69, Ile73, Val77, Leu90, Gln51 with Arg52, Glu58, Lys59, Glu79 and Arg80 (**Figure 44**).

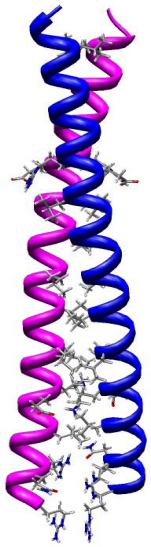


Figure 44: c-Maf homodimer interactions in the LZ region (subunit A in purple; subunit B in blue)

The CAS method allowed the retrospective free energy calculation using snapshots extracted from a molecular dynamics trajectory. The aim is to evaluate the impact in terms of the homodimer bond free energy change, of the mutation of each of the reported residues into alanine. In order to carry out a CAS analyses, each amino acidic residue involved into the binding was manually modified in the PDB file; 50 snapshots where then extracted from each trajectory of the mutated homodimer obtained with both ff03.r1 and ff99SB force field, in the absence of DNA target (salt concentration 100nM). A comparison was made between the wild type complex and the mutated ones. The interfacial amino acids were divided on the base of the variation of the free energy (ΔΔG) provoked by their substitution with an alanine: *hot spots* (ΔΔG > 4 kcal mol^{-1}), *warm spots* (2kcal mol^{-1}< ΔΔG< 4 kcal mol^{-1}) or *cold spots* (ΔΔG< 2kcal mol^{-1}). Comparable results were obtained from both force fields. Hot and warm spots are equally distributed at the homodimer interface (**Figure 45** and **46**).

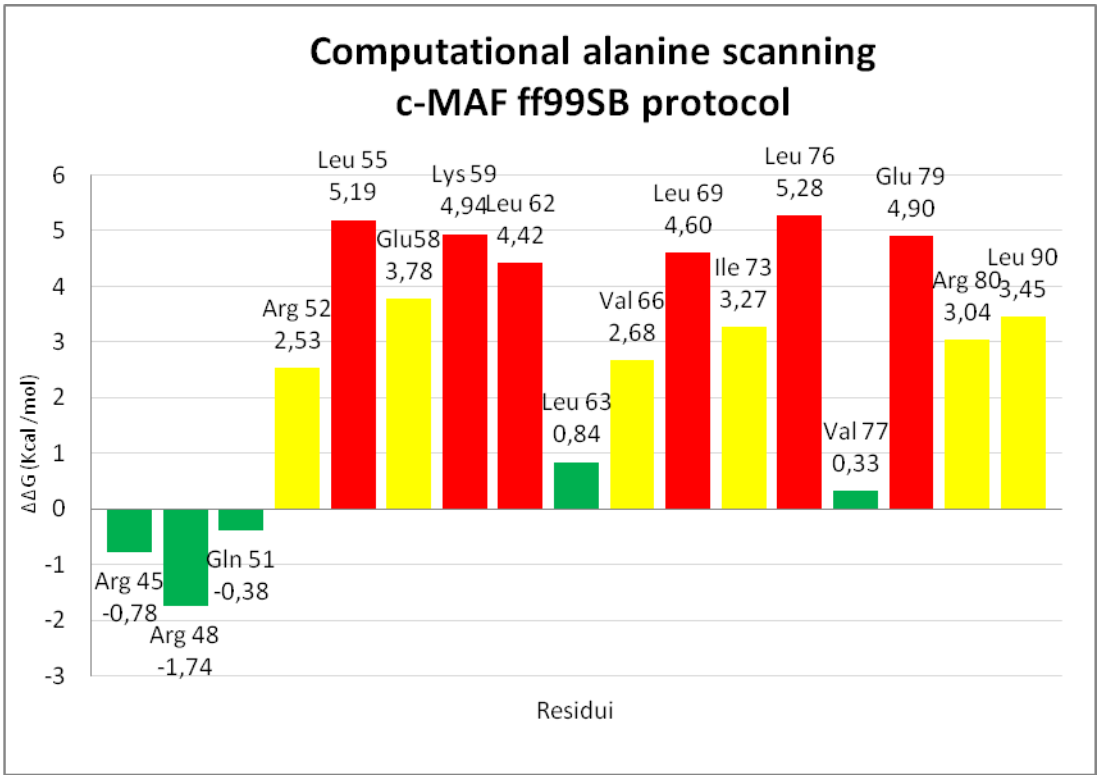


Figure 45: Alanine scanning at the c-Maf homodimer interface using ff99SB protocol

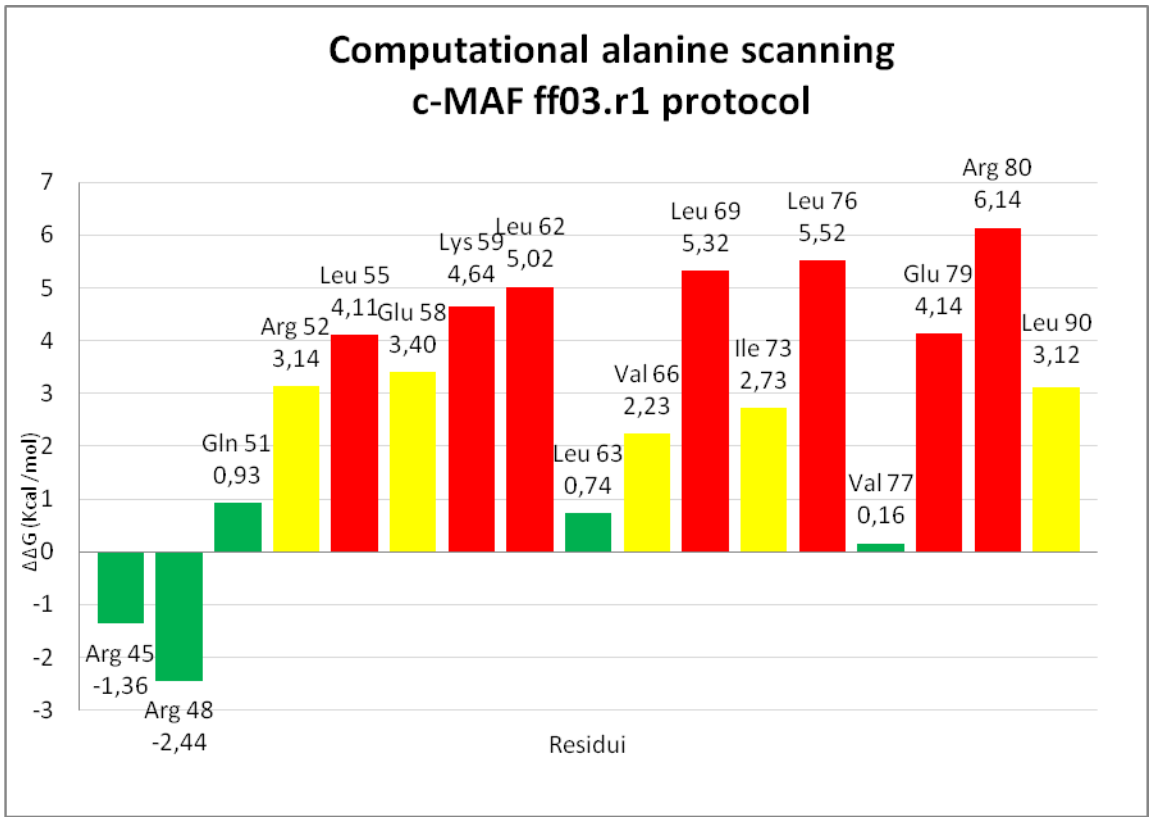


Figure 46: Alanine scanning at the c-Maf homodimer interface using ff03.r1 protocol.

The rational besides the peptidic modulator design is that of using c-Maf wild type as a template and substituting the residues, identified as cold spot from CAS analyses, in order to increase the hydrophobic and or electrostatic interactions. On the other side, the hot spots residues will not be changes. Two classes of peptides were outlined: the first one that binds the N-terminal region of the LZ and the second one that shows affinity for the C-terminal region in order to inhibit the homodimer formation interfering with the dimerization process. The starting point was the study of the c-Maf homodimer interfacial interactions localized in the LZ region closest to the DNA binding site (**Figure 47**).

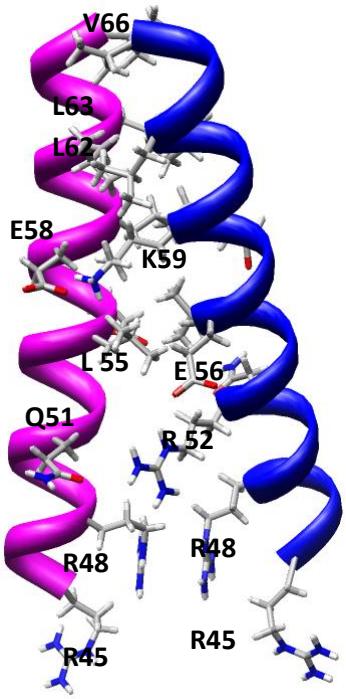


Figure 47: Detailed representation of the c-Maf homodimer interfacial interactions in the LZ region closer to DBD.

Taking c-Maf as a reference, it was decide to substitute Arg45 with a Glu, Arg48 with an acidic residue, such as Asp or Glu and Gln51 with an Asp or Glu in the aim of instauring additional electrostatic interactions compared with the wild type monomer. Even though Leu55 was identified as an hot spot during the CAS analyses, it was substituted with a basic residue and the adjacent Glu was modified into a hydrophobic leucine residues in order to keeping the hydrophobic interaction and creating an adjunctive salt bridge at the same time. Furthermore, Leu63 and Val66 were replaced with other hydrophobic residues. Peptide1, peptide 1_variant, peptide2 and peptide 2_variant were designed on the base of these considerations; in peptide 4 sequence instead residues Leu55 and Glu56 were maintained untouched (**Table 4**).

	SEQUENCE	LENGTH
Peptide wild type 1	RFKRVQQRHVLESEKNQLLQQV	22 aa
Peptide 1	EFKEVQERHVRLSEKNQLIQQV	22 aa
Peptide 1_variant	EFKDVQERHVKLSEKNQLIQQV	22 aa
Peptide 2	EFKEVQERHVRLSEKNQLI	19 aa
Peptide 2_variant	EFKDVQERHVRLSEKNQLLQQI	22 aa
Peptide 4	EFKDVQERHVLESEKNQLL	19 aa

Table 4: Sequences of the peptides designed with CAS: above the wild type monomer sequence is shown (in green are indicated the warm spots; in yellow the cold spots, in red the hot spots). The residues in purples are deviation from the wild type sequence.

Figure 48 shows the far LZ region; only one peptide was designed interacting with this sequence, peptide 5 in which Ile73 and Leu90, both identified as warm spots, were replaced with leucine residues and Val77, a cold spot, substituted with a isoleucine residue in order to increase the hydrophobic interactions.

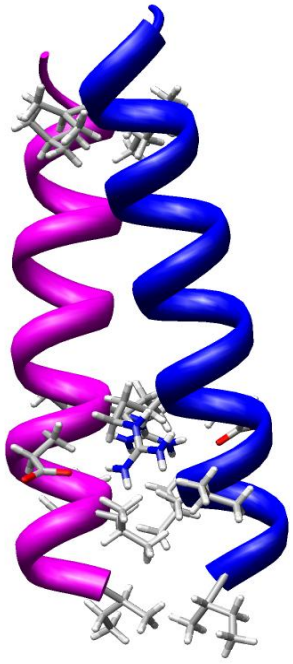


Figure 48: Detail representation of the c-Maf homodimer interaction in the far LZ region.

The peptide sequence is shown in **Table 5**:

NOME	SEQUENZA	LUNGHEZZA
Peptide wild type 2	ISRLVRERDAYKEKYEKL	18 aa
Peptide 5	LSRLLRERDAYKEKYEKI	18 aa

Table 5: Sequences of the peptides designed with CAS: above the wild type monomer sequence is shown (in green are indicated the warm spots; in yellow the cold spots, in red the hot spots). The residues in purples are deviation from the wild type sequence.

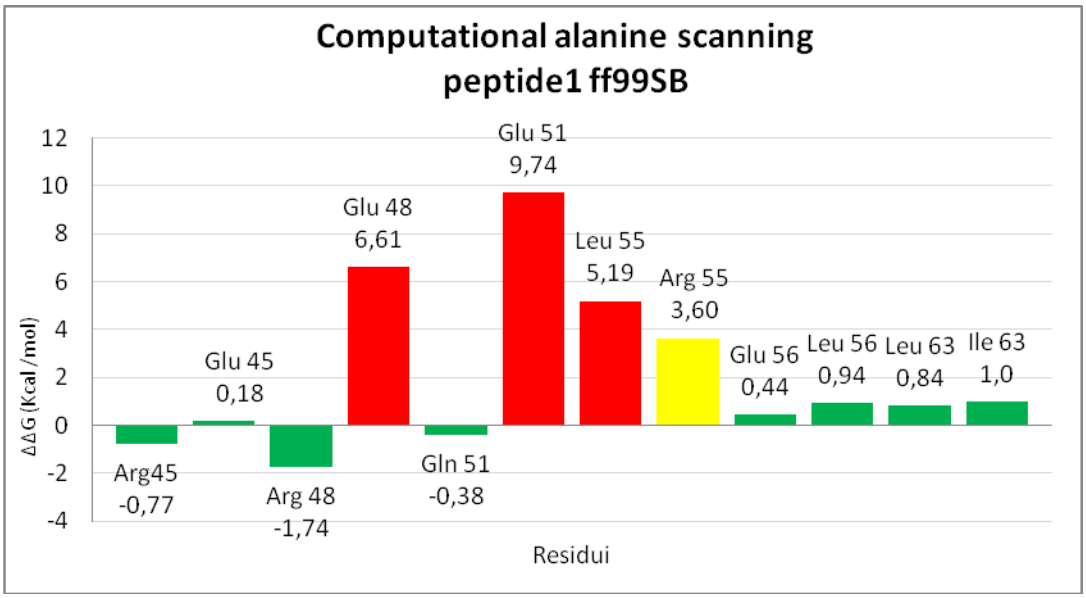
For each peptide, starting from the PDB file containing the c-Maf homodimer, the mutations were carried out only on one monomer using MOE software[27] changing one residue per time. For all the designed peptides a molecular dynamics study (same protocol applied before) was performed using ff99SB force field. The visual analyses revealed that all the peptides remained bonded to the monomer for the entire simulation time (8ns). The association free energy was calculated by MMGBSA for each peptide (solvent model igb=5; salt concentration 100nM) basing on 50 snapshots taken from each trajectory. The energetic data, shown in **Table 6** do not consider the entropic contribute.

NAME	BOND FREE ENERGY ESTIMATION	ST. DV.
Peptide wt1	-32,20 kcal/mol	±3,29
Peptide 1	-40,99 kcal/mol	±3,28
Peptide 1_variante	-49,83 kcal/mol	±7,43
Peptide 2	-43,77 kcal/mol	±3,68
Peptide 2_variante	-46,09 kcal/mol	±5,88
Peptide 4	-40,06 kcal/mol	±3,59
Peptide wt 2	-53,69 kcal/mol	±2,94
Peptide 5	-53,80 kcal/mol	±2,87

Table 6: Estimation of the bond free energy between c-Maf monomer and the designed peptides calculated by MMGBSA.

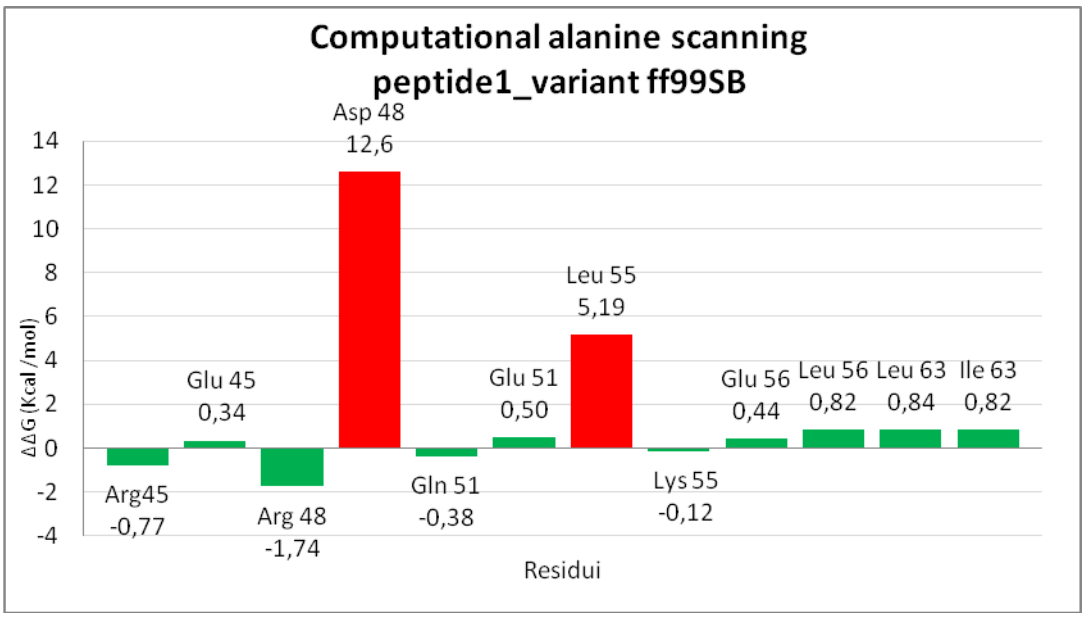
Successively, a CAS analyses was performed on the residues replaced in the designed peptides with the purpose of evaluating if the mutation into an alanine is more effective on the bond energy than the residues on the entire c-Maf wild type monomer (**Figure 49**).

Peptide 1: EFKEVQERHVRLSEKNQLIQQV (22aa)



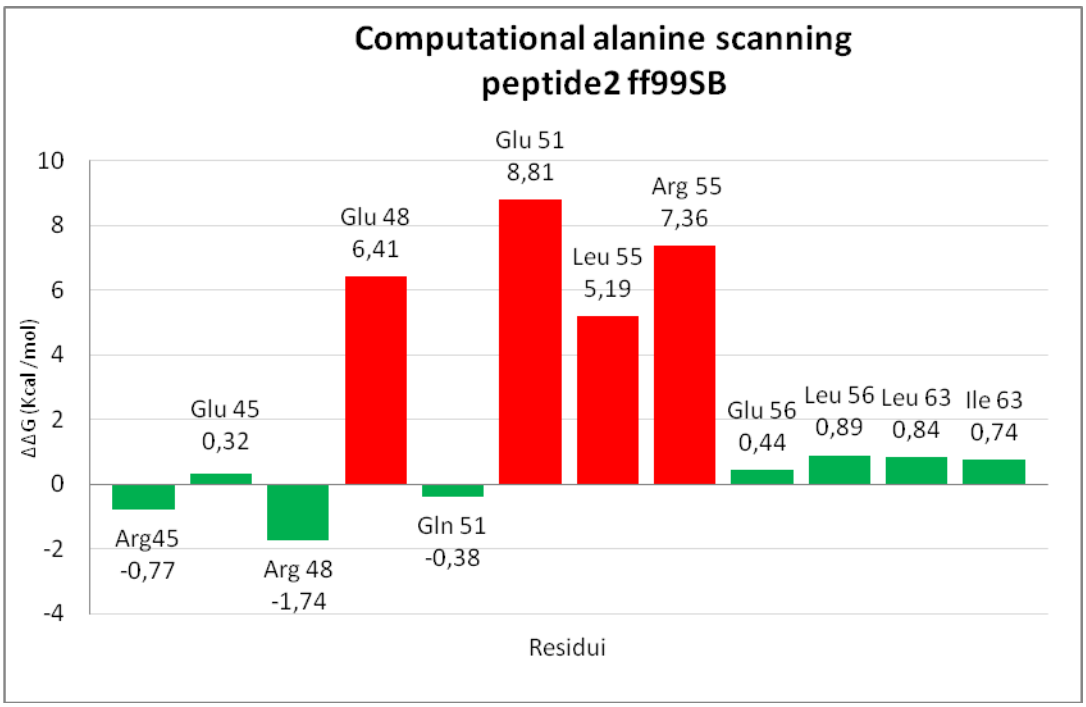
Glu45, Leu56, and Ile63 mutated into alanine have almost the same impact on the bond formation energy than Arg45, Glu56 and Leu63 of the wild type monomer; on the contrary the replacement of Glu48 and Glu51 residues provokes a much more significative change into the bond free energy than the equivalent Arg48 and Gln51 residues on the wild type monomer which means that this substitution is favorable.

Peptide 1_variant: EFKDVQERHVKLSEKNQLIQQV (22 aa)



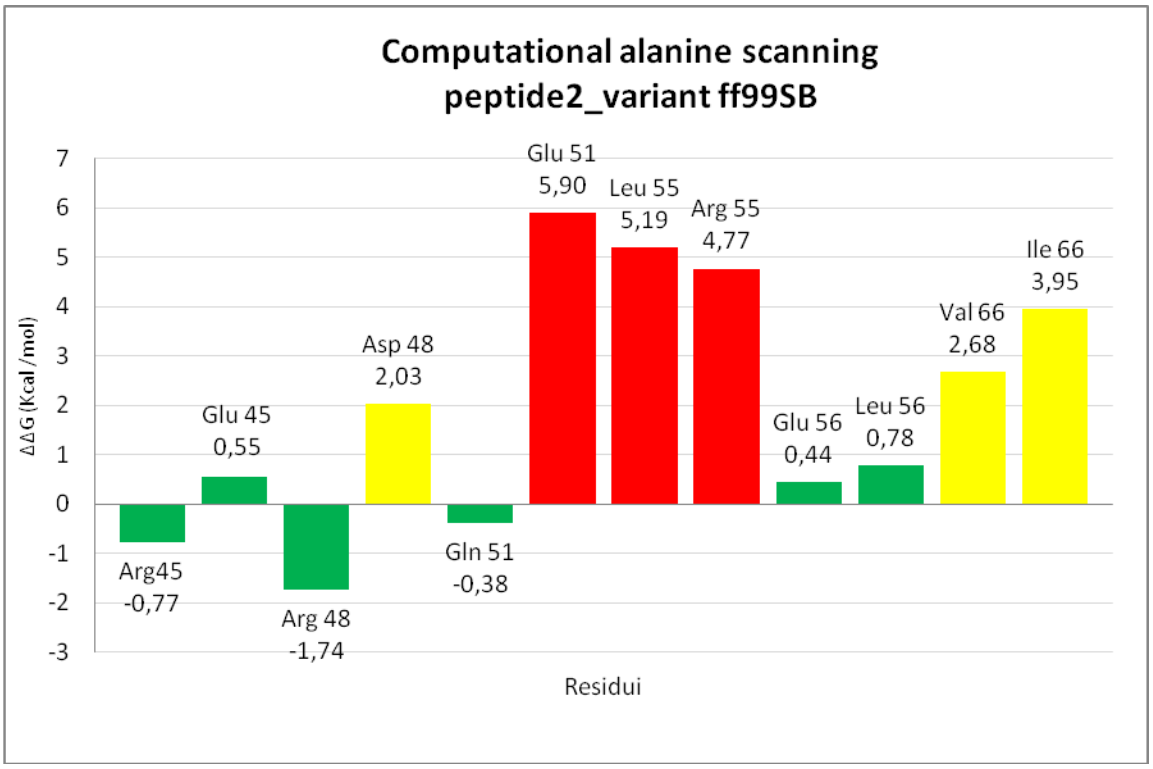
The only favorable substitution regards Arg48 residue with Asp, as a matter of fact once the replacement of Asp with alanine has a big influence on the complex formation energy.

Peptide 2: EFKEVQERHVRLSEKNQLI (19 aa)



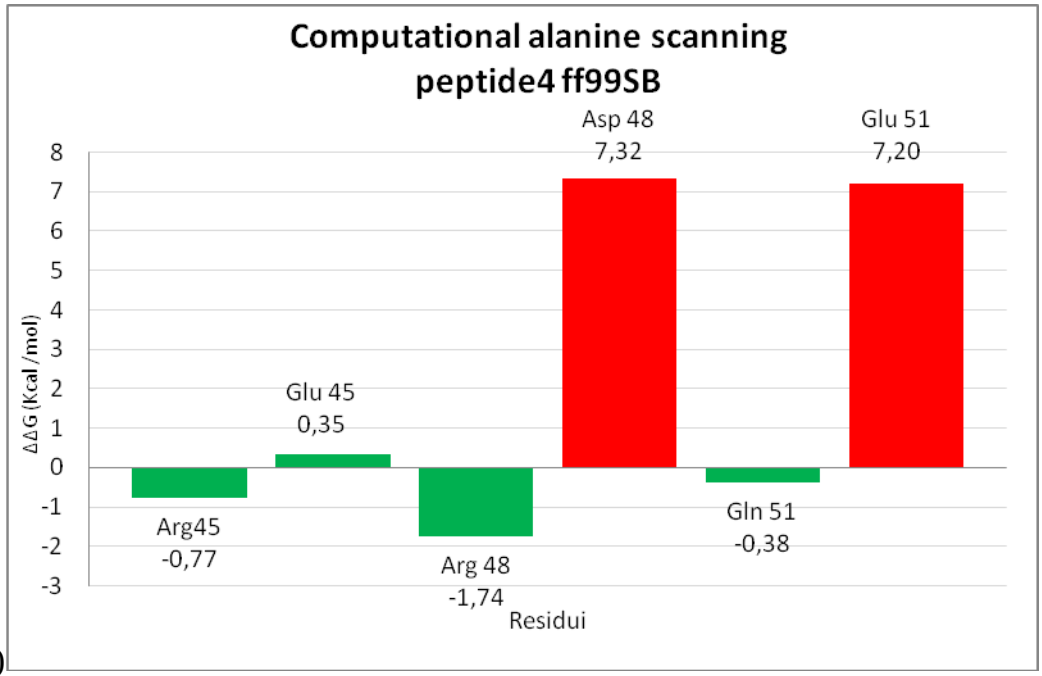
The substitution of Glu51, Glu48, and Arg55 residues with alanine weights much more on the complex formation energy than the corresponding mutations on the wild type monomer. Thus, the replacement results being useful.

Peptide 2_variant: EFKDVQERHVRLSEKNQLLQQI (22 aa)



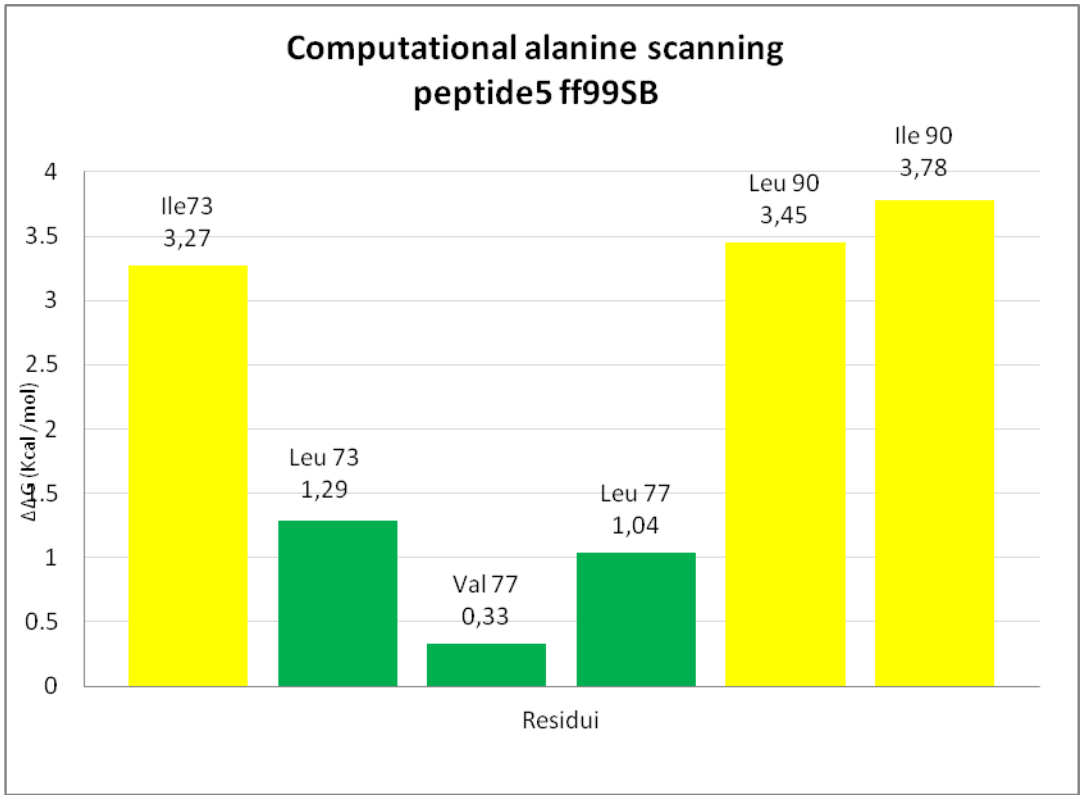
For what concerns peptide 2_variant all the substitutions result being more favorable than the corrispective residues on the wild type monomer.

Peptide 4: EFKDVQERHVLESEKNQLL (19 aa)



All the three substitutions result being favorable, moreover comparing the energetic data with peptide 2 from which peptide 4 differs only in mantaining the wild type residues Leu55 and Glu56, it could be assessed that the choice made is a potential succesful strategy.

Peptide 5: LSRLLRERDAYKEKYEKI (18 aa)



Peptide 5 is the only peptide designed for interfering in the C-terminal region of c-Maf homodimer. The replacement of Ile73 with a leucine is not particularly favorite while both Val77 and Leu90 substitution with, respectively, leucine and isoleucine, result effective.

In conclusion, using MMBGSA method it was possible to asses that both the two classes of peptides, maintain the ability of binding the c-Maf monomer. Moreover, they results to be termodinamically favorite compared to the wild type sequence. The best results were obtained with peptide 5; as in the first class of peptides, the most promising peptides outlined were peptide 1_variant; peptide2_variant and peptide2.

1.7.6. References

1) Adcock, A. S., McCammon, J. A., *Chem Rev.* **2006**, 106, 1589-1615.
2) Scheraga, H. A., Khalili, M., Liwo, A., *Annu. Rev. Phys. Chem.* **2007**, 58, 57-83.
3) Born, M., Oppenheimer, J. R., *Ann. Phys.* **1927**, 84, 457.
4) Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman, P. A., AMBER11 **2010**, University of California, San Francisco.
5) Mackerell, A. D., *J. Comput. Chem.* **2004**, 25, 1584-1604.
6) Straatsma, T. P., McCammon, J. A., *Ann. Rev. Phys. Chem.* **1992**, 43, 407-435.
7) Kollman, P., *Chem. Rev.* **1993**, 93, 2395-2417.
8) Almlof, M., Brandsdal, B. O., Aqvst, J., *J. Comput. Chem.* **2004**, 25, 1242-1254.
9) Gilson, M. K., Given, J. A., Bush, B. L., McCammon, J. A., *Biophys. J.* **1997**, 72, 1047-1069.
10) Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., Case, D. A., *J. Am. Chem. Soc.* **1998**, 120, 9401-9409.
11) Eriksson, M. A., Pitera, J., Kollman, P. A., *J. Med. Chem.* **1999**, 42, 868-881.
12) Banba, S., Guo, Z. Y., Brooks, C. L., *J. Phys. Chem.* **2000**, 104, 6903-6910.
13) Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, S., Srinivasan, J., Case, D. A., Cheatham, T. E., *Acc. Chem. Res.* **2000**, 33, 889-897.
14) Lee, M. R., Duan, Y., Kollman, P. A., *Proteins: Struct. Funct. Genet.* **2000**, 39, 309-316.
15) Pieraccini, S., Saladino, G., Cappelletti, G., Cartelli, D., Francescato, P., Speranza, G., Manitto, P. Sironi, M., *Nature chemistry* **2009**, 1, 642-648.
16) Moreira, I. S., Fernandes, P. A., Ramos, M., *Theor. Chem. Acc.* **2007**, 117, 99-113.
17) Massova, I., Kollmann, P. A., *J. Am. Chem. Soc.* **1999**, 121, 8133-8143.
18) Moreira, I. S., Fernandes, P. A., Ramos, M. J., *J. Phys. Chem. B.* **2006**, 110, 10962-10969.
19) Kurokawa, H., Motohashi, H. Kimura, M., Takagawa, H., Kanno, Y., Yamamoto, M., Tanaka, T., *Molecular and Cellular Biology* **2009**, 29, 6232-6244.
20) Yamamoto, T. M., Kyo, T., Kamiya, T., Tanaka, J. D., Engel, H., Motohashi, H., Yamamoto, M., *Genes Cells* **2006**, 11, 575-591.

21) http://swissmodel.expasy.org/

22) {HYPERLINK http://biophysics.cs.vt.edu/H}+ +

23) Miyamoto, S., Kollmann, P. A., *J. Comput. Chem.* **1992**, 13, 952-962.

24) Berendsen, H. J. C., Posta, J. P. M., Van Gunsteren, W. F., Dinola, A., Haak, J. R., *J. Chem. Phys.* **1984**, 81, 3684-3690.

25) Humphrey, W., Dalke, A., Schulten, K., *Journal of Molecular Graphics* **1996**, 14, 33-38.

26) Dlakić, M., Grinberg, A. V., Leonard, D. A., Kerppola, T. K., *The EMBO journal* **2001**, 20, 828-840.

27) MOE, Chemical Computing Group, Montreal, Quebec, Canada.

1.8. Synthesis of the c-Maf LZ modulators and circular dicroism studies

1.8.1. Peptides synthesis

The c-Maf LZ domain and the peptidic modulators outlined from *in silico* studies as described in chapter 1.7., were obtained by chemical synthesis using MW-SPPS strategies. In **table 7** are recapitulated the amino acidic sequences and the synthetic strategy adopted for the designed peptides:

PEPTIDE	SEQUENCE	ΔG estimation (kcal/mol)	MW (Da)	ADOPTED STRATEGY
LZ C-terminal	QQRHVLESEKNQLLQQVDHLKQ ISRLVRERDAYKEKYEKLV	-	5076.08	MW-SPPS
1	EFKEVQERHVRLSEKNQLIQQV	-40.99	2737.1	MW-SPPS
1_variante	EFKDVQERHVKLSEKNQLIQQV	-49,83	2695.05	MW-SPPS
2	EFKEVQERHVRLSEKNQLI	-43,77	2381.69	manual
2_variante	EFKDVQERHVRLSEKNQLLQQI	-46,09	2737.09	manual
4	EFKDVQERHVLESEKNQLL	-40.06	2340.0	MW-SPPS
Wt 2	ISRLVRERDAYKEKYEKL	-53,69	2295.64	manual
5	LSRLLRERDAYKEKYEKI	-53,80	2309.67	manual

Table 7: Recapitulating table of the synthesized peptides

1.8.2. Circular dicroism analyses

1.8.2.1. c-Maf LZ domain characterization

The CD studies were obtained in collaboration with the Pharmacy Department of Parma University. The role of the LZ domain is to allow homo or hetero-dimerization of b-ZIP factors through hydrophobic interactions due to the presence of several leucine residues. The spectrum of the entire c-Maf LZ domain was carried out in phosphate buffer (20mM, pH 7.40); the presence of two negative bands, respectively around 208nm (amidic transition p-p*) and at 222nm (ascribed to the n-p* transition), typical of α-helices, can be observed (**Figure 50**). The absence of a maximum at 198nm, attributed to the *excitation coupling* of the p-p* transitions, suggests the presence of several secondary conformations.

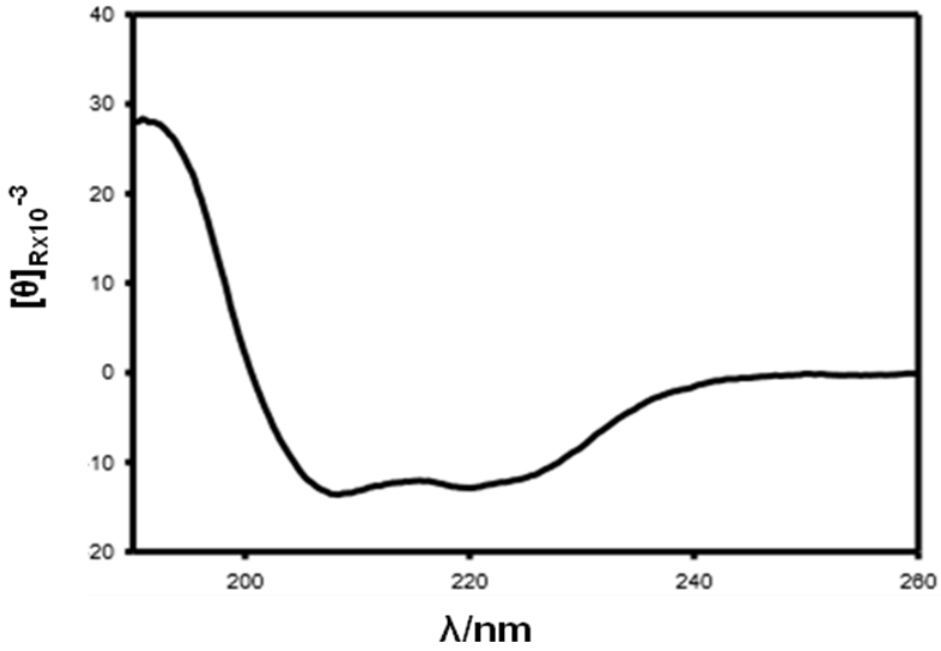


Figure 50: CD spectrum of c-Maf LZ domain in phosphate buffer pH 7.40

The LZ domain appears to be well structured in the absence of the basic region (BR) and the extended homology region (EHR) that complete the DNA binding domain of Maf proteins. The analyses of the CD spectrum secondary structure was performed using CONTIN algorithm (Dichroweb server) as shown in **Table 8.**

	HELIX	STRAND	TURNS	UNORDER
LZ c-terminal	43%	10%	15%	32%

Table 8: Calculated percentage of different secondary structure motifs obtained by Dichroweb server.

The high α-helical percentage even in the absence of the DNA target, could be ascribed to the hydrophobic interactions established by the leucine residues that allow the conformational disorder decrease through dimerization. In order to evaluate the oligomerization state in solution, essential information for optimizing the model of study for the dimerization inhibitors, the LZ domain CD spectrum was measured at different concentrations (**Figure 51**). In this study cuvettes presenting

different pathlength were used with the aim of keeping the signal/rumor ratio constant compensating the lower signal with a longer pathlength.

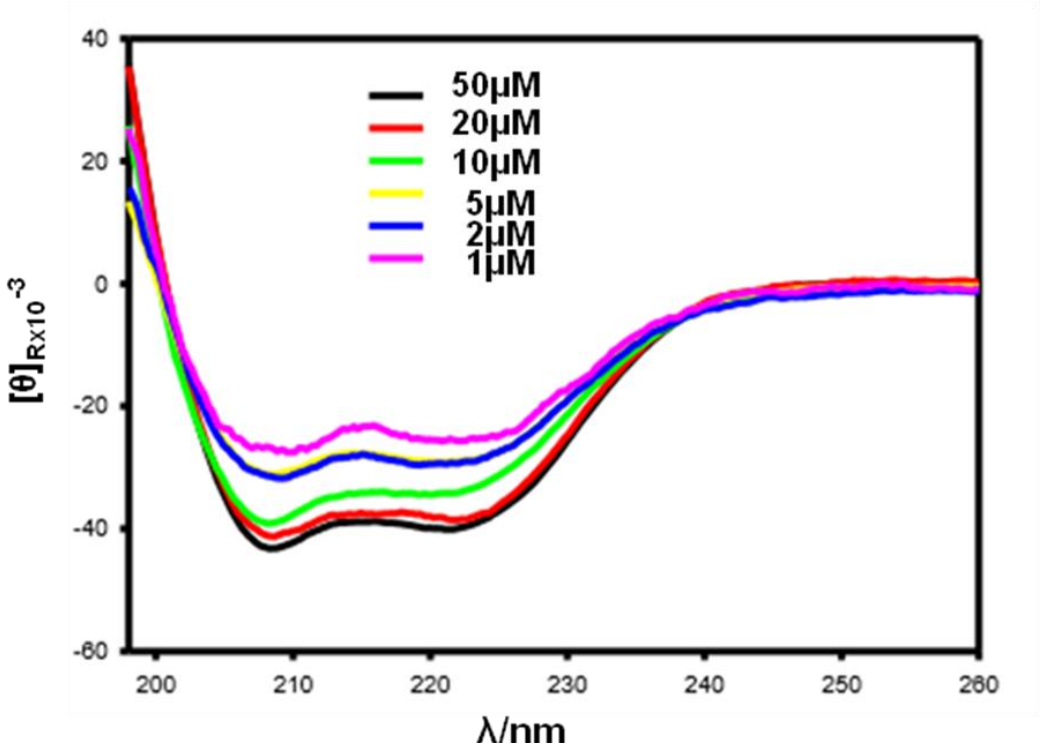


Figure 51: CD spectra of c-Maf LZ domain at different concentrations

The dilution provokes a diminishing in the α-helical content assuming a shift to the monomeric form of the LZ domain as a consequence of a loss of secondary structure. In **Figure 52** are reported the c-Maf LZ domain ellipticity values at 208nm as a function of its concentration; a deceasing tendency could be observed corresponding to a minor α-helical contents. Thus it could be asserted that the monomeric/dimeric equilibrium state of the LZ domain is directly correlated to its concentration in solution.

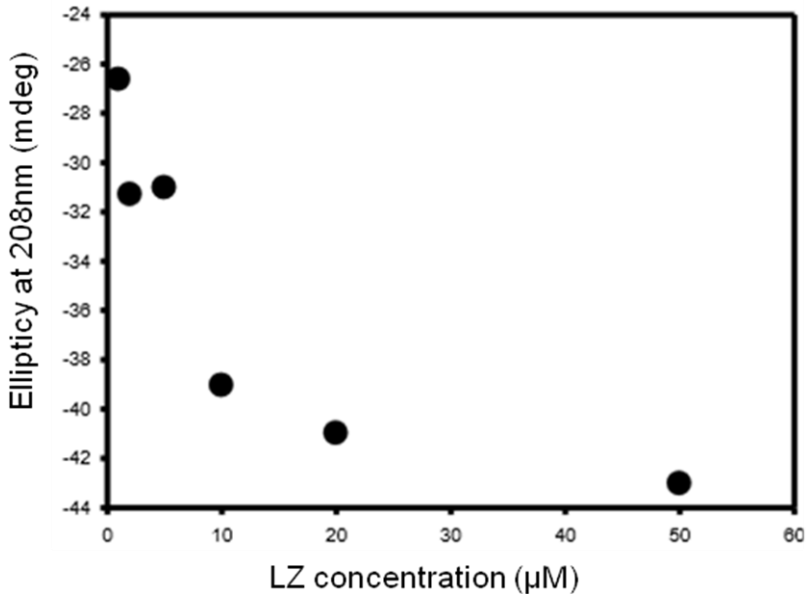


Figure 52: Ellipticity values at 208nm as a function of LZ concentration

1.8.2.2. CD analyses of dimerization inhibitors

Initially the spectra of the single peptides were measured at 20µM concentration (**Figure 53**).

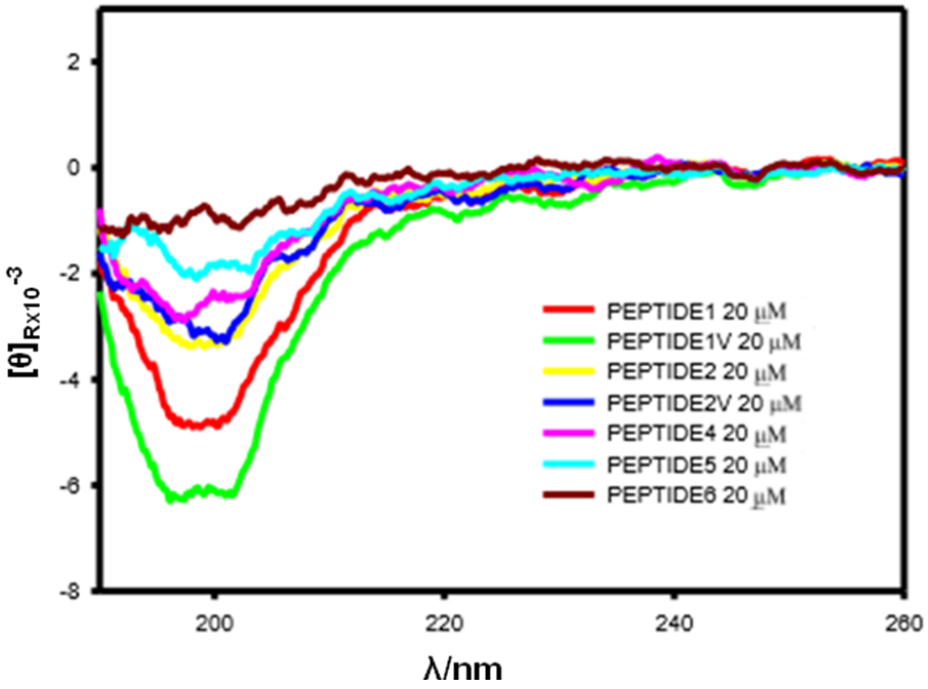


Figure 53: CD spectra of the peptides alone (concentration 20µM)

In solution the peptides exhibit a unstructured form leading to the assumption that they would not be able to form homodimers. In a second time, the interaction studies with the LZ domain were also performed measuring the spectra of the binary mixtures at equimolar concentration (20μM) after overnight incubation (**Figure 54**).

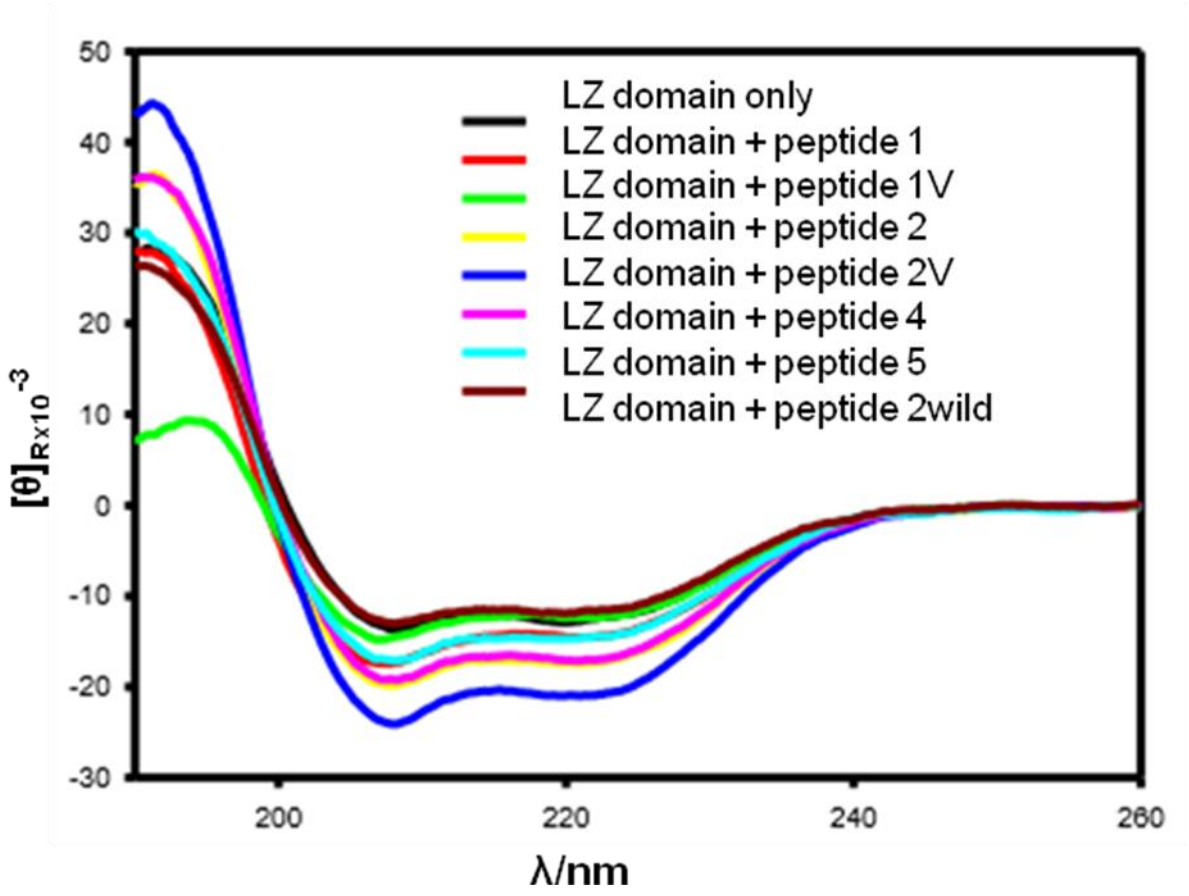


Figure 54: CD spectra of the LZ domain and its interactions with the inhibitors designed *in silico*.

The spectra of the interactions showed an α-helical content higher than the spectrum of the LZ domain alone. In order to verify if the increase is caused by the contribute of the peptide or if an interaction is effective through a three dimensional structure stabilization, the spectra of the mixtures were compared with the ones obtained from the arithmetic sum of the LZ and the peptide ellipticity contributes taken singularly (From **Figure 55** to **Figure 61**).

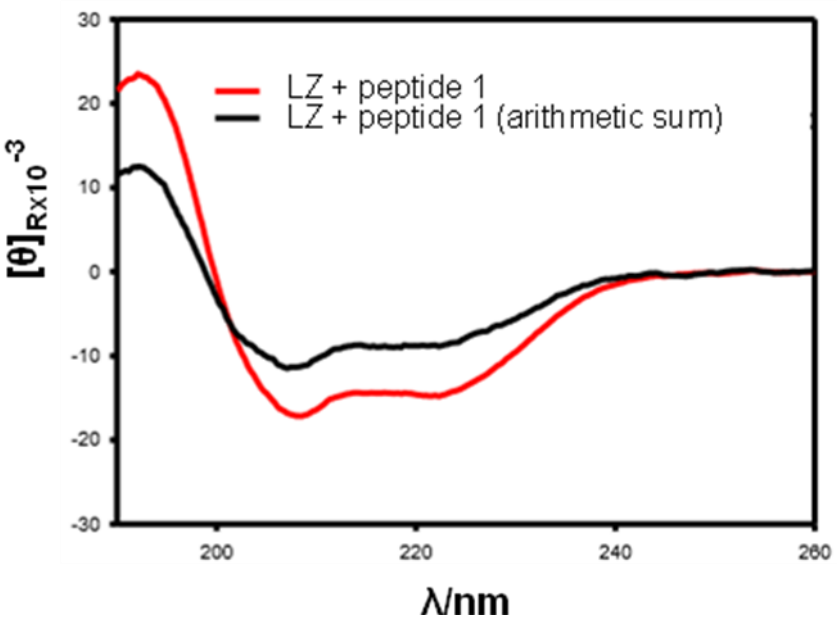


Figure 55: Comparison of the CD spectra of the LZ/peptide 1 interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

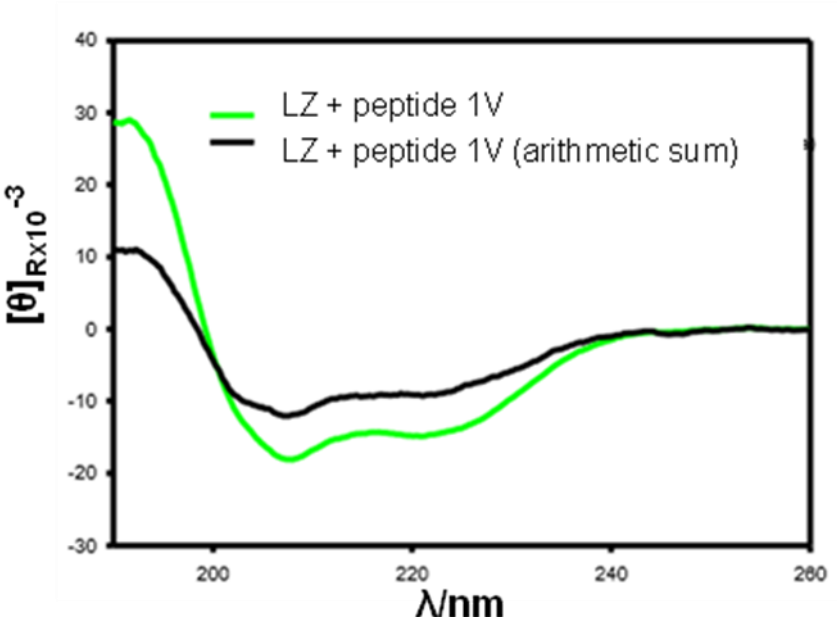


Figure 56: Comparison of the CD spectra of the LZ/peptide 1V interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

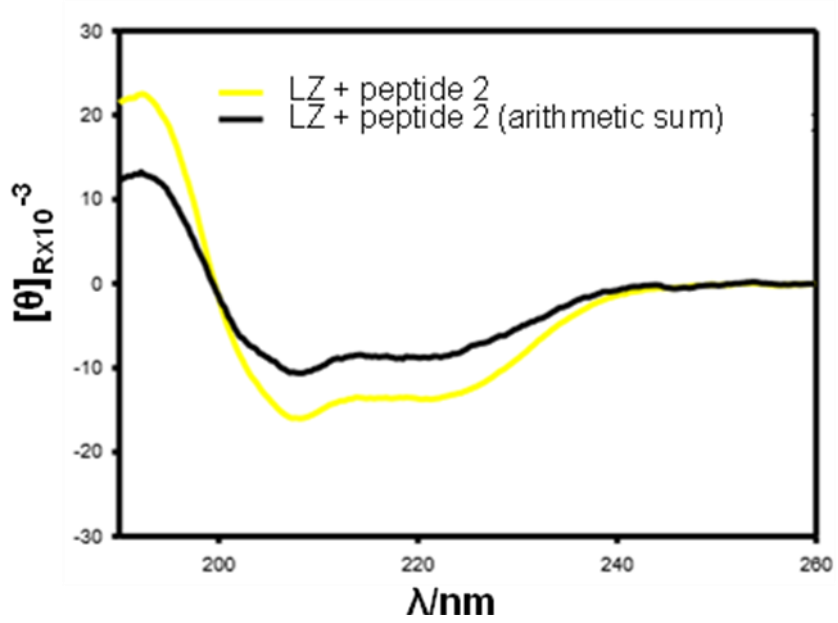


Figure 57: Comparison of the CD spectra of the LZ/peptide 2 interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

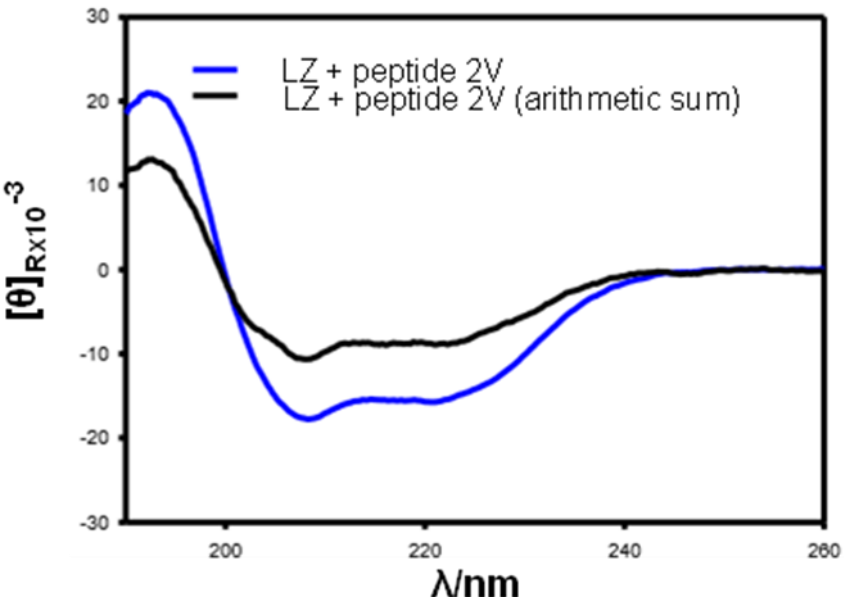


Figure 58: Comparison of the CD spectra of the LZ/peptide 2V interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

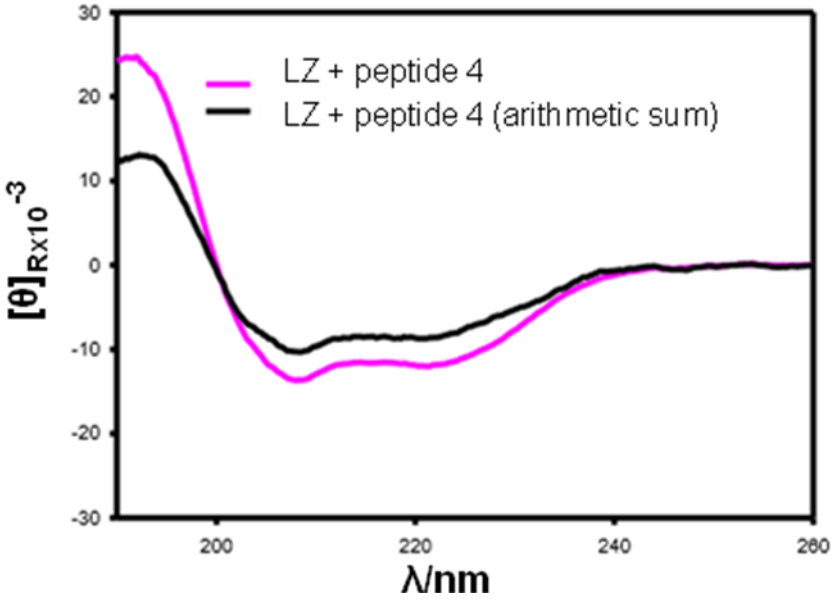


Figure 59: Comparison of the CD spectra of the LZ/peptide 4 interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

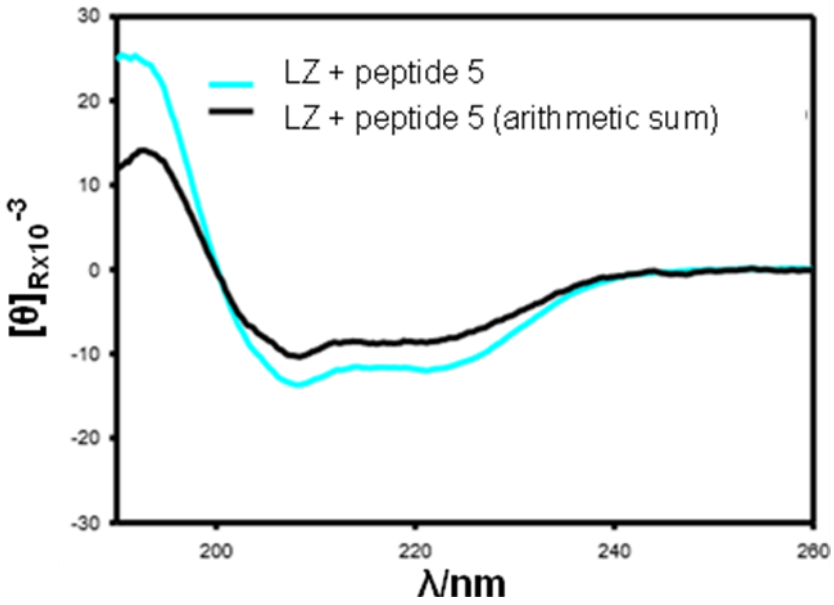


Figure 60: Comparison of the CD spectra of the LZ/peptide 4 interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

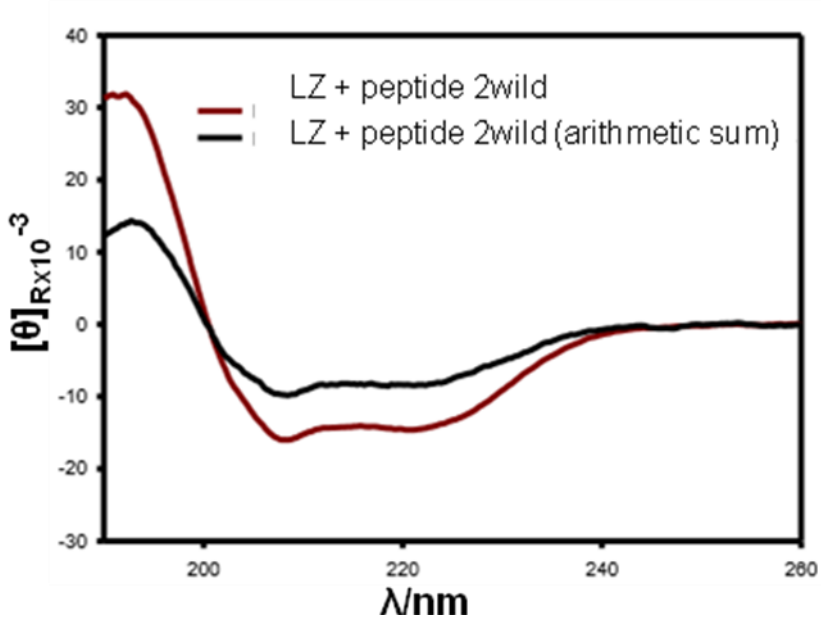


Figure 61: Comparison of the CD spectra of the LZ/peptide 2wild interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

The spectra of the mixtures exhibit an α-helical content higher than the arithmetic sum of the contributes of the peptides taken singularly, as enlightened by the intensity of the peaks at 208nm and 222nm, thus showing that a successful interaction occurs between the LZ domain and the inhibitors. An ulterior confirmation was obtained by a secondary structure analyses of the mixtures carried out using Dichroweb server (CONTIN server) (**Table 9**).

	HELIX	STRAND	TURN	UNORDER
LZ domain alone	43%	10%	15%	32%
LZ + peptide 1	64%	1%	9%	31%
LZ + peptide 1V	59%	1%	9%	31%
LZ + peptide 2	69%	1%	10%	20%
LZ + peptide 2V	74%	1%	9%	16%
LZ + peptide 4	67%	1%	10%	22%
LZ + peptide 5	64%	1%	10%	25%
LZ + peptide 2wild	58%	2%	11%	29%

Table 9: Secondary structure analyses of the LZ domain/inhibitors mixtures carried out by Dicroweb server.

An increase of the α-helical percentage and a consequent decrease of the unordered fraction is noticeable for each mixture endorsing the hypothesis that a positive interaction occurs trough a LZ domain structure consolidation. The most promising peptide results to be peptide 2V (74% α-helical contents). Peptide 2V inhibitory activity was investigated; in **Figure 62** LZ ellipticity is displayed as a function increasing concentration of inhibitor. The results show a decreasing tendency in the LZ domain ellipticity supporting the hypothesis of an effective dimerization inhibition.

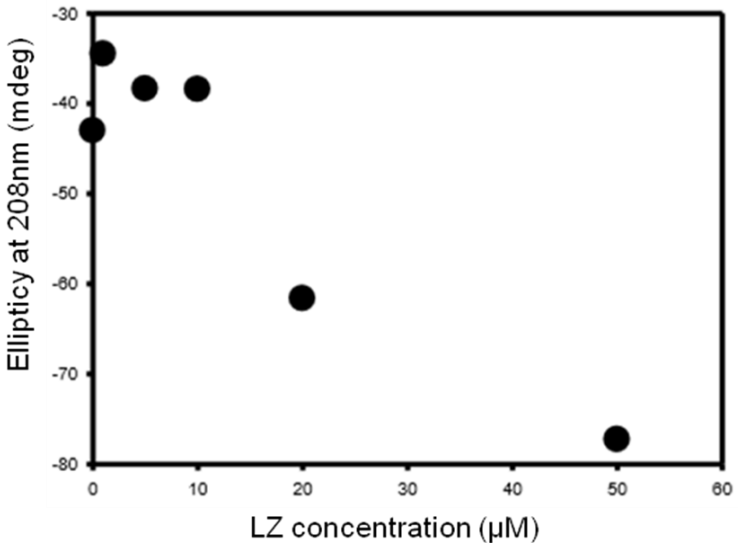


Figure 62: LZ domain ellipticity values plotted as a function of peptide 2V inhibitor concentration.

With the aim of investigate the specificity of the dimerization inhibitors for c-Maf LZ domain, the CD spectra of the peptide with MafA LZ, which presents a high level of homology with c-Maf, were registered in the same experimental conditions (**Figure 63**).

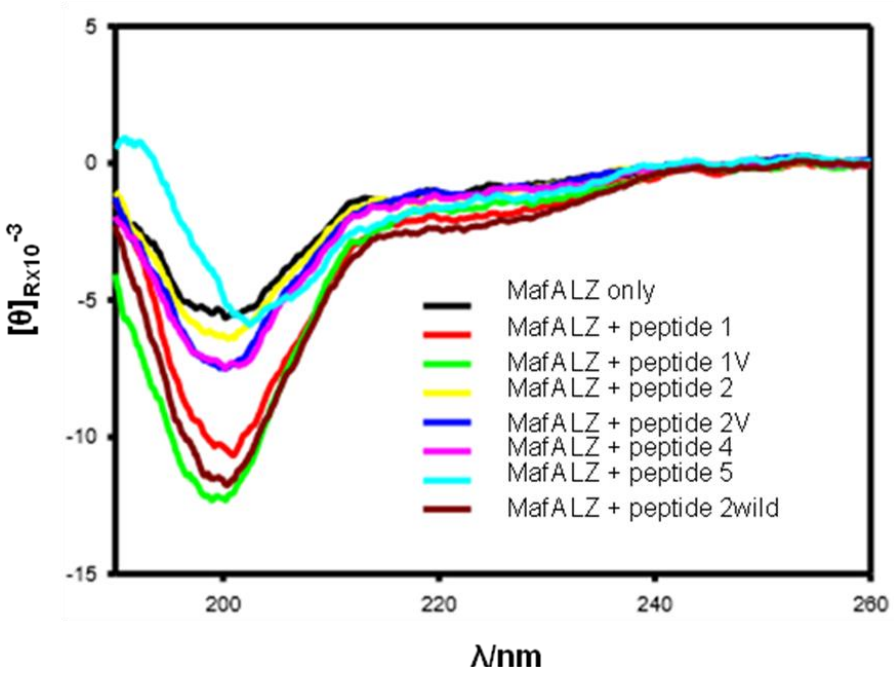


Figure 63: CD spectra of MafA LZ domain with the c-Maf LZ domain inhibitors.

It could be observed that MafA LZ domain itself is not structured however the presence of the inhibitory peptides not only doesn't improve the secondary structure stability. An increase in the peak at 200nm is observed due to the sum of the ellipticity contributes of both the peptides as it is confirmed by the comparing the spectrum of the mixture with the one obtained from the arithmetic sum of the contributes of the two peptides (**Figure 64**).

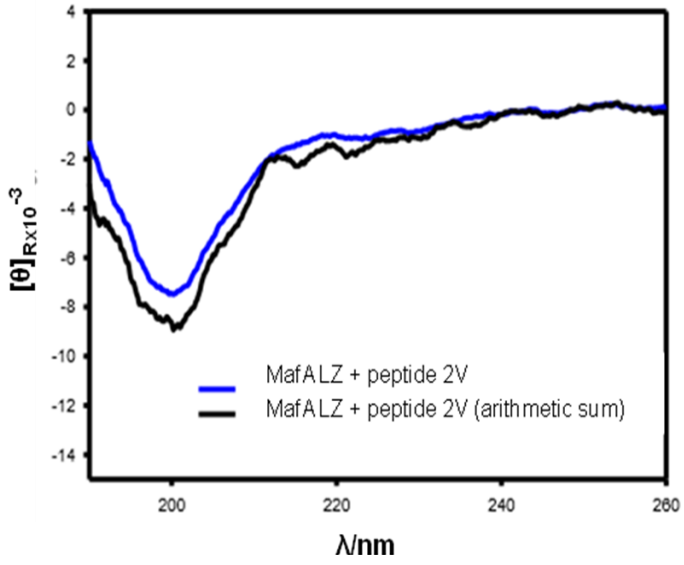


Figure 64: Comparison of the CD spectra of the MafA LZ/peptide 2wild interaction mixture and the arithmetic sum of the separated contributes of the two peptides.

From this study it can be evicted that the inhibitory activity of the peptides is specific for their target c-Maf, the successful interaction occurs through a three dimensional stabilization that implies an increase in the α-helical contents of the complex.

1.9. Conclusions

In this work our goal was to study the DNA binding of Mafs proteins bLZ transcription factors whose sequence is well conserved within Maf family members. In particular, we focused on a large Maf, c-Maf whose deregulation plays a fundamental role into a major pathology, currently untreatable, multiple myeloma. The overexpression of c-Maf is observed in the 50% of multiple myeloma cases associated with poor prognosis. Because of the lack of large Mafs crystallographic data available in literature, the creation of an c-Maf homology model was made decisive for the design of modulators. Moreover, thanks to the DBD similarity, and in order to have a better understanding of the target-interaction mode of Mafs DBD, we decided to extend the study to another large Maf protein, MafA, which is involved into the regulation of insulin synthesis, secretion and metabolism.

Starting from the crystallographic data available on PDB regarding the complex G-Maf, a small Maf, and its DNA target, MARE25, an homology model of the complex c-Maf/T-MARE was outlined. Successively, performing molecular dynamics studies, we observed the folding behavior of c-Maf in the presence and absence of its DNA target T-MARE. Comparing the results with the ones obtained from G-Maf/MARE25 complex we noticed a common behavior: in the absence of the DNA target, the EHR region, which represents, together with the basic region, the DNA binding domain, shows a progressive unfolding tendency diminishing its helical content. On the other hand, in the presence of the DNA target, the DBD maintains a folded structure indicating that the interaction is effective through a three-dimensional conformation stabilization. Our observations are consistent with the CD experiments data previously reported in literature. We succeeded in obtaining the chemical synthesis of the entire 75mer MafA DNA binding domain adopting MW-SPPS strategies. The folding properties of the entire MafA DBD and its fragmentations, and the interaction mode with its DNA target (the human insulin promoter) were also investigated using circular dicroism. These results gave us an ulterior confirmation of the previously enlightened tendency: a disorder-order transition is at the base of the recognition mechanism of Mafs DBD and DNA target sequence. In a second time, using computational alanine scanning method, we carried out the *in silico* design of c-Maf peptidic modulators that act as inhibitors of dimerization which is fundamental for DNA binding and transcription activation. The peptides were obtained by chemical synthesis and an activity screening was carried out by circular dicroism. These studies confirm the inhibitory activity and specificity of the peptides towards their target, c-Maf LZ domain. However, in this preliminary phase, the peptides pharmacokinetics and

pharmacodynamics aspects were not evaluated. The future developments of this project will be focused on the pharmacokinetics parameters improvement by creating peptidomimetics modulators.

1.10. Materials and methods

1.10.1. Materials

Fmoc protected amino acids, Rink amide PEG MBHA resin, HBTU and DIPEA were purchased from Iris Biotech Gmbh. HOBT, piperidine, acetic anhydride, triisopropylsilane (TIS), thioanisole (TAN), phenol, TFA, and HPLC-grade solvents were from Sigma Aldrich. 1-Methyl-2-pyrrolidinone (NMP) was obtained from BDH/VWR/PROLABO and 1,2-ethanedithiol (EDT) from Fluka. The DNA sequences were purchased at Primm Srl. All peptides were prepared by MW-SPPS using a Liberty Microwave Peptide Synthesizer (CEM Corporation) and purified by preparative RP-HPLC using Jasco BS-997-01 equipment and a DENALI C-18 column from GRACE VYDAC (10 lm, 250 9 22 mm). Two mobile phases were used: A = 95 % water, 5 % ACN, 0.1 % TFA; B = 95 % ACN, 5 % Water, 0.1 % TFA. The peptides were analyzed by analytical HPLC using a Jasco BS-997-01 equipment and detector series Star 800 from VARIAN. The mass spectra were recorded on a LCQ Advantage spectrometer from Thermo Finnigan for ESI–MS analysis and on Autoflex 3 from Bruker Daltonics for MALDI-TOF analysis. The CD spectra were recorded at room temperature using a Jasco J-810 spectropolarimeter and a 0.01 cm quartz cell (Hellma Suprasil).

1.10.2. MW-SPPS

Peptides were synthesized C-terminally amidated and N-terminally acetylated using Rink amide PEG MBHA resin with a loading of 0.35 $mmolg^{-1}$ and on a 0.1 mM scale. Amino acid side-chain protection was as follows: tBu for Asp, Glu, Tyr, Thr, Ser; Pbf for Arg; Boc for Lys; Trt for Asn, Gln and His. Protocol A (for peptide up to 28 residues): single couplings were performed with Fmoc-amino acid (5 equiv.) activated in situ with HBTU (5 equiv), HOBt (5 equiv.) dissolved in DMF, and DIPEA 1M in NMP (10 equiv.); each coupling was achieved using MW irradiation (5 min, 75 ºC, 20W; for His and Cys 50ºC, 2 min at 0W and 4 min at 25W). Fmoc cleavage was accomplished by treating the peptidyl-resin with 20% piperidine in DMF (3 min, 75ºC, 40 W). Protocol B (for Arg and, starting from the 35th residue, for all amino acids): double couplings were performed with Fmoc-amino acid (5 equiv.) activated in situ with HBTU (5 equiv.), HOBt (5 equiv.) dissolved in DMF, and DIPEA 1M in NMP (10 equiv.). The first coupling was performed at room temperature (1 h and without MW irradiation);

the second one was achieved using MW energy (5 min, 75ºC, 25W). Fmoc cleavage was accomplished by treating the peptidyl-resin with 20% piperidine in DMF (3 min, 75ºC, 40 W).

1.10.3. N-terminal acetylation

Synthesized peptides were manually acetylated at the N-terminus on resin using acetic anhydride (10 equiv.) and DIPEA (10 equiv.) in DCM for 30 min at room temperature.

1.10.4. Cleavage from the resin

The peptides were manually cleaved from the resin with a mixture of TFA/water/TIS (90:5:5 v/v) for 6 h at room temperature or, when Cys and Met were present with the mixture TFA/water/TIS/TAN/EDT (90:3:5:5:3 v/v). The crude peptides were precipitated from ice-cold diethyl ether and recovered by centrifugation at 4ºC for 5 min (4,500 rpm). Three diethyl ether washes/centrifugation cycles were carried out to efficiently remove the scavengers. If Cys was present, a mixture of t-butyl-methyl ether/ petroleum ether (1:1 v/v) was used to avoid oxidation.

1.10.5. Peptide purification and characterization

HPLC and MS data of synthesized peptides

The peptides were analyzed by analytical HPLC using two mobile phases: A = 95% Water, 5 % ACN, 0.1 % TFA, B = 95% ACN, 5% Water, 0.1% TFA and a gradient of: 95% A for 5 min., then 95-50% A over 50 minutes. UV detection was made at 220 nm. The mass spectra were recorded on a LCQ Advantage spectrometer from Thermo Finnigan for ESI-MS analysis and on Autoflex 3 from Bruker Daltonics for MALDI-TOF (sinapinic acid or 4-CHCA matrix) analysis (**Table 10**).

Human MafA DNA binding domain					
EERFSDDQLVSMSVRELNRQLRGFSKEEVIRLKQKRRTLKNRGYAQSCRFKRVQQRHILESEKCQLQSQ VEQLKLEVGR					
Name	**Chain Length**	**$MW_{calculate}$ (Da)**	**MW_{found} (Da)**	**HPLC T_R (min)**	**Mg purified**
H3 short	248-263	2039.49	2038.18	14.08	4,0
H1-H2	223-243	2592.85	2592.30	18.19	9,0
LZP	277-301	3020.44	3019.65	16.91	16,0
BR	248-275	3531.19	3532.30	15.27	18,3
H2-H3short	236-263	3537.16	3536.12	16.54	35,7
EHR-BR	227-278	6365.33	6360.5	23.84	14,6
EHR-H3	227-263	4500,18	4494,6	18,80	9,3
H3	248-301	6622.71	6621.30	16.39	23,0
DBD	227-301	9001,48			

Table 10: Recapitulatory analyses of the synthesized peptides.

.

1.10.6. General procedure for DNA annealing

The two complementary DNA oligomers were diluted in HPLC-quality water, heated at 95ºC and gradually cooled (10ºC every 10 min) to room temperature. The obtained duplex was stored at -20C before using.

1.10.7. Circular dichroism on MafA

Peptide stock solutions were prepared in HPLC-quality water (100 M, 1.5mL). Spectra were obtained from 195 to 250nm with a 0.1nm step and 1 s collection time per step, taking three averages. The spectrum of the solvent was subtracted to eliminate interference from cell, solvent, and optical equipment. The CD spectra were plotted as mean residue ellipticity h (degree 9 cm2 9 dmol-1) versus wave length k (nm). Noise-reduction was obtained using a Fourier-transform filter program from Jasco.

1.10.8. Computational part

Molecular dynamics protocols with AMBER11:

The generation of *prm top* files (containing molecular topology description and force field parameters) and *inpcrd* files (describing the atomic coordinates) was performed with tleap. The solvation model used is TIP3P, the water box has a parallelepipeds shape.

Hydrogens minimization phase:

- Minimization *conjugate gradient* cycle number: 5000 (max cyc.);
- Minimization *steepest descent* cycle number: 1000 (max cyc.);
- Periodic condition (ntb=1);
- *Cut-off*= 8;
- Output frequency: 100.

Water and ions minimization phase:

- Minimization *conjugate gradient* cycle number: 5000 (maz cyc.);
- Minimization *steepest descent* cycle number: 2000 (max cyc.);
- Periodic condition (ntb=1);
- *Cut-off*= 8;

- Output frequency: 100 (ntpr, ntwe, ntwx).

NVT and NTP equilibration:

- Freedom of movement just for water and ions;
- Initial temperature 0K, final temperature 300K (temp0 and temp*i*);
- Steps number: 3000/ Time step 0.0015ps (nstlim and dt);
- Shake (ntc= 2);
- Periodic condition (ntb= 1);
- *Cut-off*= 8;
- Output frequency: 500 (ntpr, ntwe, ntwx).

Side chains, water and ions minimization phase:

- Minimization *conjugate gradient* cycle number: 5000 (max cyc.);
- Minimization *steepest descent* cycle number: 2500 (max cyc.);
- Periodic condition (ntb=1);
- *Cut-off*= 8;
- Output frequency: 100 (ntpr, ntwe, ntwx).

6 heating steps:

- From 0K to 50K (5ps; backbone restraint: 10kcal/mol), from 50K to 100K (5ps; backbone restraint: 9 kcal/mol), from 100K to 150K 85ps; backbone restraint: 8 kcal/mol), from 150K to 200K (5ps; backbone restraint: 6kcal/mol) and from 250K to 300K (backbone restraint: 5kcal/mol).
- Periodic conditions (ntb=1);
- Output frequency: 500 (ntpr, ntwe, ntwx);
- Temperature control (Berendsen thermostat, ntt=1);
- For each phase: time step 0.0005ps; nstlim 10000;
- Shake (ntc= 2);
- *Cut-off*= 8.

NVT dynamics for 100ps (backbone restrictions: 5kcal/mol):

- Periodic conditions (ntb= 1);
- Output frequency: 500 (ntpr, ntwe, ntwx);
- Shake (ntc= 2);
- Time step 0.002ps; nstlim 50000;
- *Cut-off*= 8;
- Temperature control (Berendsen thermostat, ntt=1).

4 dynamics phases NPT (for 100ps, backbone restraint: 5kcal/mol; 4kcal/mol, 3kcal/mol, 2kcal/mol; from 1000ps at 1kcal/mol without restraint):

- Periodic conditions (ntb= 2);
- Output frequency: 500 (ntpr, ntwe, ntwx);
- Temperature control (Berendsen thermostat, ntt=1).
- Shake (ntc= 2);
- *Cut-off*= 8;
- Time step 0.002ps; nstlim 50000.

8 dynamics phases NPT for 1000ps without backbone restraint:

- Periodic conditions (ntb= 2);
- Output frequency: 500 (ntpr, ntwe, ntwx);
- Temperature control (Berendsen thermostat, ntt=1).
- Shake (ntc= 2);
- *Cut-off*= 8;
- Time step 0.002ps; nstlim 500000.

<u>MMGBSA calculation protocol</u>

- Endframe= 50 (time at which interrupting the snapshots extrapolation from each trajectory);
- Verbose= 1 (variable describing the output file information amount);
- Igb= 5 (Generalized Born method; solvent model 5);
- Saltcon= 0.100 (salt concentration).

Alanine scanning calculation protocol

- Startframe= 1 (starting time for extracting snapshots from the trajectory);
- Endframe= 50 (time at which interrupting the snapshots extrapolation from each trajectory);
- Verbose=1 (variable specifying the output file information amount);
- Igb= 5 (Generalized Born method, solvent model 5);
- Saltcon= 0.100 (Salt concentration).

Chapter 2:

Construction of Ratiometric Fluorescent Sensors by Ribonucleopeptides

2.1. Introduction: fluorescent biosensors design strategies

Molecular tools for shedding light on the complex interplay between macromolecules, signaling molecules, and biologically important ions inside the cells play a central role in molecular and cell biology. In the past two decades, a lot of attention has been devoted by chemists and biologists to develop a "biosensor" that allows the real-time tracking of a small molecule of interest in living cells. A biosensor consists of a receptor component to capture a target ligand and a signal transduction component to convert the ligand-binding event into measurable signals, such as fluorescence, chemiluminescence, colorimetric, electrochemical, and magnetic responses. Especially recent advances in the development of fluorescent biosensors have made an outstanding contribution to elucidating not only the roles of individual biomolecules, but also the dynamic intracellular relationships between these molecules.[1] However, rational design strategies of fluorescent biosensors are not as mature as they look. Despite the fact that various sensors have been developed to date, a lengthy and laborious trial-and-error procedure is usually required to obtain a sensor with optimized functions since the interplay between recognition and signal-transduction event is unique for each individual sensor. An insatiable request for the establishment of a more universal and versatile strategy continues to provide an attractive alternative, so-called modular strategy, which permits facile preparation of biosensors with tailored characteristics by a simple combination of a receptor and a signal transducer. Various kinds of fluorescent biosensors constructed by synthetic receptors[2-9] and biological macromolecular receptors such as proteins[10] and aptamers[11] have been reported to date. The construction of fluorescent biosensors generally relies on the rational design strategy as follows. The first step requires an effort to find a macromolecular receptor with appropriate affinity and specificity to the target. The second step integrates the signal transduction function induced by the molecular recognition event into the receptor. Because the native biological receptor usually lacks an inherent property of the signal transduction function, foreign reporter moieties such as an auto-fluorescent protein (AFP) and a synthetic fluorophore must be introduced at the appropriate position of the receptor component.

2.1.1. Auto-fluorescent protein (AFP) based biosensors

Auto-fluorescent proteins (AFPs) such as green fluorescent protein (GFP) from the jellyfish[12] are well-established and versatile reporter proteins for monitoring gene expression profiles[13] and protein localizations[14] in a variety of systems. It is noteworthy that AFPs exhibit spontaneous fluorescence emission in cells by the autocatalytic formation of the chromophore after translation.[15,16] Therefore, AFPs can be endogenously expressed in cells or tissues just by transfection of the plasmid DNA without interfering with their fluorescence properties and damaging the cells. In addition to the application of AFPs as a reporter tag, various kinds of AFP-based biosensors have recently been developed by fusion of receptor proteins or mutation of AFPs. There are practically two strategies for the construction of AFP-based biosensors; (a) analyte-sensitive sensors and, (b) conformation-sensitive sensors.[17] The design of analyte-sensitive sensors is based on AFP variants, whose fluorescent properties are directly affected by the interaction between a target molecule and a chromophore moiety in AFP. Initially, pH and halide-sensitive AFP variants have been developed exploiting the intrinsic pH sensitivity of GFP mutants[18-20] and the high pK_as of YFP mutants.[21-23] Mutations in close proximity to the GFP chromophore or the barrel structure of BFP lead to the specific biosensor for Hg^{2+} [24] or Zn^{2+},[25] respectively. In this type of sensor, the receptor function is directly integrated into the chromophore itself by the alteration of the chemical nature around the chromophore. The conformation-sensitive sensors are designed so that the conformational change of the receptor associated with the ligand-binding event transduces to a significant fluorescence response of AFPs. This design strategy is more versatile than that for the analyte-sensitive sensor, because this type of sensor could be applied to a variety of native receptor proteins. Actually, biosensors for Ca^{2+} [26-37] as well as for small organic molecules such as ATP[38], cAMP[39-42], cGMP[43-45], tryptophan[46], glutamate[47,48], and inositol phosphates[49,50] have been reported based on this strategy. This type of sensors could be classified in three categories; (1) dual AFP-fused fluorescent resonance energy transfer (FRET)-based biosensors, (2) single circularly permuted (cp) AFP-based biosensors, and (3) split AFP-based biosensors.

2.1.1.1. Dual AFP-fused FRET-based biosensors

FRET is a physicochemical phenomenon which only occurs when two fluorophores are in sufficient proximity (<10 nm) of each other, and the emission spectrum of the donor overlaps the excitation spectrum of the acceptor.[51] In the AFP-based FRET strategy, CFP and YFP mutants have been favorably utilized as a FRET donor and an acceptor, respectively[52] (**Figure 1**). Engineering with two AFPs in combination with the receptor protein affords a sensor protein that responds to dynamic fluctuation of intracellular ligand concentration by a ratiometric fluorescence change. The feasibility of this strategy strongly depends on the magnitude of the structural change of the receptor. In the case of the receptor with a large structural change upon binding the substrate, this strategy would be the most straightforward way to integrate the signal transduction function into the receptor of interest, although serious concerns have been pointed out that the obtained FRET signals do not simply reflect the change in the expected distance of FRET pairs.[52-53] The trailblazing work for this sensor was reported by Miyawaki et al., in which a genetically encoded calcium indicator composed of two different colored AFP mutants and calmodulin, a Ca^{2+} receptor, has been constructed.[30] The receptor complex that accompanies the dissociation or the association of multiple subunits upon ligand binding was also suitable for the design of FRET-based biosensors (Figure 1b). Zaccoro and coworkers constructed a ratiometric fluorescent biosensor for cyclic adenosine monophosphate (cAMP) on the basis of an intermolecular FRET system between regulatory (R) and catalytic (C) subunit of protein kinase A (PKA).[41-42] This biosensor can detect the rise of intracellular cAMP concentration by the decrease in the FRET efficiency induced by the dissociation of C subunit from R subunit.

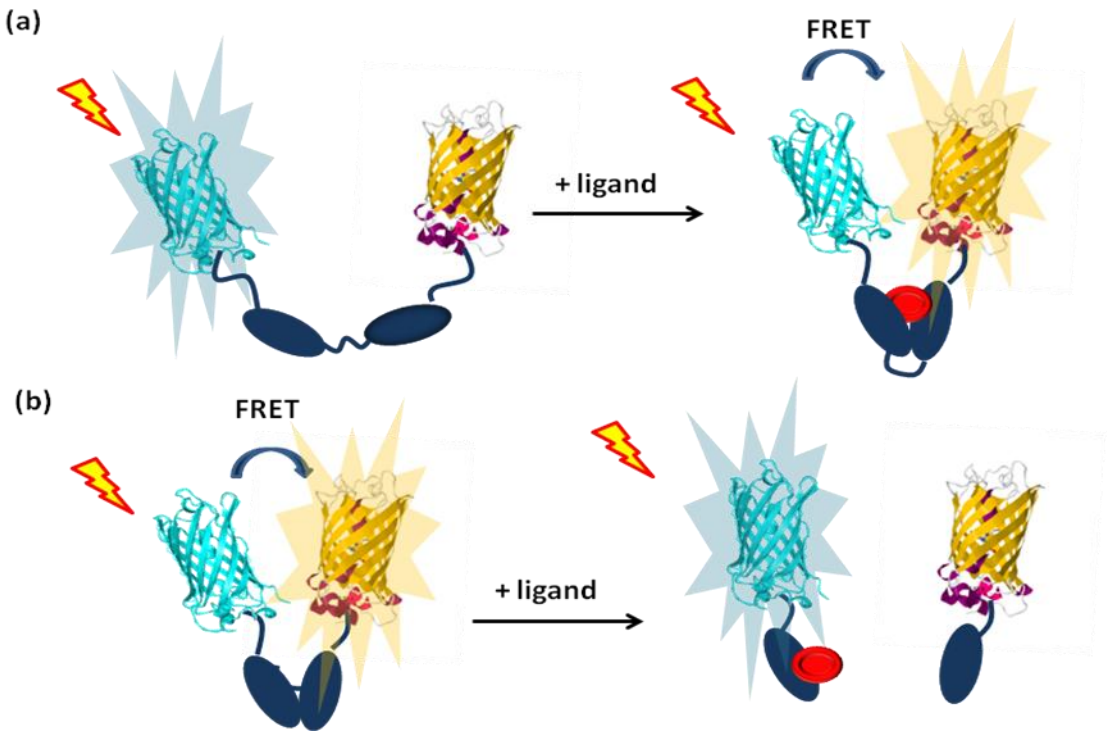


Figure 1: Schematic illustration shows a concept of ligand sensing by dual AFP-fused FRET-based biosensors. Currently, CFP and YFP mutants are preferentially selected as FRET donor and acceptor, respectively. (a) Intramolecular FRET-biosensors exploit the protein domains with a large structural change upon the ligand-binding event. (b) Intermolecular FRET-based biosensors accompany the dissociation or association of multiple subunits upon the ligand-binding event. The dissociation-type FRET-based biosensor is only depicted in this figure.

2.1.1.2. Single circularly permuted (cp) AFP-based biosensors

As an alternative strategy, non-FRET biosensors based on a single circularly permuted (cp) AFP have been developed for targeting Ca^{2+},[32-34,37] cGMP,[53] H_2O_2,[54,55] Zn^{2+},[56] and an inositol phosphate derivative.[50] cpAFPs were constructed by connecting original N and C termini by a short peptide linker and regenerating the novel N and C termini at a specific position. Although the dual AFP-fused FRET-based biosensor is the most facile and robust strategy among AFP-based biosensors, the application of the strategy for FRET-based sensors is essentially difficult in the case of a receptor that undergoes just a slight structural change upon binding to the substrate. In such a case, a sophisticated manipulation of the interplay between the small perturbation of the receptor conformation and the alteration of photochemical property of the AFP chromophore in the ligand-binding event must be required. In the cpAFP-based biosensors, the receptor protein could be placed at the residues near the chromophore that would critically affect the photochemical property of the chromophore in AFP. Morii and coworkers developed a novel cpAFP-based sensor for D-*myo*-inositol-1,3,4,5-tetrakisphosphate (Ins (1,3,4,5) P_4), from a newly designed split pleckstrin homology (PH) domain of Bruton's tyrosine kinase (Btk) and a single cpGFP[50] (**Figure 2**). The resulting split PH domain-cpGFP conjugate, Btk-cpGFP, exhibited bimodal absorption spectra corresponding to the protonated and deprotonated states of the chromophore in GFP. Interestingly, the Btk-cpGFP realized a ratiometric fluorescence detection of Ins (1,3,4,5) P_4 by the excitation of each distinct absorption band, and retained the ligand affinity and the selectivity of the original PH domain.

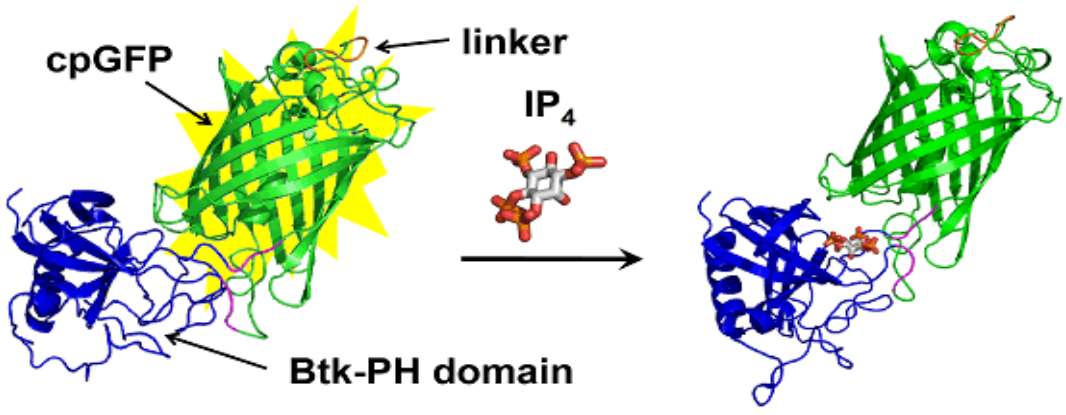


Figure 2: Schematic illustration shows a fluorescent biosensor for inositol tetrakisphospgate based on the split Btk PH domain-cpGFP conjugate.[50] The original N and C termini are linked with a short peptide linker (orange), and the novel terminal of cpGFP (purple) is fused to the split Btk PH domain (blue). The conformational change if the PH domain induced by the ligand-binding event was transduced to the structural perturbation at the chromophore of conjugated GFP, and then resulted in the ratiometric fluorescence change of cpGFP.

2.1.2. Chemically-modified protein based sensors

Another strategy for constructing fluorescent biosensors is a structure-based design of a protein-based biosensor covalently modified with a synthetic fluorophore. Genetically encoded biosensors using AFPs can be expressed basically in all types of cells, allowing straightforward visualization of intracellular target molecules. In contrast, these protein-based biosensors require the invasive technique for translocating across the plasma membrane, such as electroporation,[57-59] lipofection,[60.61] microinjection,[62] and tagging arginine-rich sequences.[63,64] However, this type of biosensors could be more advantageous in some respects. The relatively smaller size of the synthetic fluorophore is likely to contribute to the less perturbation of the property of the original receptor protein. Furthermore, since the amount of intracellular biosensors can be controlled easily compared to the genetic expression system, the interference to the molecular geography of the analyte in cells could be suppressed minimally. In the synthetic fluorophore-attached biosensors, the principle of the signal transduction mechanism is generally based on the alteration of microenvironment of the fluorophore during the ligand-binding event. Especially, polarity-sensitive fluorescent probes are the most widely utilized due to the abrupt change in hydrophobicity in the vicinity of the protein surface.[10] The central issue for the construction of these biosensors is the way to introduce the fluorophore into the receptor protein site-selectively. Current methodologies for the site-selective incorporation of synthetic fluorophores into protein are divided into three groups; (1) incorporation of a thiol reactive fluorophore by engineering of a mutant receptor with a unique cysteine residue, (2) a site-specific unnatural amino acid mutagenesis with an expanded genetic code, and (3) a post-photoaffinity labeling modification.

In the first method all the original cysteine residues must be initially substituted with other amino acids followed by the introduction of a unique cysteine residue at the specific position to avoid the non-specific labeling. The incorporation site of a fluorophore could be determined by the three-dimensional structure of the receptor protein obtained from crystallographic analysis. Morii and coworkers constructed novel biosensors for inositol 1,4,5-trisphosphate (Ins (1,4,5) P3) and 1,3,4,5-tetrakisphosphate (Ins (1,3,4,5) P4) by utilizing the pleckstrin homology (PH) domain of phospholipase C (PLC) δ_1[65] and the general receptor for phosphoinositides 1 (GRP1) (**Figure 3**), respectively. In these biosensors, a synthetic fluorophore was attached at the proximity of the ligand-binding site so that the changes in orientation of the fluorophore induced by the substrate binding leads to a sufficient fluorescence response. This structure-based design of synthetic fluorophore-modified biosensors is a

powerful method to produce biosensors with high selectivity and appropriate affinity to target inositol derivatives in living cells.[66,59,64]

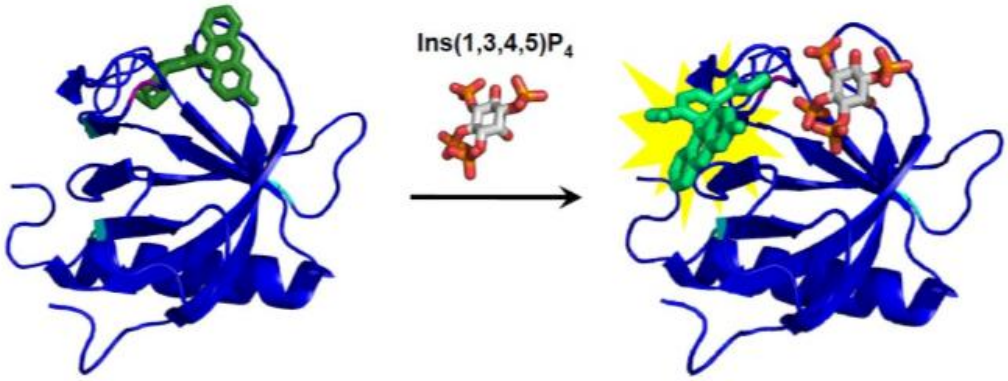


Figure 3: Schematic illustration shows a fluorescent biosensor for Ins (1.3.4,5) P_4 based on the GRP1 PH domain covalently modified with a fluorescein (green) as a reporter probe. The original cysteine residues(cyan) were replaced with other amino acids. The fluorophore was designed to be oriented near the binding pocket. The position labeled by the fluorescein at Glu82 is indicated in magenta. The local environmental change of the fluorophore induced by the ligand-binding event was transduced to the fluorescence change.

Unlike the post-labeling of unique cysteine residues, a mutagenesis technique for direct incorporation of synthetic fluorophores as unnatural amino acids into desired positions in proteins has been developed. Such a site-specific 0mutagenesis with an expanded genetic code that employed an amber suppression method[67,68] or a four-base codon method[69] in cell-free translation systems has provided a variety of fluorescently modified proteins.[70-72] Hohsaka and coworkers synthesized unnatural amino acids modified with BODIPY derivatives and incorporated two of them into different positions of calmodulin as a donor and acceptor pair for FRET using two four-base codons.[72] The doubly modified calmodulin exhibited a substantial FRET signal in response to the conformational change of calmodulins induced by the addition of the calmodulin binding peptide.

When a three dimensional structure of a receptor protein is not available, it is difficult to convert such a receptor into a fluorescent biosensor by applying the above mentioned methods. An approach enabling a site-specific incorporation of a signal transducer proximal to the binding pocket of intact protein, for which little or no structural information is available, is also highly desirable.

2.1.3. Signaling aptamers

Combinatorial chemistry with in vitro evolution, known as SELEX (systematic evolution of ligands by exponential enrichment), offers an effective strategy for generating DNA or RNA receptors (aptamers) with appropriate affinity and specificity for various targets ranging from small molecules to proteins.[73-77] Because most of the structurally characterized aptamers undergo induced-fit type conformational changes upon binding of their cognate ligands,[78] the introduction of signal transduction function can be accomplished by taking advantages of the ligand dependent change in the local environment around the attached fluorophores. Even if does not exist a report of a specific or an appropriate protein receptor for the substrate of interest, aptamers for the substrate can potentially be generated through in vitro selection. In this respect, the design of aptamer-based fluorescent sensors represents an attractive and promising alternative to the protein-based sensors. Additionally, development of the various types of fluorescence-signaling aptamers has been accelerated by the facile preparation of aptamers modified with synthetic fluorophores in a site-selective manner.

2.1.4. References

1) Tainaka, K., Sakaguchi, R., Hayashi, H., Nakano, S., Liew, F.F., and Morii T., *Sensors* **2010**, 10, 1355-1376.
2) De Silva, A. P., Gunaratne, H. Q., Gunnlaugsson, T., Huxley, A. J., McCoy, C. P., Rademacher, J. T., and Rice, T. E., *Chem. Rev.* **1997**, 97, 1515-1566.
3) Johnson, I., *Histochem. J.* **1998**, 30, 123-140.
4) Terai, T., and Nagano, T., *Curr. Opin. Chem. Biol.* **2008**, 12, 515-521.
5) Domaille, D. W., Que, E. L., and Chang, C. J., *Nat. Chem. Biol.* **2008**, 4, 168-175.
6) Cao, H., and Heagy, M. D., *J. Fluoresc.* **2004**, 14, 569-584.
7) Gomes, A., Fernandes, E., and Lima, J. L., *J. Fluoresc.* **2006**, 16, 119-139.
8) Soh, N., *Anal. Bioanal. Chem.* **2006**, 386, 532-543.
9) Nolan, E. M., and Lippard, S. J., *Acc. Chem. Res.* **2009**, 42, 193-203.
10) Wang, H., Nakata, E., and Hamachi, I., *Chem. Biochem.* **2009**, 10, 2560-2577.
11) Liu, J., Cao, Z., and Lu, Y., *Chem. Rev.* **2009**, 109, 1948-1998.
12) Shimomura, O., Johnson, F. H., and Saiga, Y., *J. Cell. Comp. Physiol.* **1962**, 59, 223-239.
13) Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C., *Science* **1994**,263, 802-805.
14) Lippincott-Schwartz, J., Snapp, E., and Kenworthy, A., *Nat. Rev. Mol. Cell. Biol.* **2001**, 2, 444-456.
15) Heim, R., Prasher, D. C., and Tsien, R. Y., *Proc. Natl. Acad. Sci. USA* **1994**, 91, 12501-12504.
16) Remington, S. J., *Curr. Opin. Struct. Biol* **2006**, 16, 714-721.
17) Zhang, J., Campell, R. E., Ting, A. Y., and Tsien, R. Y., *Nat. Rev. Mol. Cell Biol.* **2002**, 3, 906-918.
18) Kneen, M., Farinas, J., Li, Y., and Verkman, A. S., *Biophys. J.* **1998**, 74, 1591-1599.
19) Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G., and Tsien, R. Y., *Proc. Natl. Acad. Sci. USA* **1998**, 95, 6803-6808.
20) Miesenbock, G., De Angelis, D. A., and Rothman, J. E., *Nature* **1998**, 394, 192-195.
21) Wachter, R. M., Yarbrough, D., Kallio, K., and Remington, S. J., *J. Mol. Biol.* **2000**, 301, 157-171.
22) Wachter, R. M., and Remington, S. J., *Curr. Biol.* **1999**, 9, R628-629.
23) Jayaraman, S., Haggie, P., Watcher, R. M., Remington, S. J., and Verkman, A. S., *J. Biol. Chem.* **2000**, 275, 6047-6050.
24) Chapleau, R. R., Blomberg, R., Ford, P. C., and Sagermann, M., *Protein Sci.* **2008**, 17, 614-622.
25) Barondeau, D. P., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D., *J. Am. Chem. Soc.* **2002**, 124, 3522-3524.

26) Mank, M., and Griesbeck, O., *Chem. Rev.* **2008**, 108, 1550-1564.

27) Garaschuk, O., Griesbeck, O., and Konnerth, A., *Cell Calcium* **2007**, 42, 351-361.

28) Palmer, A. E., Jin, C., Reed, J. C., and Tsien, R. Y., *Proc. Natl. Acad. Sci. USA* **2004**, 101, 17404-17409.

29) Palmer, A. E., Giacomello, M., Kortemme, T., Hires, S. A., Lev-Ram, V., Baker, D., and Tsien, R. Y., *Chem. Biol.* **2006**, 13, 521-530.

30) Miyawaki, A., Llopis, J., Heim, N., McCaffery, J. M., Adams, J. A., Ikura, M., and Tsien, R. Y., *Nature* **1997**, 388, 882-887.

31) Mank, M., Reiff, D. F., Heim, N., Friedrich, M. W., Borst, A., and Griesbeck, O., *Biophys. J.* **2006**, 90, 1790-1796.

32) Nakai, J., Ohkura, M., and Imoto, K., *Nat. Biotechnol.* **2001**, 19, 137-141.

33) Souslova, E. A., Belousov, V. V., Lock, J. G., Stromblad, S., Kasparov, S., Bolshakov, A. P., Pinelis, V. G., Labas, Y. A., Lukyanov, S., Mayr, L. M., and Chudakov, D. M., *BMC Biotechnol.* **2007**, 7, 37.

34) Baird, G. S., Zacharias, D. A., and Tsien, R. Y., *Proc. Natl. Acad. Sci. USA* **1999**, 96, 11241-11246.

35) Romoser, V. A., Hinkle, P. M., and Persechini, A., *J. Biol. Chem.* **1997**, 272, 13270-13274.

36) Truong, K., Sawano, A., Mizuno, H., Hama, H., Tong, K. I., Mal, T. K., Miyawaki, A., and Ikura, M., *Nat. Struct. Biol.* **2001**, 8, 1069-1073.

37) Nagai, T., Sawano, A., Park, E. S., and Miyawaki, A., *Proc. Natl. Acad. Sci. USA* **2001**, 98, 3197-3202.

38) Imamura, H., Nhat, K. P., Togawa, H., Saito, K., Iino, R., Kato-Yamada, Y., Nagai, T., and Noji, H., *Proc. Natl. Acad. Sci. USA* **2009**, 106, 15651-15656.

39) DiPilato, L. M., Cheng, X., and Zhang, J., *Proc. Natl. Acad. Sci. USA* **2004**, 101, 16513-16518.

40) Nikolaev, V. O., Bunemann, M., Schmitteckert, E., Lohse, M. J., and Engelhardt, S., *Circ. Res.* **2006**, 99, 1084-1091.

41) Zaccolo, M., De Giorgi, F., Cho, C. Y., Feng, L., Knapp, T., Negulescu, P.A., Taylor, S. S., Tsien, R. Y., and Pozzan, T. A., *Nat. Cell Biol.* **2000**, 2, 25-29.

42) Zaccolo, M., and Pozzan, T., *Science* **2002**, 295, 1711-1715.

43) Sato, M., Nakajima, T., Goto, M., and Umezawa, Y., *Anal. Chem.* **2006**, 78, 8175-8182.

44) Honda, A., Adams, S. R., Sawyer, C. L., Lev-Ram, V., Tsien, R. Y., and Dostmann, W. R., *Natl. Acad. Sci. USA* **2001**, 98, 2437-2442.

45) Nikolaev, V. O., Gambaryan, S., and Lohse, M. J., *Nat. Methods* **2006**, 3, 23-25.

46) Kaper, T., Looger, L. L., Takanaga, H., Platten, M., Steiman, L., and Frommer, W. B., *PLoS Biol.* **2007**, 5, e257.

47) Okumono, S., Looger, L. L., Micheva, K. D., Reimer, R. J., Smith, S. J., and Frommer, W. B., *Proc. Natl. Acad. Sci. USA* **2005**, 102, 8740-8745.

48) Tsien, R. Y., *FEBS Lett.* **2005**, 579, 927-932.

49) Hirose, K., Kadowaki, S., Tanabe, M., Takeshima, H., and Iino, M., *Science* **1999**, 284, 1527-1530.

50) Sakaguchi, R., Endoh, T., Yamamoto, S., Tainaka, K., Sugimoto, K., Fujieda, N., Kiyonaka, S., Mori, Y., and Morii, T., *Bioorg. Med. Chem.* **2009,** 17,7381-7386.

51) Sapsford, K. E., Berti, L., and Medintz, I. L., *Angew. Chem. Int. Ed.* **2006**,45, 4562-4589.

52) Piston, D. W., and Kremers, G. J., *Trends Biochem. Sci.* **2007**, 32, 407-414.

53) Nausch, L. W., Ledoux, J., Bonev, A. D., Nelson, M. T., and Dostmann, W. R., *Proc. Natl. Acad. Sci. USA* **2008**, 105, 365-370.

54) Belousov, V. V., Fradkov, A. F., Lukyanov, K. A., Staroverov, D. B., Shakhbazov, K. S., Terskikh, A. V., and Lukyanov, S., *Nat. Methods* **2006**, 3, 281-286.

55) Dooley, C. T., Dore, T. M., Hanson, G. T., Jackson, W. C., Remington, S. J., and Tsien, R. Y., *J. Biol. Chem.* **2004**, 279, 22284-22293.

56) Mizuno, T., Murao, K., Tanabe, Y., Oda, and M., Tanaka, T., *J. Am. Chem. Soc.* **2007**, 129, 11378-11383.

57) Marrero, M. B., Schieffer, B., Paxton, W. G., Schieffer, E., and Bernstein, K. E., *J. Biol. Chem.* **1995**, 270, 15734-15738.

58) Fenton, M., Bone, N., and Sinclair, A. J., *J. Immunol. Methods* **1998**, 212, 41-48.

59) Sakaguchi, R., Tainaka, K., Shimada, N., Nakano, S., Inoue, M., Kiyonaka, S., Mori, Y., and Morii, T., *Angew. Chem. Int. Ed.* **2009**, 48.

60) Zelphati, O., Wang, Y., Kitada, S., Reed, J. C., Felgner, P. L., and Corbeil, J., *J. Biol. Chem.* **2001**, 276, 35103-35110.

61) Zheng, X., Lundberg, M., Karlsson, A., and Johasson, M., *Cancer Res.* **2003**, 63, 6909-6913.

62) Abarzua, P., LoSardo, J. E., Gubler, M. L., and Neri, A., *Cancer Res.* **1995**, 55, 3490-3494.

63) Wadia, J. S., and Dowdy, S. F., *Adv. Drug. Deliv. Rev.* **2005**, 57, 579-596.

64) Sugimoto, K., Nishida, M., Otsuka, M., Makino, K., Ohkubo, K., Mori, Y., and Morii, T., *Chem. Biol.* **2004**, 11, 475-485.

65) Morii, T., Sugimoto, K., Makino, K., Otsuka, M., Imoto, K., and Mori, Y., *J. Am. Chem. Soc.* **2002**, 124, 1138-1139.

66) Nishida, M., Sugimoto, K., Hara, Y., Mori, E., Morii, T., Kurosaki, T.,and Mori, Y., *EMBO J.* **2003**, 22, 4677-4688.
67) Wang, L., and Schultz, P. G., *Angew. Chem. Int. Ed.* **2005**, 44, 34-66.
68) Xie, J., and Schultz, P. G., *Nat. Rev. Mol. Cell. Biol.* **2006**, 7, 775-782.
69) Honsaka, T., and Sisido, M., *Curr. Opin. Chem. Biol.* **2002**, , 6, 809-815.
70) Anderson, R. D., III; Zhou, J., and Hecht, S. M., *J. Am. Chem. Soc.* **2002**, 124, 9674-9675.
71) Taki, M., Honsaka, T., Murakami, H., Taira, K., and Sisido, M., *J. Am. Chem. Biol.* **2002**, 124, 14586-14590.
72) Kajihara, D., Abe, R., Iijima, I., Komiyama, C., Sisido, M., and Honsaka, T., *Nat. Methods* **2006**, 3, 923-929.
73) Ellington, A. D., *Curr. Biol.* **1994**, 4, 427-429.
74) Ellington, A. D., and Szostak, J. W., *Nature* **1990**, 346, 818-822.
75) Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M., *Annu. Rev. Biochem.* **1995**, 64, 763-797.
76) Osborne, S. E., and Ellington, A. D., *Chem. Rev.* **1997**, 97, 349-370.
77) Wilson, D. S., and Szostak, J. W., *Annu. Rev. Biochem.* **1999**, 68, 611-647.
78) Westhof, E., and Patel, D. J., *Curr. Opin. Struct. Biol.* **1997**, 7, 305-309.

2.2. Structural aspects of RNP aptamers

In vitro selection provides one of the most powerful strategies for obtaining functional RNA molecules.[1-5] The selection and evolution technique has produced RNA aptamers that specifically bind small molecules[1-8] or ribozymes that catalyze a variety of chemical reactions[1-5,9,10] from randomized pools of oligo-ribonucleotides. Recent advances in the in vitro selection method have provided signaling nucleic acid aptamers[11-16] applicable for biosensors and allosteric ribozymes regulated by both ligands and proteins.[17-21] The key feature of RNA to perform such a variety of function has been attributed to the fact that a random pool of RNA provides a large number of different shapes that facilitate the chance to select RNA molecules with desired functions.[22] RNA performs a wide range of functions in conjugation with proteins in all cells. One of such examples is found in the three-dimensional structures of RNA-protein complexes, particularly in that of the ribosome.[23,24] The ribosome is the ribonucleoprotein that performs the remarkable task of selecting the amino acids to be added to the growing polypeptide chain by reading successive messenger RNA codons. It has been considered that RNA, not protein, is responsible for catalysis of the peptide transferase activity and that the ribosomal proteins stabilize and orient the RNA into a specific, active structure.[22,25] Such protein cofactors are required by many RNA enzymes, including RNase P and most group I and group II introns, for catalysis under physiological conditions.[26] The remarkable diversity of RNA-protein complexes suggests that its miniature analogue, a ribonucleopeptide, serves as an alternative framework for the design of tailor-made receptors and enzymes. Morii et al. developed a strategy that allows assembly of a ribonucleopeptide receptor by means of the structure-based design[27,28] and the in vitro selection method. On the basis of the Rev response element (RRE) RNA and HIV-1 Rev peptide complex,[29] an RNA subunit, referred to as RRENn, was designed consisting of two functionally separated regions, a possible ligand-binding region with 20 randomized nucleotides, successively extended to 40 nucleotides, and an adjacent stem region that serves as a binding site for the Rev peptide (**Figure 4**).

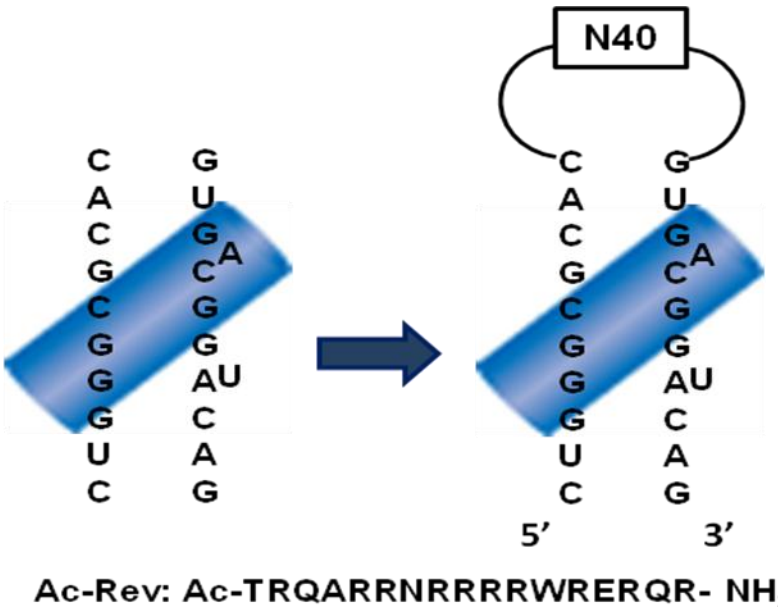


Figure 4: Schematic illustration of the structure of Rev-RRE complex (left) and Rev-RRENn ribonucleotide complex (right). The amino acid sequence of the Rev peptide (34-50, Ac-Rev) is shown at the bottom.

Gel mobility shift assay of the RRENn and the Rev peptide (Ac- Rev) confirmed an RNA-peptide complex formation, indicating that the introduction of the additional ligand binding region did not interfere with the specific complex formation between RRENn and the Rev peptide. Ribonucleopeptide receptors for ATP were isolated from a pool of RNA sequences (4^{20}) by the in vitro selection method (**Figure 5**).

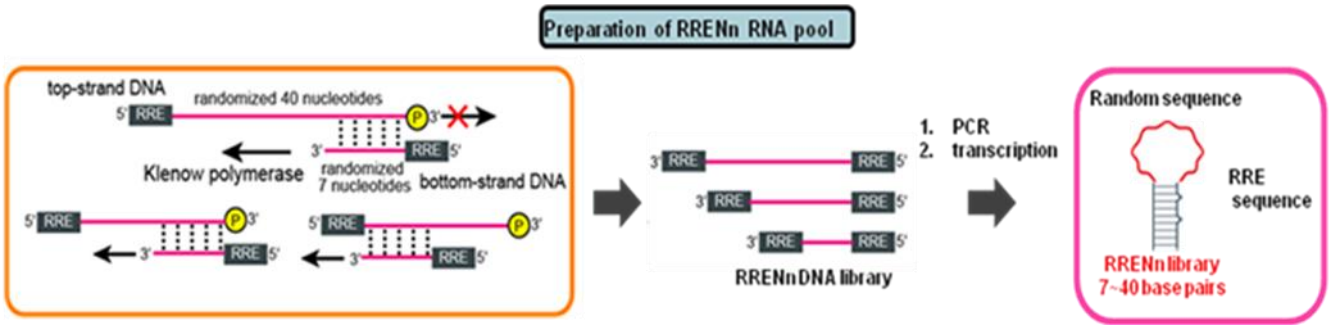


Figure 5: Schematic representation of the *in vitro* selection method applied to the preparation of RRENn RNA pool.

2.2.1. Rev peptide-RRE RNA recognition mode

RNA-protein interactions are central features of many fundamental biological processes, including translation, mRNA processing, and transcription. The HIV-1 Rev protein is an RNA-binding protein that regulates viral gene expression by affecting the relative amounts of spliced and un-spliced mRNAs that are exported to the cytoplasm. Rev mediates its function by binding to the RRE RNA located within the *env* gene of HIV.[30] Biochemical analyses have identified a high-affinity Rev-binding site in stem-loop IIB of the RRE (**Figure 6A**).[31] A short α-helical peptide, corresponding to the arginine-rich RNA-binding domain of Rev (**Figure 6B**), bind specifically to this small RNA element,[32] providing an effective scaffold for the construction of RNP receptors.

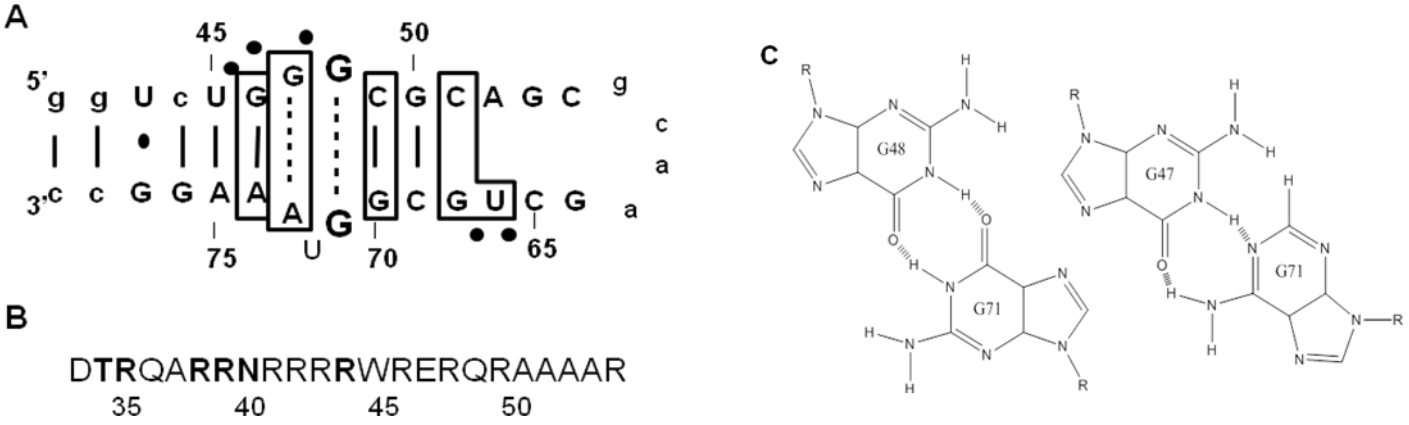


Figure 6: Minimal RRE-Rev complex. **A)** Sequence of RNA derived from stem-loop IIB of wild-type RRE. Boxed nucleotides were invariant in RNA selection studies of the Rev-RRE interaction in vitro, and the bold letters indicate a G-G base pair that was found to covary to an A-A base pair. Filled circles represent phosphates that interfere with Rev protein binding when ethylated. The numbering in the internal loop region is the same as for wild-type RRE, and non-wild-type nucleosides, in lowercase, are not numbered sequentially. **B)** Sequence of the Rev peptide, which contains amino acids 34 through 50 of the RNA-binding domain. Bold letters indicate amino acids that results in a > 10-fold decrease in binding specificity upon mutation to alanine. **C)** Hydrogen-bonding arrangement of two purine-purine base pairs in the internal loop of the RRE determined from previous NMR analyses.

The Rev peptide binds the major groove of the RNA in an α helical conformation, as indicated by circular dichroism studies.[32,33] DNA-binding proteins often use α helices to recognize specific bases in the major groove. However, the major groove of an A-form RNA helix is deep and too narrow to accommodate an α helix, thus requiring non-Watson-Crick interactions to widen the groove. In the Rev-RRE complex, two purine-purine base pairs locally open the major groove, a process which

appears to be facilitated by a pronounced distortion of the RNA backbone that results primarily from formation of the G48-G71 base pair in a locally parallel-stranded orientation.[34] This distortion results in an S-shaped architecture to the backbone from nucleotides G70 to A73, an undertwisting of the base pairs in the internal loop, and an opening of the major groove by ~5 Aº (**Figure 7**).

Figure 7: RNA major groove width and depth. Views down the α helix axis of Rev (blue) and GCN4 (green) with the major grooves of the RNA (or DNA) aligned with a standard A-form RNA duplex (red).

The GG base pair is important for Rev binding, yet it is not in direct contact with the peptide through hydrogen bonds, suggesting that it provides a critical structural feature for specific binding. The Rev α helix penetrates much more deeply into the major groove than is typical of DNA-binding proteins. In contrast to the shallow major groove of a B-form DNA helix, the RRE internal loop retains the characteristic deep groove of A-form RNA, and the Rev peptide must bind deep within the groove to allow side chains to reach the bases. In addition to binding deep within the groove, the Rev peptide has an extensive interface with the RNA, interacting over three to four turns of the α helix. Four amino acids make important base- specific contacts in the major groove (**Figure 8**). The residues Arg35 and Arg39 interact with nucleotides U66, G67, and G70 on one side of the groove, and Asn40 and Arg44 interact with nucleotides U45, G46, G47, and A73 on the opposite side of the groove. Almost all of these nucleotides were invariant in vitro selection experiments or showed chemical modification

interference of Rev binding.[35] These important nucleotides all have functional groups within reasonable hydrogen-bonding distance of these four amino acids; however, the precision of the side chain positions is insufficient to determine exact hydrogen bond arrangements for these contacts. Nevertheless, Asn40 is clearly coplanar with the purine-purine G47-A73 base pair and makes hydrogen bonds to groups in the major groove. All four of the amino acids that make potential base-specific contacts are critical for peptide-binding specificity,[32] and mutation of any of these three arginines to lysine results in a significant loss of binding activity in vivo, suggesting the existence of hydrogen bonds rather than simple electrostatic contacts.[33] In addition to the bases specific interactions, Thr34 and several arginines contact the phosphate backbone. In the average structure, the y-OH of Thr34 is a potential hydrogen bond donor to a phosphodiester oxygen ofG47 (**Figure 8A**). Mutation ofThr34 results in a >10-fold decrease in binding specificity, and chemical modification of the G47 phosphate reduces binding specificity.[32,35] Interestingly, Thr34 is also forming an "N-cap" structure at its NH_2-terminus that presumably stabilizes the α helix.[36] In addition to Thr, there are six arginines (Arg38, Arg4l, Arg42, Arg43, Arg46, and Arg48) that are positioned to make either hydrogen-bonding or simple electrostatic interactions with the phosphate backbone (**Figure 8B**).

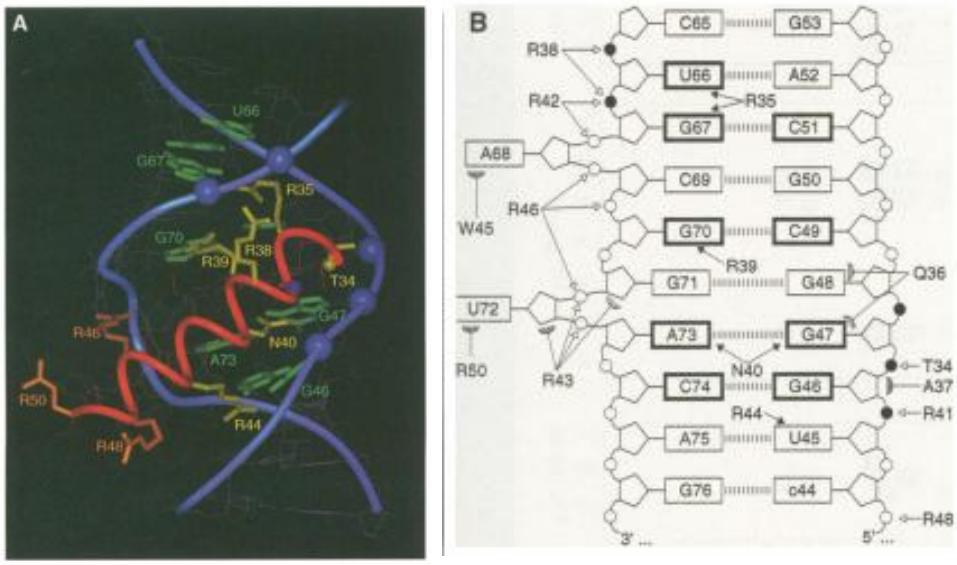


Figure 8: RNA-peptide contacts. **A)** View of the core binding site (nucleotides 44 to 53, 65 to 76, and amino acids 33 to 50). Amino acids that give > 10-fold or >3-fold decrease in binding specificity are shown in yellow and orange, respectively. The RNA is shown in blue, with the invariant nucleotides from in vitro selection studies colored green and phosphates that interfere with binding when ethylated as blue spheres. **B)** Schematic of specific RNA-peptide interactions. Black arrows, open arrows and hatched arcs indicate base-specific, phosphate backbone and van der Waals contacts, respectively. Thick boxes and black circles indicate important nucleotides and phosphates, respectively.

Arginine-38 is positioned near the phosphates of U66 and G67, which both give chemical modification interference,[35] and it is likely that Arg[38] makes bridging hydrogen bonds between the two ŋ-nitrogens and the phosphodiester oxygens of U66 and G67, because it cannot be functionally substituted by lysine.[33] Arginine-38 and Thr[34] make phosphate contacts flanking the region of the base-specific contacts and probably stabilize this region of the peptide. An especially arginine-rich face of the α helix makes phosphate contacts on the 3' side of the hairpin containing two single-nucleotide bulges, which brings many phosphates in close proximity. In particular, Arg[46] probably stabilizes the unfavorable electrostatic interactions near the U72 bulge, where the phosphorus atoms of G70 and U72 are -4 Å apart. Mutation of Arg[46] to alanine, but not to lysine, results in a sevenfold decrease in binding specificity, consistent with a specific electrostatic contact.[33] In addition, Arg[46], Arg[48], and Arg[50] at the COOH-terminus of the peptide, which mutational data suggest are moderately important for specific binding,[32] are making phosphate and Van der Waals contacts that may help orient the isolated α helix in the groove. The arginine-rich motif (ARM) is defined solely by a short region containing a high density of arginines. There are now two structural examples of short ARM peptides bound to RNA, bovine immunodeficiency virus Tat-TAR (trans-acting region) and HIV-1 Rev-RRE, and each peptide adopts a different secondary structure (β hairpin and α helix, respectively). Therefore, the ARM does not represent a unique structural motif.[37] Despite the differences, both peptides bind deeply in the major groove, and the high density of arginines characteristic of the ARM may be important for charge neutralization as the peptide penetrates the deep groove and forms a large interface. This feature probably also allows these 10- to 20-amino acid ARM peptides to bind as structurally independent elements, which is another characteristic feature of the ARM. The distinctive RNA structures formed by the binding sites undoubtedly also contribute to the folding and binding of these small ARM peptides. [38]

2.2.2. SELEX, Systematic Evolution of Ligands by Exponential Enrichment

The probability that a given sequence polynucleotide or peptide chain will fold to form a stable three-dimensional structure with a given ligand binding is unknown, but is generally thought to be very low. A related problem, which has been equally difficult to address theoretically, is to estimate the number of fundamentally different classes of structures capable of carrying out a given binding or catalytic function. These questions are important for theories of the origin and early evolution of life, as the first biological catalysts presumably arose from random sequence polymers.[39-41] Szostak et al. addressed these questions by developing methods for the synthesis of large numbers of random sequence RNA molecules, and for the isolation of molecules with specific ligand binding properties from such populations.[42]

Systematic Evolution of Ligands by Exponential Enrichment (SELEX), also referred to as *in vitro selection* or *in vitro evolution*, is a powerful process to select and isolate receptors for target recognition from a combinatorial library. An *in vitro* selection experiment comprises a number of sequential steps, the first of which is the generation of a nucleic acid library of random sequences. This starting pool of mainly nonfunctional RNA or DNA sequences is generated using a standard DNA-oligonucleotide synthesizer.[43] The design of such libraries involves the synthesis of a short defined sequence, followed by a random region of variable length and another defined sequence at the 5'-end. The length of the random region is normally between 20 to 40 basepair, which creates a library with a large number of random sequences (10^{15} to 10^{16}). This pool of synthetic single-stranded DNA is amplified in the polymerase chain reaction (PCR). Generating several copies of each DNA in its double-stranded form. By *in vitro* transcription, a corresponding library of RNA molecules can be generated which can then be used for the *in vitro* selection. If the transcription reaction contains nucleoside triphosphate derivatives that are chemically modified but still are substrates for RNA polymerases, libraries of modified RNAs can be generated in which sequences are equipped with a broad variety of additional chemical functionalities, normally not present in natural nucleic acids.[44-46] Aptamers selected from chemically modified libraries can, in some cases, be completely resistant toward degradation by nucleases. The additional functional groups can lead to ligands with novel physical and chemical properties or can provide additional handles to be utilized for functional improvement. Alternatively, mixtures of ssDNA or chemically modified ssDNAs can be generated by omitting the transcription step in the SELEX procedure. The selection of DNA aptamers adds the challenge to separate both strands in order to obtain libraries of ssDNA.[47] DNA aptamers are slightly more stable toward nuclease digestion

but seem to be less competent to fold into ligand-binding scaffolds. Moreover, RNA aptamers have the advantage of being actively transcribed *in vivo* from suitable templates, whereas DNA aptamers have to be introduced externally. By using combinatorial nucleic acid synthesis, remarkably complex libraries comprising up to 10^{16} different RNA, ssDNA, modified RNA, or modified ssDNA sequences can be produced at once. This number exceeds the diversity of antibodies raised by the immune system by several orders of magnitude. From this huge diversity, the challenge is to select active molecules from oligonucleotide libraries of such complexity. The high pool complexity ensures the presence of oligonucleotide-structures that are complementary to virtually any shape. Thus, if the pool is incubated with the immobilized ligand, binding sequences will tightly attach to it, so that nonbinding ones can easily be removed. The bound material can be subsequently collected, amplified and used to perform the next selection cycle. Depending on the nature of the ligand, this procedure has to be repeated until the desired ligand-specific aptamers have been enriched (**Figure 9**).

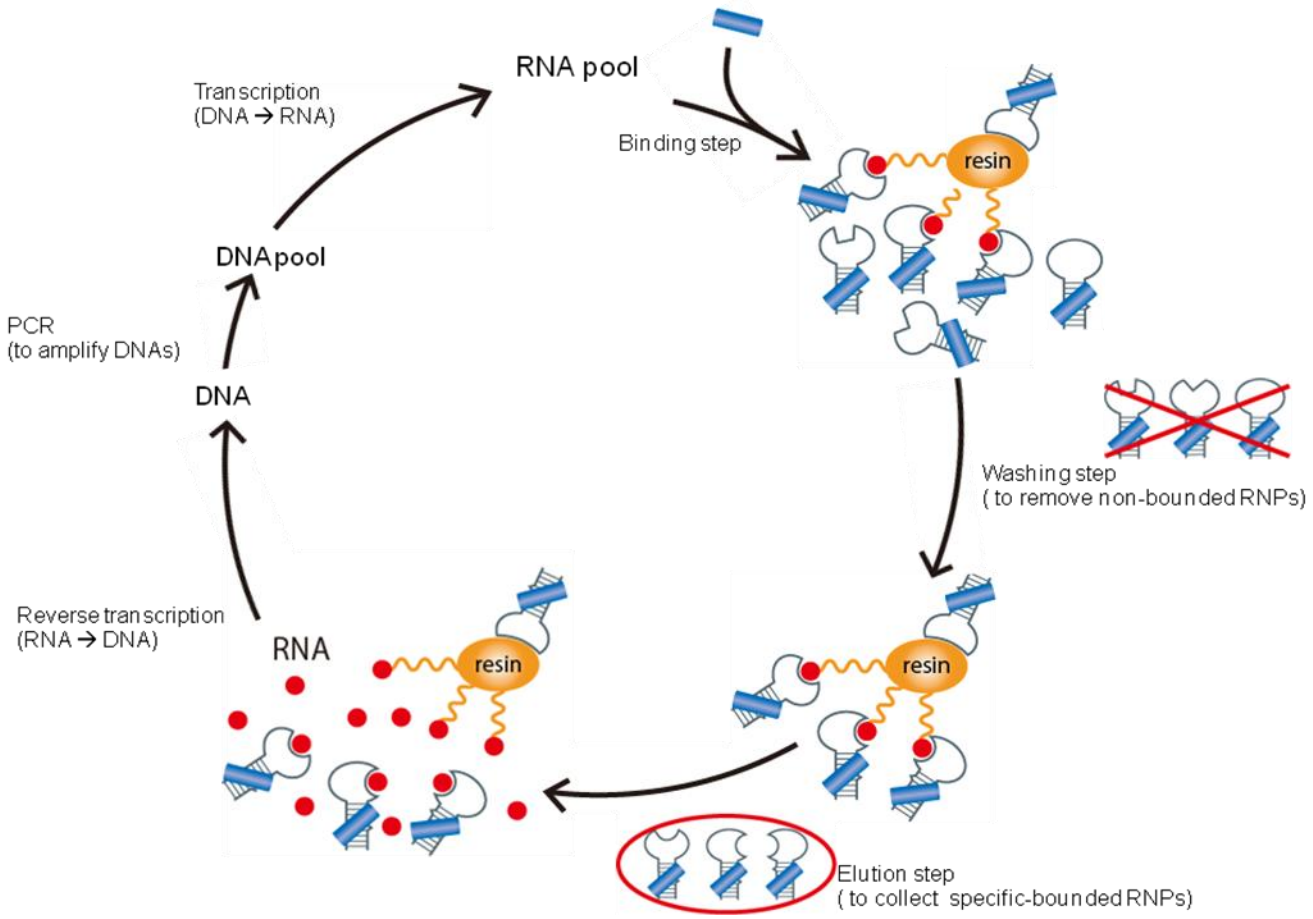


Figure 9: Schematic representation of RNA receptor selection using *in vitro* selection strategy. RNA sequences that have specific recognition to target cells are evolved to enrich the selection pools. The enriched pools are cloned and the positive clones are sequenced to determine those RNA receptors.

Particular aptamers can then be obtained from the enriched library by cloning and their sequence can be elucidated by standard sequencing protocols. Besides the possibility to perform the selection itself with modified nucleic acid sequences, which depends on whether the replicating enzymes accept the modified templates and monomer substrates, individual aptamers can also be equipped in a modular manner with additional functions[48-50] and therefore can be specifically tailored for many potential applications in biotechnology, molecular medicine and molecular biology.

2.2.3. Preliminary data

2.2.3.1. Aptamers selection and functional evaluation

Application of the selection and evolution technique is not limited to obtain functional macromolecules solely composed of RNA or DNA. It can also be used on ribonucleopeptides (RNP), such as complexes of the Rev peptide and RRE (Rev Responsive Element) RNA (**Figure 10**), which have a known three-dimensional structure and the interaction mode is described in the previous paragraph. Comparison of the characteristics of the RNA aptamers and RNPs targeting the same ligand provides an opportunity to understand the structural features that influence the fundamental molecular recognition mechanisms. To establish a proof of principle for the RNP-based selection strategy, Morii et al. initially chose ATP as a target for selection from the RNP library[51,52] and compared the characteristics of this ATP-binding RNP with those of the reported ATP-binding RNA aptamer[53] and other RNA aptamers targeting the ATP analogues cAMP,[54] FAD,[55] NAD^+ and S-adenosyl methionine (SAM).[56] These aptamers share quite similar consensus sequences for recognizing the adenosine moiety. The isolated ATP-binding RNPs,[57] in contrast, reveal consensus sequence that is different from the ones already reported for the ATP-binding aptamers. Given the fact that the above mentioned RNA aptamers are all selected against the ATP analogues by using the target ligand tethered through the C-8 position of the adenine ring, ATP-binding RNPs obtained by the selection against ATP immobilized through the 2' or 3'-OH group of ribose may reveal a different consensus sequence because in vitro selection would afford diverse

binding structures that vary significantly in accordance with the nature of ligands.[5,58,59] In fact, Burke e Gold have suggested that an entirely different RNA aptamer would be expected for the ATP-analogues if the adenosine has been linked to the resin trough a position other than the C-8.[56] It has been postulated that RNA aptamers obtained by in vitro selection from random sequence pools tend to be the simplest, most abundant structures for a given ligand.[60] RNA aptamers can specifically bind to the adenine ring of ATP by A:U Watson-Crick and Hoogsteen base pairing, or the triplex type U:A:U formation by both Watson-Crick and Hoogsteen base paring. However, structural analyses of the Sassanfar aptamer-AMP complex[61,62] revealed that the aptamer did not realize the base selectivity by simple Watson-Crick or Hoogsteen base pairing. In this chapter, we described the isolation of ATP-binding RNP, in which the RNA consensus sequence is different from those reported for RNA aptamers targeting the ATP analogues and have evaluated their affinity and selectivity for the ATP analogues.

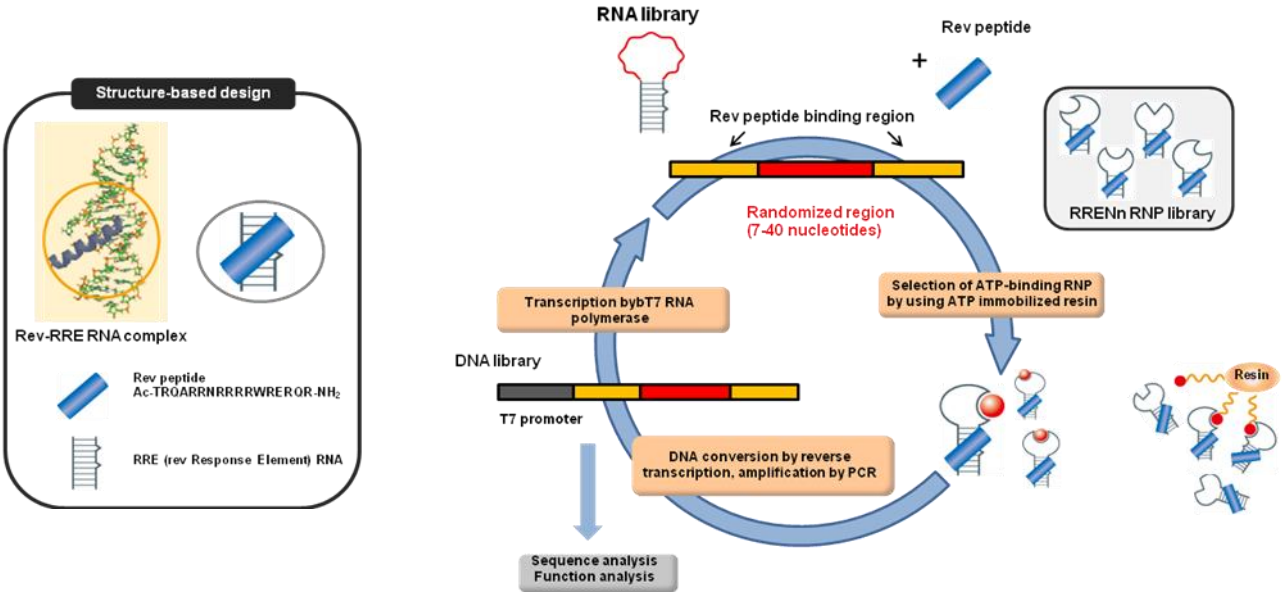


Figure 10: In vitro selection method for RNP receptors for ATP. Rev peptide and RRE RNA complex of which the three-dimensional structure and the interaction mode of each subunit have been established was utilized as the stable scaffold of RNP receptors.RNP library was constructed by introducing the randomized sequence adjacent to RRE sequence and complex with the Rev peptide. ATP-binding RNP were isolated by using the ATP-immobilized agarose resin and amplify by PCR after conversion to RNA-DNA duplex by reverse transcription to construct the DNA library for the next round. Selection and amplification was repeated until the saturation of the binding ratio to the substrate ATP on the resin. Then the sequence and affinity of RNP receptors were evaluated.

2.2.3.2. In *vitro* selection of ATP-binding RNP receptors from Nn library

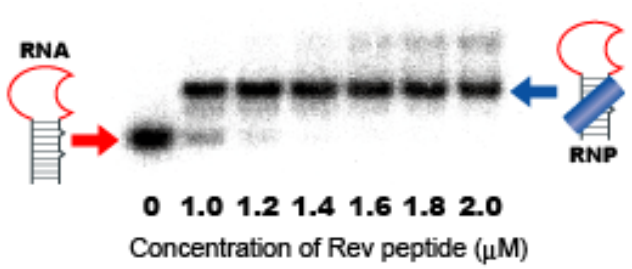
ATP-binding RNP receptors have been isolated from the RREN30 RNP library,[56] in which RNA has a randomized region with a fixed length of 30 nucleotides (**Table 1**). Because only a limited number of nucleotides would participate in ATP-binding, these RNPs contain nucleotides that are not directly involved in formation of the binding pocket. Therefore, an RRENn RNP library[51] that possesses 7 to 40 nucleotides in the randomized region was used to select ATP-binding RNP receptors with less than 30 nucleotides in the ATP-binding region. From this library it is thus possible to obtain ATP-binding RNP with the minimum of nucleotides for ATP binding. ATP attached through the ribose 2'- or 3'-hydroxyl groups via a 22 atom spacer to agarose beads was used for the selection. After the selection from the RRENn RNP library, RNAs with the randomized region of various lengths from 17 to 34 nucleotides were found (**Table 2**). Comparison of the nucleotide sequence of each RNP revealed a consensus sequence 5'-GUAGUGG---UGUGUGUG-3' that was the same as the one found in ATP-binding RNP selected from the RREN30 RNP library[56] (**Table 1** and **Table 2**). Thus, RNPs selected from the RRENn RNP library are likely to share a similar ATP biding mode with the RNPs obtained from the RREN30 RNP library. This consensus sequence differs from those reported for the ATP aptamer,[52] cAMP aptamer,[53] adenine aptamer,[63] adenine riboswitch,[64] and S-adenosyl methionine responsive riboswitch.[65] The nucleotides in the non-consensus region between the 5'-half and the 3'-half of the consensus sequence are different for each RNP. Furthermore, there is a variable region connecting the RRE and the consensus region, which often contains complementary nucleotides expected to form base pairs. The affinity of RNP selected from the RRENn RNP library for the substrate ATP was analyzed by converting the RNP receptor to the fluorescent RNP sensor by utilizing 7-methoxycoumarine-labeled Rev (7mC-Rev) for the RNP complex formation.[51,56] Titration of the fluorescent RNP with increasing concentration of ATP gave the titration curve for the complex formation of RNP and ATP, from which the equilibrium dissociation constants (K_D) for the complex of ATP and RNP could be deduced. The K_D values for the ATP binding ranged from 0.2 to 15μM and are listed in **Table 2**. For the further structural study, we focused on An16RNP from the RRENn RNP library because its RNA subunit An16 had a relatively small ATP-binding site of 19 nucleotides with the 5'-GUAGUGG---UGUGUGUG-3' consensus sequence and it showed a relatively high affinity to ATP (K_D=1.1μM). Formation of the RNP complex by An16RNA and the Rev peptide was verified by titration of An16RNA with the Rev peptide. A gel mobility shift assay showed that An16RNA (1μM) formed a 1:1 RNP complex (87%) in the presence of 1.2 equivalent of Rev (**Figure 11**).

RNA	Variable region	Consensus region	Non-consensus region	Consensus region	Variable region	k_D (μM)
A15	GCAGU	**GUAGUGG**	UU	**UGUGUGUG**	AUUGCUUA	*8.8*
A35	UCUUCU	**GUAGUGG**	UU	**UGUGUGUG**	AGUUGUG	*16*
A30	UAUACC	**GUAGUGG**	UU	**UGUGUGUG**	GGUUGG	*5.4*
A14	UACUGC	**GUAGUGG**	UU	**UGUGUGUG**	GCGGUUU	*8.8*
A16	UUGGCAC	**GUAGUGG**	UU	**UGUGUGUG**	GUGUUU	*34*
A23	UUAGAC	**GUAGUGG**	UUU	**UGUGUGUG**	GUCUGC	*20*
A01	CCUUC	**GUAGUGG**	UGU	**UGUGUGUG**	GUUGUUG	*7.7*
A26	UUCCG	**GUAGUGG**	UUG	**UGUGUGUG**	CGGUUUU	*156*
A32	UUGUAU	**GUAGUGG**	AUA	**UGUGUGUG**	AUGCCCG	*4.8*
A02	UUGCAU	**GUAGUGG**	GUA	**UGUGUGUG**	AUGUAU	*2.0*
A17	UUAGAU	**GUAGUGG**	GUA	**GUGUGUG**	UUUUCUG	*12*
A25	UGCUG	**GUAGUGG**	GUA	**UGUGUGUG**	CGGCAUU	*1.5*
A21	UUUGC	**GUAGUGG**	UUUU	**UGUGUGUG**	GCUGUA	*6.0*
A31	CAGAUU	**GUAGUGG**	CUUU	**UGUGUGUG**	AAUC	*>1000*
A34	U	**GUAGUGG**	UGCCUGUGAUGGC	**UGUGUGUG**	A	*5.8*

Table 1: Nucleotide sequences of ATP-binding RNP receptors isolated from RREN30 RNP library. The consensus sequence is indicated in bold character. The irregular nucleotides in the variable region are shaded in gray. The equilibrium dissociation constant (K_D) of each RNP was obtained from the fluorescence titration of RNP complexed with 7-methoxy coumarin-labeled Rev (7mC-Rev) by ATP at 20ºC.

RNA	RRE region	Variable region	Consensus region		Consensus region	Variable region	RRE region	Length	k_D (μM)
	10		20		30		40		
An16	GGUCUGGGCGCA	C	**GUAGUGG**	UG	**UGUGUGUG**	G	UGACGGUACAGGCC	19nt	*1.1*
An33	----------	C	**GUAGUGG**	AA	**UGUGUGUG**	G	----------	19nt	*6.6*
An15	----------	AU	**GUAGUGG**	UU	**UGUGUGUG**	AU	----------	21nt	*<1.0*
An29	----------	GCU	**GUAGUGG**	UU	**UGAGUGUG**	AGU	----------	23nt	*<1.0*
An18	----------	AC	**GUAGUGG**	GU	**UGAGUGUG**	GU	----------	21nt	*<1.0*
An34	----------	AC	**UUAGUGG**	UU	**UGUGUGUG**		----------	21nt	*1.5*
An17	----------		**UUAGUGG**	UU	**UGUGUGUG**		----------	17nt	*6.2*
An19			**UUAGUGG**	UUU	**UGUGUGUG**			18nt	*9.4*
An10	----------		**UUAGUGG**	UUU	**UGUGUGUG**	U	----------	20nt	*15*
An09	----------		**GAAGUGG**	UUU	**UGAGUGUG**	GU	----------	22nt	*1.4*
An01	----------	AAGCUG	**GUAGUGG**	GUA	**AGCGUGUG**	ACAGCAAGAU	----------	34nt	*4.7*

Table 2: Nucleotide sequences of ATP-binding RNP receptors isolated from the RRENn RNP library. The consensus region is indicated in bold with the irregular nucleotides marked in pink. The non-consensus region is boxed in red. Nucleotides expected to form base pairs in the variable region are shaded in grey. The equilibrium dissociation constant (K_D) for the ATP complex of each RNP was obtained from the fluorescence spectral changes of each RNP complexed with 7-methoxy coumarin-labeled Rev (7mC-Rev) by titration with ATP at 4ºC.



An16 RNA 1 μM Rev peptide (μM)	0	1	1.2	1.4	1.6	1.8	2
Non-specific (%)	0	0	0	13	19	24	28
RNP (%)	0	73	87	80	76	76	72
RNA (%)	100	27	13	7	5	0	0

Figure 11: The autoradiogram of the gel shift assay shows that An16RNA forms a 1:1 complex with the Rev peptide in a buffer containing 10mM Tris-HCl (PH 7.6), 100mM NaCl, 10mM $MgCl_2$, 0.005% Tween 20, and 6% sucrose at 4ºC. Peptide concentrations added to An16RNA (1μM) are shown under each lane (0, 1.0, 1.2, 1.4, 1.6, 1.8 and 2.0μM). The table shows the percentage of band intensity for free RNA, the 1:1 RNP complex (RNP) and RNP complexes with more than one equivalent of Rev (Non-specific).

2.2.3.3. Functional evaluation of An16 ATP-binding RNP receptor

Information on the affinity and selectivity of An16RNP for its target can be obtained from the comparison of the K_D values of An16RNP with various ligands and correlating the change in affinity with the structural differences of ligands. An16RNP revealed low affinities for UTP, CTP, or GTP, but high affinities for AMP, ADP, dATP, adenosine (Ado) and adenine that were similar as for ATP (**Figure 12** an **13**).

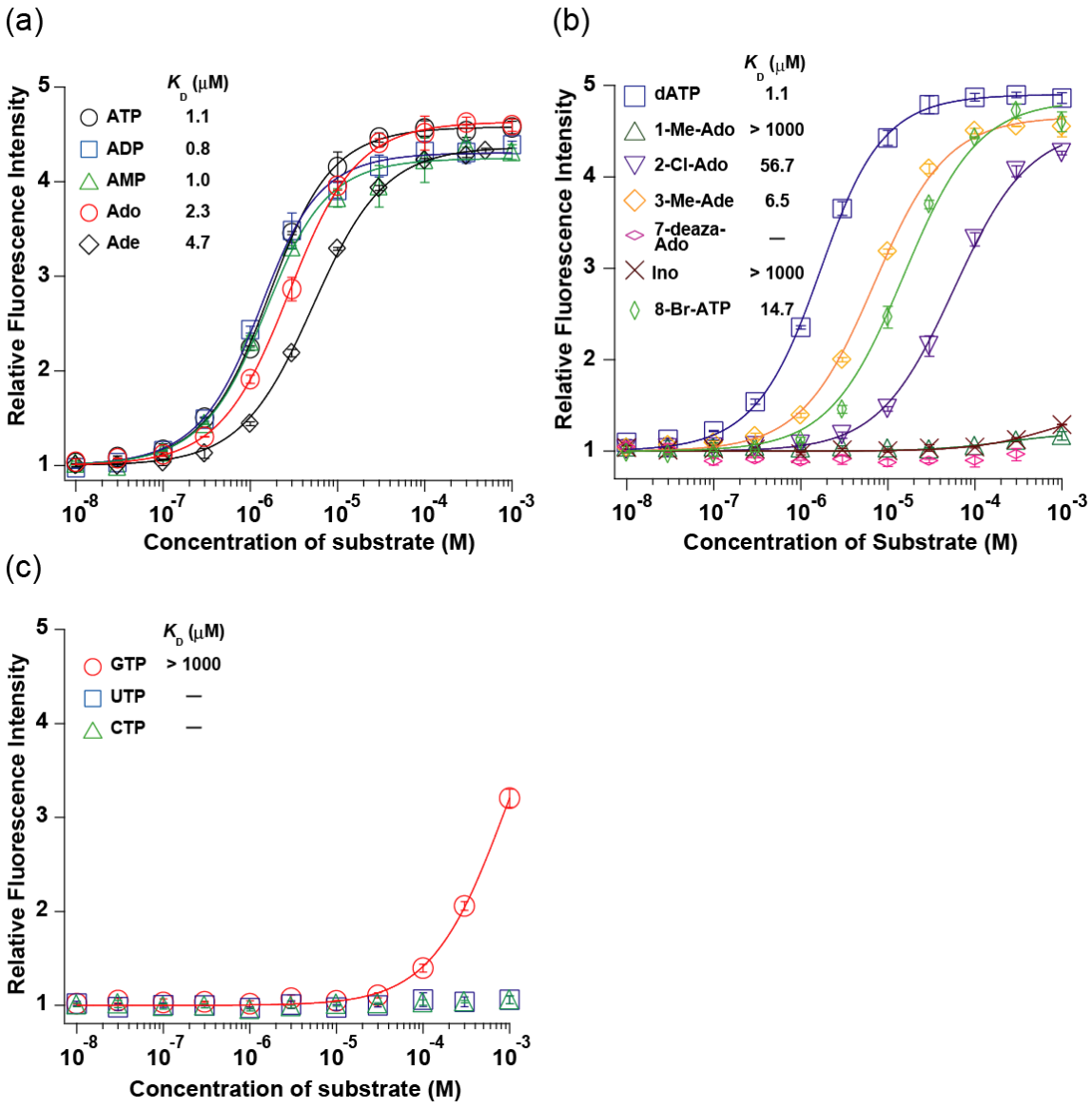


Figure 12: Saturation curves for the relative fluorescence intensity changes of An16RNP complexed with 7mC-Rev by titration with (a) ATP, ADP, AMP, Ado and Ade, with (b) dATP, 1-Me-Ado, 2-Cl-Ado, 3-Me-Ade, 7-deaza-Ado, Ino, and 8-Br-ATP, and with (c) GTP, UTP, and CTP at 4ºC indicate that an16RNP responds sensitively to the adenine base but moderately to the phosphate group and the ribose moiety of the ATP analogues.

This indicates that the adenine ring dominantly contributes in the complex formation. Because the affinities for Ado and adenine were slightly lower than for ATP, An16RNP could interact with the α-phosphate group to a certain extent. As An16RNP barely showed affinity to the adenosine analogues 1-methyl-adenosine (1-Me-Ado), inosine (Ino) and 7-deaza-adenosine (7-deaza-Ado), N-1 and N-7 atoms and the 6-amino group are essential for An16RNP to bind the adenine ring (**Figure 13**).

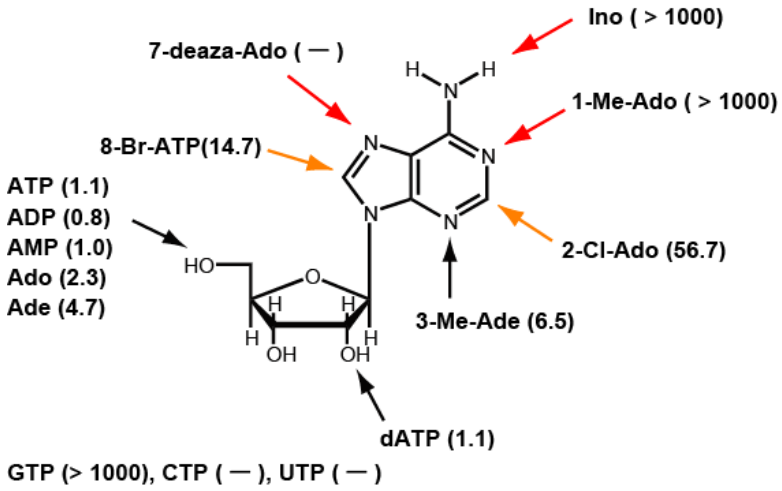


Figure 13: A schematic illustration summarizes the affinity and the selectivity of the ATP-binding An16RNP for ATP analogues. Equilibrium dissociation constants of An16RNP for ATP, ADP, AMP, Ado, adenine, 1-Me-Ado, 2-Cl-Ado, 3-Me-Ade, 7-deaza-Ado, Ino, 8-Br-ATP, dATP, GTP, CTP and UTP obtained by the fluorescence titration of An16RNP complexed with 7mC-Rev with the indicated ligand are shown in between brackets (μM).

The affinity for 3-methyl-adenine (3-Me-Ade), on the other hand, is almost similar as for adenine, indicating that An16RNP does not contact the adenine N-3 position. The lower affinity for 2-chloro-adenosine (2-Cl-Ado) is possibly caused by a steric hindrance or an electrostatic repulsion. The ATP aptamer selected by Sassanfar *et al.*,[52] mainly recognizes the N-1, N-3 and the 6-amino group by the direct interaction of a guanine base of the aptamer with the adenine base of the substrate ATP.[60-62] The cyclic-AMP aptamer selected by Koizumi *et al.*,[53] interacts with N-1, N-7 and the 6-amino group of the adenine ring, which are also the determinants for the ATP complex formation of An16RNP. Additionally, the cAMP aptamer is sensitive to the nucleotide conformation and prefers to bind cAMP in the *anti* conformation of the adenine ring. The affinity of An16RNP for the 8-bromo-ATP (8-Br-ATP) was found to be lower than for ATP. Because 8-Br-ATP preferentially takes the *syn* conformation when free in solution,[66] the observed lower affinity is possibly caused by the unfavorable

energetic cost to change the conformation of the substrate from *syn* to *anti*. Alternatively, An16RNP might bind 8-Br-ATP in its *syn* conformation with reduced complex stability due to the steric hindrance at the ribose moiety. These results indicated that An16RNP recognizes mainly the adenine base moiety and interacts less with the phosphate group and the ribose moiety (**Figure 13**). An16RNP had exquisite properties as ATP-binding receptor and revealed nearly quantitative formation of the RNP complex with the Rev peptide. An16RNP showed higher affinity for ATP than An16RNA alone[50] (**Figure 14**). These characteristics made An16RNP as an ideal model to study the structure of the ATP-binding RNP receptor.

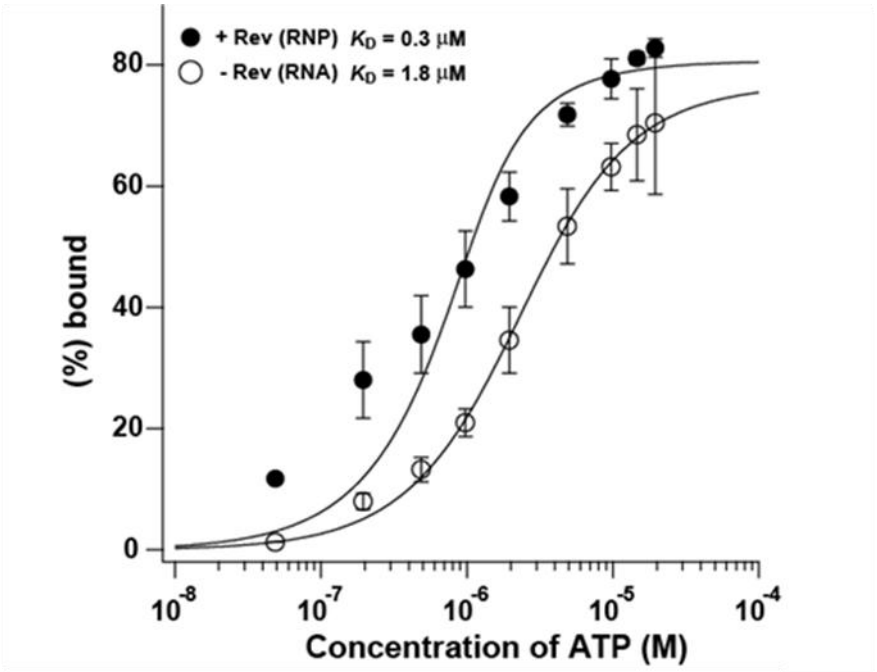


Figure 14 Titration curves show the ATP-binding of 5'-^{32}P-labeled An16RNA in the absence (open circles) or presence (closed circles) of the Rev peptide in a buffer containing 10mM Tris-HCl (pH 7.60), 100mM NaCl, 10mM $MgCl_2$, and 0.005% Tween 20 at 4ºC.

2.2.4. References

1) Ellington, A. D., and Szostak, J. W., *Nature* **1990**, 346, 818-822.
2) Ellington, A. D., *Curr. Biol.* **1994**, 4, 427-429.
3) Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M., *Annu. Rev. Biochem.* **1995**, 64, 763-797.
4) Osborne, S. E., and Ellington, A. D., *Chem. Rev.* **1997**, 97, 349-370.
5) Wilson, D.S., and Szostak, J. W., *Annu. Rev. Biochem.* **1999**, 68, 611-647.
6) Sassanfar, M., Szostak, J. W., *Nature* **1993**, 364, 550-553.
7) Burgstaller, P., Famulok, M., *Angew. Chem., Int. Ed. Engl.* **1994**, 33, 1084-1087.
8) Burke, D. H., Gold, L., *Nucleic Acids Res.* **1997**, 25, 2020-2024.
9) Bartel, D. P., Szostak, J. W., *Science* **1993**, 261, 1411-1418.
10) Prudent, J. R., Uno, T., Schultz, P. G., *Science* **1994**, 264, 1924-1927.
11) Potyrailo, R. A., Conrad, R. C., Ellington, A. D., Hieftje, G. M., *Anal. Chem.* **1998**, 70, 3419-3425.
12) Jhaveri, S., Rajendran, M., Ellington, A. D., *Nat. Biotechnol.* **2000**, 18, 1293-1297.
13) Brody, E. N., Gold, L., *J. Biotechnol.* **2000**, 74, 5-13.
14) Yamamoto, R., Baba, T., Kumar, P. K., *Genes Cells* **2000**, 5, 389-396.
15) Lee, M., Walt, D. R., *Anal. Biochem.* **2000**, 282, 142-146.
16) Hamaguchi, N., Ellington, A. D., Stanton, M., *Anal. Biochem.* **2001**, 294, 126-131.
17) Tang, J., Breaker, R. R., *Chem. Biol.* **1997**, 4, 453-459.
18) Tang, J., Breaker, R. R.., *Nucleic Acids Res.* **1998**, 26, 4214-4221.
19) Araki, M., Okuno, Y., Hara, Y., Sugiura, Y., *Nucleic Acids Res.* **1998**, 26, 3379-3384.
20) Koizumi, M., Soukup, G. A., Kerr, J. N. Q., Breaker, R. R., *Nat. Struct. Biol.* **1999**, 6, 1062-1071.
21) Robertson, M. P., Ellington, A. D., *Nucleic Acids Res.* **2000**, 28, 1751-1759.
22) Fontana, W., Konings, D. A. M., Stadler, P. F., Schuster, P., *Biopolymers* **1993**, 33, 1389-1404.
23) Ban, N., Nissen, P., Hansen, J., Moore, P. B., Steitz, T. A., *Science* **2000**, 289, 905-920.
24) Nissen, P., Hansen, J., Ban, N., Moore, P. B., Steitz, T. A., *Science* **2000**, 289, 920-930.
25) Puglisi, J. D., Blanchard, S. C., Green, R., *Nat. Struct. Biol.* **2000**, 7, 855-861.
26) Caprara, M. G., Nilsen, T. W., *Nat. Struct. Biol.* **2000**, 7, 831-833.
27) Pomerantz, J. L., Sharp, P. A., Pabo, C. O., *Science* **1995**, 267, 93-96.
28) McColl, D. J., Honchell, C. D., Frankel, A. D., *Proc. Natl. Acad. Sci. U.S.A* **1999**, 96, 9521-9526.
29) Battiste, J. L., Mao, H., Rao, N., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., Williamson, J. R., *Science* **1996**, 273, 1547-1551.

30) Gibson, J. J., *Perception of the Visual World* **1950** (Houghton Miffin, Boston).
31) Warren, W. H., Hannon, D. J., *Nature* **1988**, 336, 162.
32) Royden, C. S., Banks, M. S., Crowell, J. A., *ibid.* **1992**, 360, 583.
33) Judge, S. J., *Vision Res.* **1980**, 20, 535.
34) Sakata, H., Shibutani, H., Kawano, K., *J. Neurophysiol.* **1983**, 49, 1364.
35) Longuet-Higgins, H. C., Prazdny, K., *Proc. R. Soc. London Ser. B* **1980**, 208, 385.
36) Warren, W. H., *Perception of Space and Motion* **1995**, chap. 3, 263-325.
37) Duffy, C. J., Wurtz, R. H., *J. Neurosci.* **1995**, 15, 5192.
38) Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., Williamson, J. R., *Science* **1996**, 273, 1547-1551.
39) Cech, T. R., *Proc.natn. Acad. Sci. U.S.A.* **1986**, 83, 4360-4363.
40) Orgel, L. E., *J. theor. Biol.* **1986**, 123, 127-149.
41) Benner, S. A., Elligton, A. D. & Tauer, A., *Proc. natn. Acad. Sci. U.S.A.* **1989**, 86, 7054-7058.
42) Elligton, A. D., Szostak, J. W., *Nature* **1990**, 346, 818-822.
43) Green, R., Ellington, A. D., Bartel, D. P., Szostak, J. W., *Methods* **1991**, 2, 75.
44) Eaton, B. E., Pieken, W. A., *Annu. ReV. Biochem.* **1995**, 64, 837.
45) Eaton, B. E., *Curr. Opin. Chem. Biol*. **1997**, 1, 10.
46) Brody, E. N., Gold, L. J., *Biotechnol.* **2000**, 74, 5.
47) Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., Toole, J. J., *Nature* **1992**, 355, 564.
48) Eaton, B. E., Gold, L., Hicke, B. J., Janjic, N., Jucker, F. M., Sebesta, D. P., Tarasow, T. M., Willis, M. C., Zichi, D. A., *Bioorg. Med. Chem*. **1997**, 5, 1087.
49) Green, L. S., Jellinek, D., Bell, C., Beebe, L. A., Feistner, B. D., Gill, S. C.; Jucker, F. M., Janjic, N., *Chem. Biol.* **1995**, 2, 683.
50) Hicke, B. J., Marion, C., Chang, Y. F., Gould, T., Lynott, C. K., Parma, D., Schmidt, P. G., Warren, S. J., *Biol. Chem.* **2001**, 276, 48644.
51) Morii, T., Hagihara, M., Sato, S., Makino, K., *J. Am. Chem. Soc.* **2002**, 124, 4617-4622.
52) Hasegawa, T., Hagihara, M., Fukuda, M., Nakano, S., Fujieda, N., Morii, T., *J. Am. Chem. Soc.* **2008**, 130, 8804-8812.
53) Sassanfar, M., Szostak, J. W., *Nature* **1993**, 364, 550-553.
54) Koizumi, M., Breaker, R. R., *Biochemistry* **2000**, 39, 8983-8992.
55) Burgstaller, P., Famulok, M., *Angew. Chem. Int. Ed. Engl.* **1994**, 33, 1084-1087.
56) Burke, D. H., Gold, L., *Nucleic Acids Res.* **1997**, 25, 2020-2024.

57) Hagihara, M., Fukuda, M., Hasegawa, T., Morii, T., *J. Am. Chem. Soc.* **2006**, 128, 12932-12940.
58) Davis, J., Szostak, J. W., *Proc. Natl. Acad. Sci. USA* **2002**, 99, 11616-11621.
59) Carothers, J. M., Oestreich, S. C., Davis, J. H., Szostak, J. W., *J. Am. Chem. Soc.* **2004**, 126, 5130-5135.
60) Lorsch, J. R., Szostak, J. W., *Acc. Chem. Res.* **1996**, 29, 103-110.
61) Dieckmann, T., Suzuki, E., Nakamura, G. K., Feigon, J., *RNA* **1996**, 2, 628-640.
62) Jiang, F., Kumar, R. A., Jones, R. A., Patel, D. J., *Nature* **1996**, 382, 183-186.
63) Meli, M., Vergine, J., De'cout, J. L., Maurel, M. C., *J. Biol. Chem.* **2002**, 277, 2104-2111.
64) Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., Patel, D., *J. Chem. Biol.* **2004**, 11, 1729-1741.
65) Montage, R. K., Batey, R. T., *Nature* **2006**, 441, 1172-1175.
66) Uesugi, S., Ikehara, M., *J. Am. Chem. Soc.* **1977**, 99, 3250-3253.

2.3. Construction of modular ribonucleopeptide (RNP) based biosensors

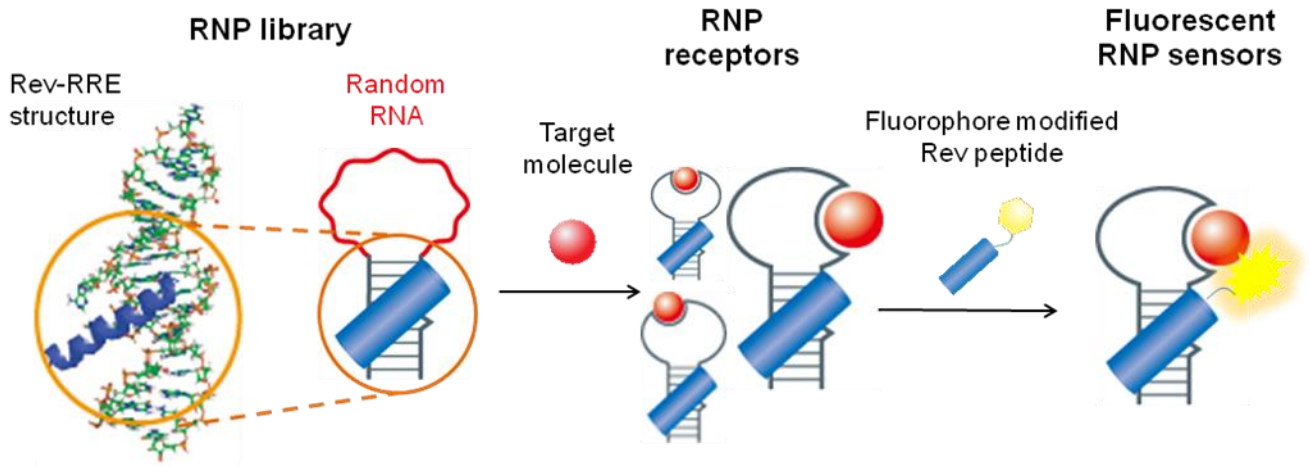


Figure 15: Schematic illustration of the stepwise construction of RNP receptors and fluorescent RNP sensors. Randomized RNA sequences are introduced as loop into the stable RNP complex to afford the RRENn RNP library.[1] By applying the in vitro selection method on this library, RNP receptors that bind specifically to various target ligands are obtained. RNP receptors are subsequently converted to RNP sensors by introducing a fluorophore at the N-terminus of the Rev peptide.[1,2]

2.3.1. Modular strategies for tailoring RNP fluorescent biosensors

A modular strategy that permits a facile preparation of biosensors with tailored characteristics by a simple combination of a receptor and a signal transducer has recently emerged as a new paradigm for a versatile design of fluorescent biosensors. Stojanovic and coworkers have reported a modular design of signaling aptamers based on the allosteric regulation of binding events.[3] These chimeric aptamers composed of two modular aptamers, one for the target recognition and another for holding a reporter dye, displayed robust fluorescence responses to three different targets. At paragraph 2.2. we introduced the innovative concept of simultaneous preparation of fluorescent biosensors with diverse functions based on a framework of ribonucleopeptide (RNP)[4] using the stable complex between Rev peptide and RRE RNA as scaffold.[5-8] The RNP receptors obtained by in vitro selection have an assortment of RNA structures and reveal different affinity to the target molecule, and therefore are considered as an RNP receptor library (**Figure 16**).

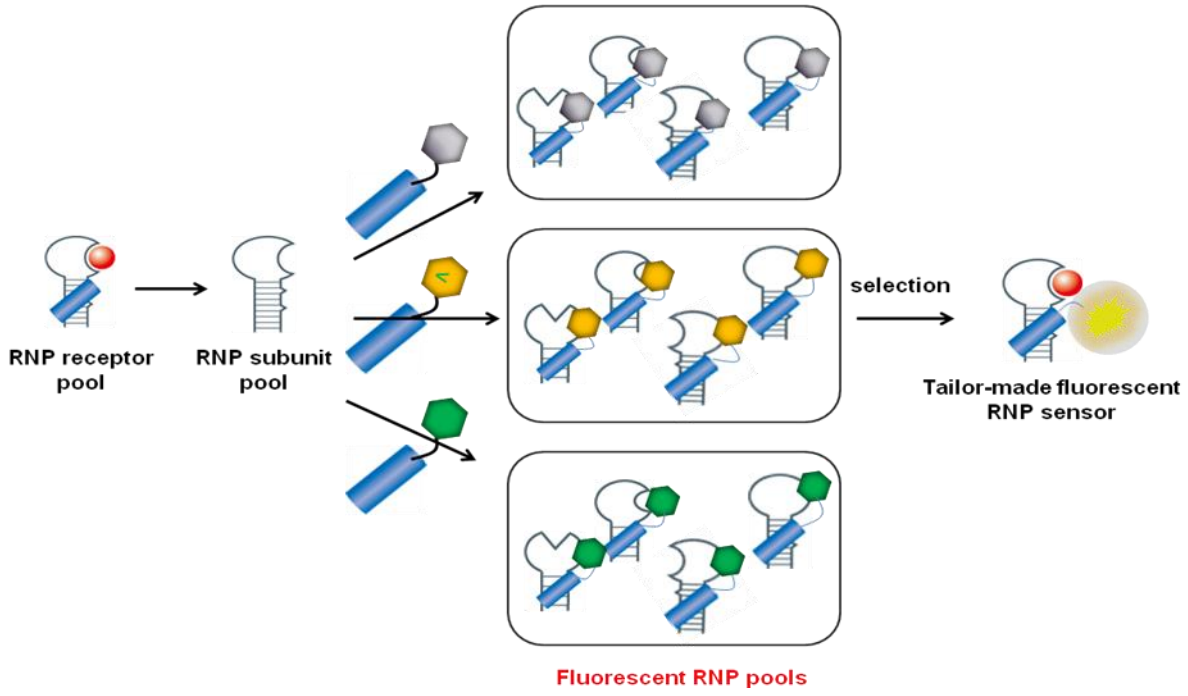


Figure 16: Schematic illustration for a screening strategy of a tailor-made RNP fluorescent sensor. Combination of the RNA subunit library of the RNP receptor and several fluorophore-labeled Rev peptide subunits generates combinatorial fluorescent RNP receptor libraries, from which RNP sensors with desired optical and/or binding properties are screened.

From the structural analyses of ATP-binding RNP receptors, the binding mode for substrate and the interaction mode of each nucleotide were evaluated.[6] The ATP-binding RNP receptors show structural changes upon substrate binding, as a result a change in the fluorescence spectrum is also observed (**Figure 17**).

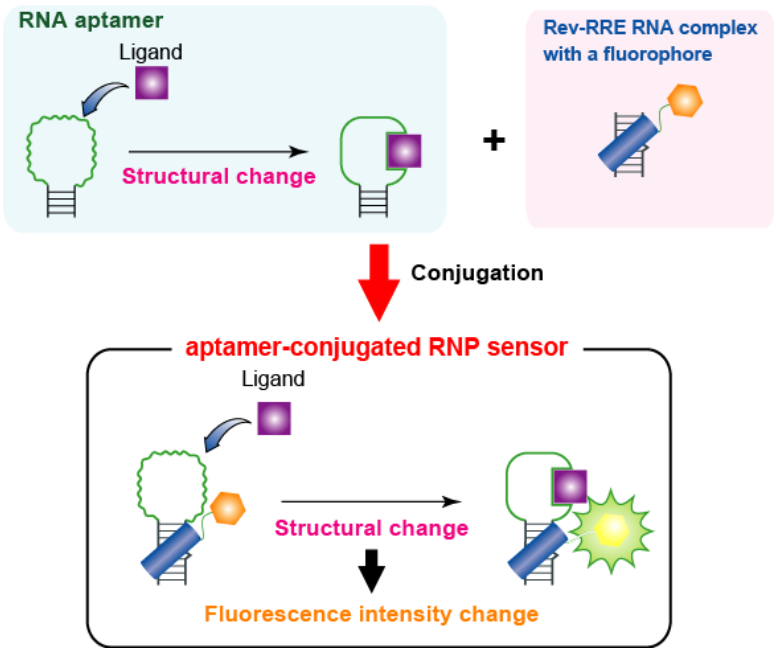


Figure 17: The concept of aptamer-conjugated fluorescent sensor by using fluorophore modified ribonucleopeptides. The aptamers that induced structural changes upon substrate binding can be converted into a fluorescent sensor by introducing the RNP motif modified with a fluorophore.

In conclusion, it can be postulated that the structural changes upon substrate binding and the stability acquirement of the triple complex of Rev, RNA, and the substrate occur in a comprehensive manner. Therefore, every aptamer, constructed independently, that shows structural changes upon the substrate binding event, has a potential for an easily conversion into fluorescent sensor (**Figure 17**). The transmission of the structural change can be achieved fusing the aptamer to the Rev-RRE complex and subsequently, converting it into a fluorescent sensor. This methodology can be applied for the convenient construction of aptamer-based sensors with desired fluorescent properties taking advantages from the possibility of freely choosing the fluorophore that matches to the target detection. In order to verify the successful application of this strategy, I studied the development of seminaphthorhodafluor (SNARF)-tethering RNP fluorescent biosensors. This work is material of a paper published in 2012 (*Org. Biomol. Chem.* **2012**, 10, 8767-8769).

2.3.2. Design of seminaphthorhodafluor (SNARF)-tethering RNP fluorescent biosensors

As for fluorescence detection, ratiometric sensing presents several advantages, such as enhanced dynamic range, precise correction without being affected by environmental variables, and convenient visual monitoring that make it suitable for bio analytical measurements.[9] For the purpose of expanding the suitability of this technique, we developed an innovative strategy for obtaining universally applicable ratiometric fluorescent sensors with modular RNP structure by using targets specific RNA aptamers as a receptor module and ratiometrically detectable fluorophore-attached Rev as a reporter module. Seminaphthorhodafluor (SNARF)[9] (**Figure 18 C**) is a representative red-emitting fluorophore that shows pH dependent dual-emission properties. RNPs that tether SNARF through the Rev peptide were designed in order to sensitively detect micro-environmental changes associated with substrate binding.[10-19]

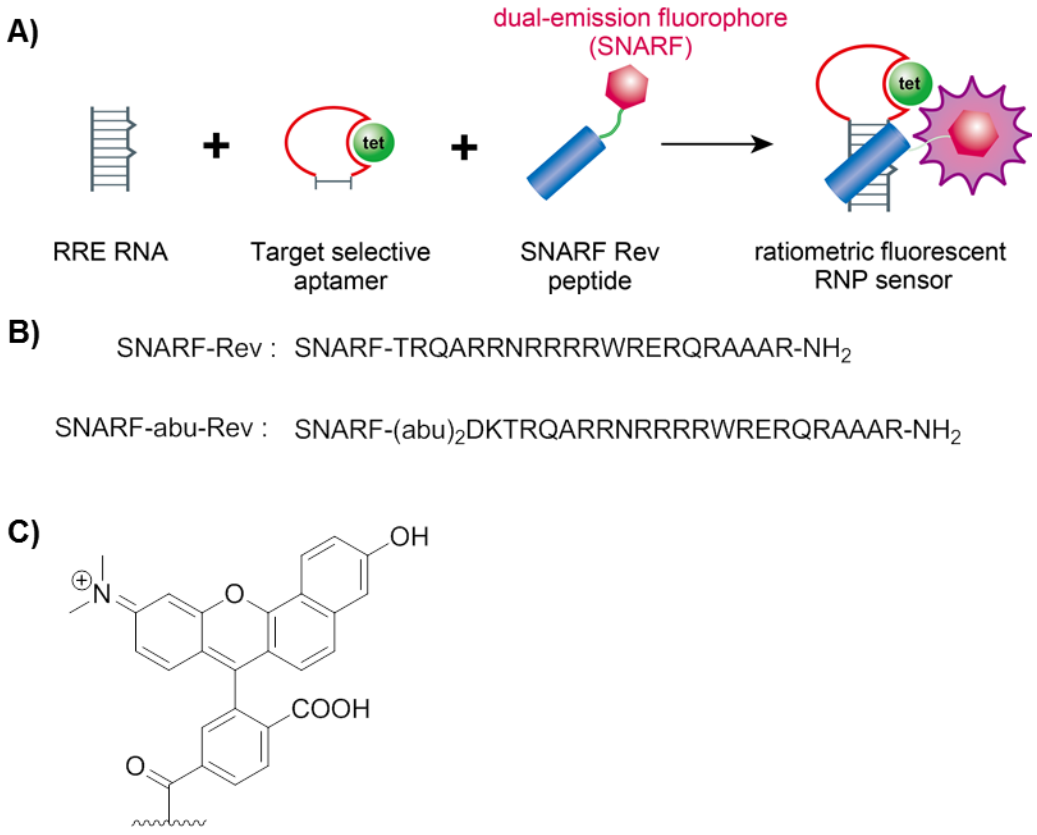


Figure 18: A) Schematic illustration of fluorescent ratiometric biosensor based on RNP. **B)** Peptide sequences of SNARF-Rev derivatives used in this study. **C)** Structure of SNARF (6-isomer).

Two types of SNARF-attached Rev peptide ((**Figure 18 B**), which incorporate linkers of different length, were designed. The first one (SNARF-Rev) was obtained by directly modifying the N-terminus of Rev peptide with 6-carboxy-SNARF isomer (**Figure 18 C**). The second one (SNARF-abu-Rev) contemplated the insertion of a spatial linker (γ-aminobutyric acid) between the N-terminus of Rev peptide and SNARF (**Figure 18 B**). These Rev-peptides were characterized by UV-visible, HPLC, fluorescent spectroscopy and mass spectrometry (See Materials and Methods).[19]

2.3.3. Tetracycline selective sensors

As first trial, tetracycline selective ratiometric fluorescent sensors were designed. Tetracycline is an antibiotic, which inhibits prokaryotic translation by interfering with binding of the aminoacyl-tRNA to the ribosomal A-site. It is a widely used therapeutic agent of low toxicity, active against most common pathogens, which has also been used at sub-therapeutic levels in animal feed to stimulate weight gain, as well as for prophylactic disease control.[20-22] To construct tetracycline-binding RNA receptor module of RNP, tetracycline binding aptamer cb28, isolated by Berens *et al.*, via *in vitro* selection,[23] was fused to RRE RNA (tet-RRE) (**Figure 19**).[24,19]

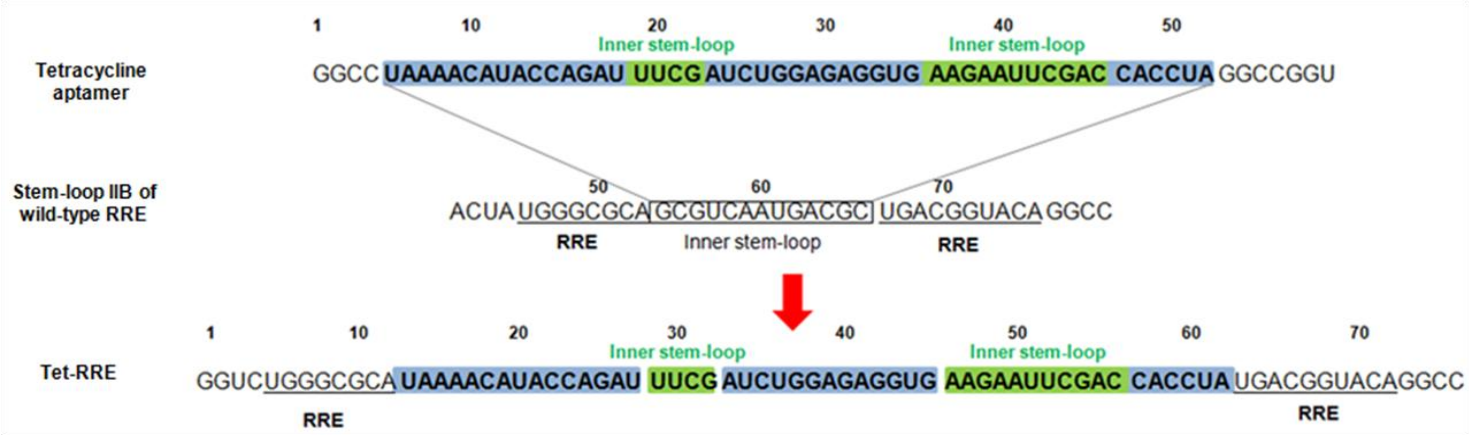


Figure 19: Nucleotide sequences of the tetracycline aptamer, RRE RNA, and modular tetracycline-binding RNP (tet-RRE RNA). The tetracycline aptamer fused to the RRE sequence (tet-RRE) was complexed with 6-carboxy-SNARF-modified Rev (SNARF-Rev and SNARF-abu-Rev).

SNARF modified Rev-peptides were then complexed with the tet-RRE providing fluorescent RNPs (tet-RRE/SNARF-Rev). The sensing capability and selectivity of tet-RRE/SNARF-Rev RNP were initially evaluated by fluorescent titration experiments with tetracycline.

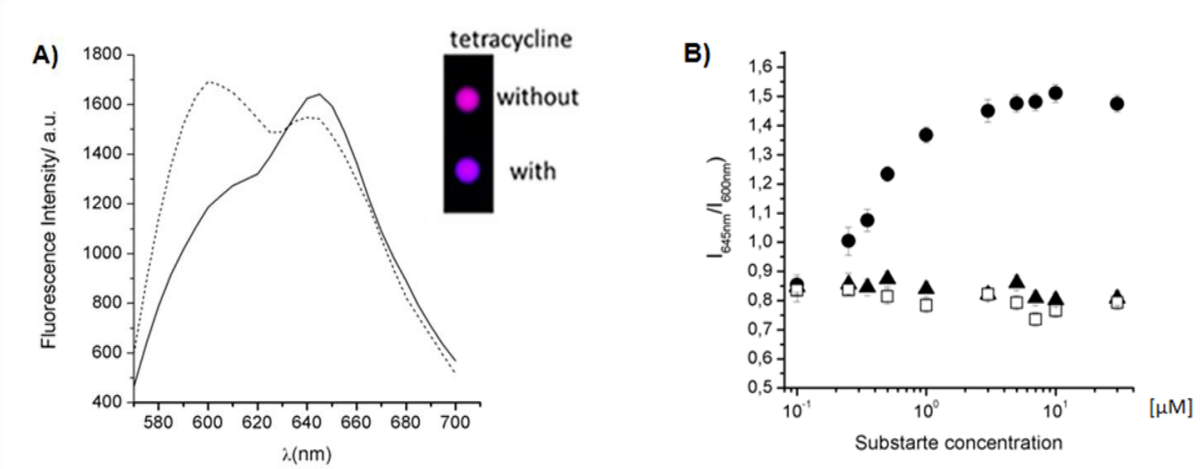


Figure 20: A) Comparison of the fluorescence spectrum of tet-RRE/ SNARF-Rev with (plane line) or without (dash line) tetracycline (30 μM). (inset) Fluorescent images of tet-RRE/SNARF-Rev with or without tetracycline (30 μM) filtered at 580 ± 25 nm (red) and 640 ± 20 nm (blue). **B)** Fluorescence titration plots of the emission intensity ratio (645 nm / 600 nm) of tet-RRE/ SNARF-Rev vs. antibiotics concentration. Tetracycline (black filled circle), streptomycin (black filled triangle), chloramphenicol (black opened square).

Figure 20A shows the fluorescence spectra of tet-RRE/SNARF-Rev with or without tetracycline in aqueous buffered conditions (pH 7.60). A typical ratiometric fluorescence change, that consists of a decrease in the emission at 600nm, which is due to the acidic phenol form of SNARF, and a simultaneous increment of the emission at 645nm, which is due to its basic phenolate form (excitation wavelength 534nm), is apparently observed (**Figure 20 A and Figure 21**).[9,19]

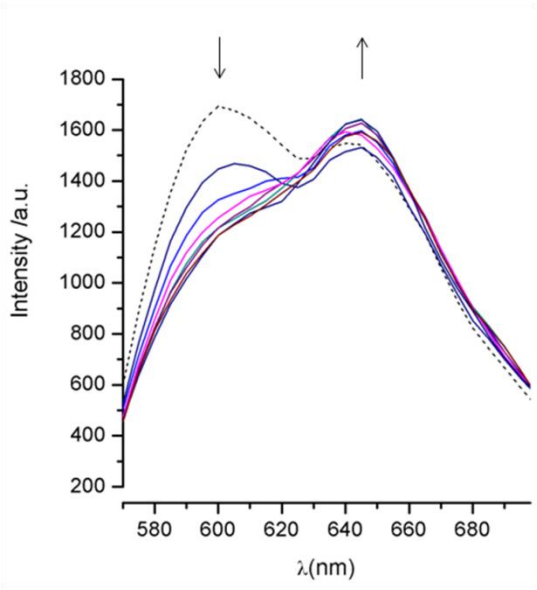


Figure 21: Fluorescent spectral change of tet-RRE/SNARF-Rev upon the addition of tetracycline (0-30μM). The arrows show the direction of change of the spectrum at that position as tetracycline concentration is increased.

A change in the fluorescent colour of tet-RRE/SNARF-Rev was clearly detected in the presence of tetracycline (inset of **Figure 20 A**). The emission intensity ratio *R* (I_{645nm}/I_{600nm}) plot showed good saturation behaviour (differential R=0.7) and afforded a dissociation constant much lower than 1μM. On the other hand, no significant spectral change was observed when tet-RRE/SNARF-Rev was titrated with streptomycin or chloramphenicol (**Figure 20 B**). These results indicate that ratiometric fluorescent change could successfully read-out the binding of tet-RRE/SNARF-with tetracycline.

In order to clarify the mechanism of the ratiometric fluorescent change, subsequently, a fluorescent pH titration of the SNARF unit on tet-RRE/SNARF-Rev in the absence or presence of tetracycline was accomplished (**Figure 22**). The pH-dependent fluorescence emission spectrum of tet-RRE/SNARF-Rev alone is moderately shifted compared to the one achieved in the presence of tetracycline (**Figure 23**), indicating that the SNARF fluorophore displays more acidic pK_a upon the substrate binding.[19]

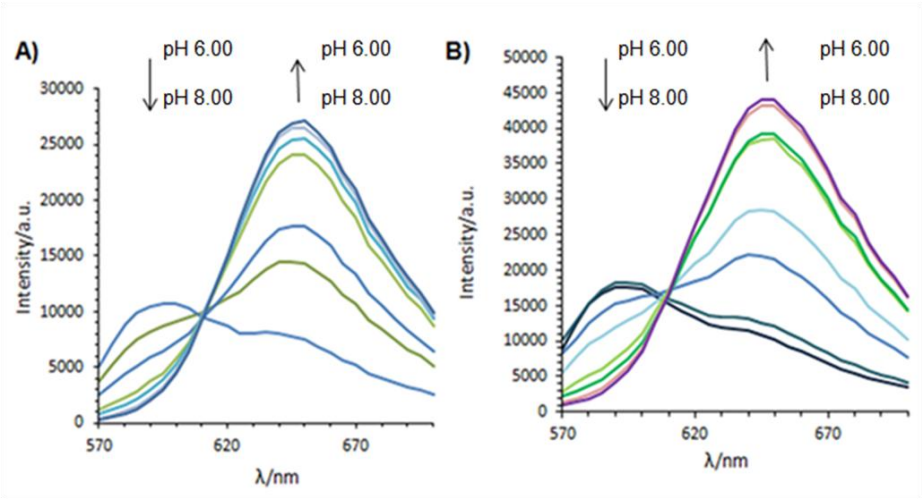


Figure 22: A) Fluorescence spectral change of tet-RRE/ SNARF-Rev at different pH values (6.00 to 8.00). **B)** Fluorescence spectral change of tet-RRE/ SNARF-abu-Rev at different pH conditions (pH 6.00 to 8.00).

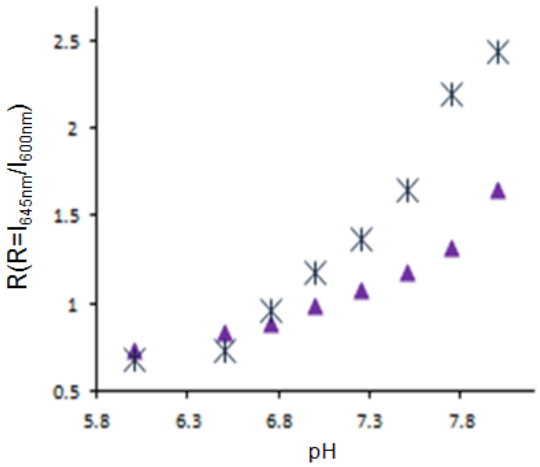


Figure 23: A pH titration (pH 6.00 to 8.00) of SNARF alone (light blue snow contraction line), SNARF-Rev peptide (filled red circle), tet-RRE/SNARF-Rev RNP (filled purple triangle) and tet-RRE/SNARF-Rev upon 3 µM tetracycline addition (dark snow construction line).

This pK_a shift towards acidic pH region suggests that the basic phenolate form of SNARF unit is stabilized. These results strongly suggest that the ratiometric fluorescent emission change of tet-RRE/SNARF-Rev upon the tetracycline binding is ascribed to the pK_a shift of the SNARF unit as a result of micro-environmental change of SNARF, which is induced by the conformational change of RNA upon tetracycline binding.[7] Another interesting consideration comes from the observation of a more remarkable response of tet-RRE/SNARF-Rev RNP (**Figure 24**) compared to the attached-linker one (Tet-RRE/SNARF-abu-Rev) (differential R=0.5). The result indicates that the proximity to the binding pocket is an important factor affecting the efficiency of these sensors.[19]

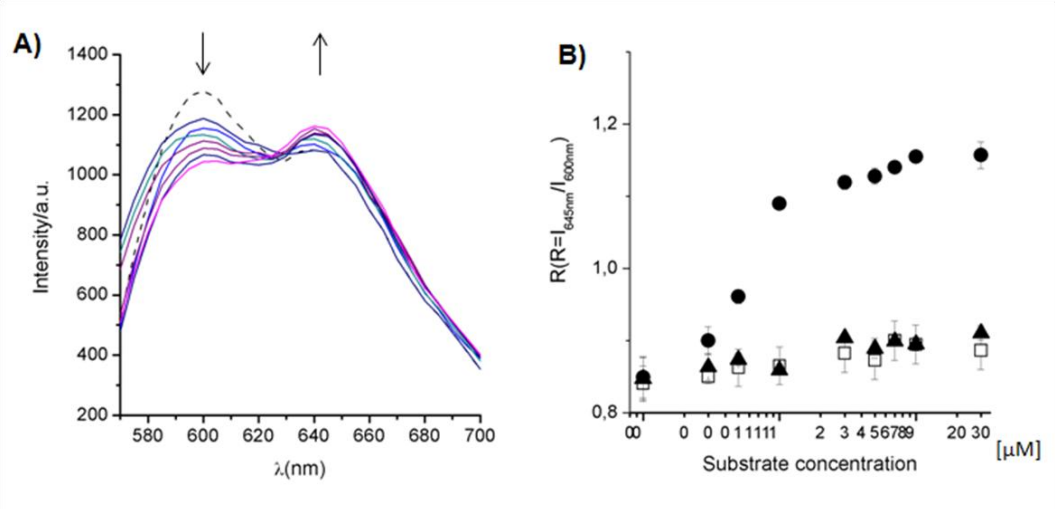


Figure 24: A) Fluorescence spectral change of tet-RRE/ SNARF-abu-Rev upon the addition of tetracycline (0-30 μM). **B)** Fluorescence titration plot of the emission intensity ratio (645 nm / 600 nm) of tet-RRE/ SNARF-abu-Rev vs. the antibiotic compounds concentration. Tetracycline (black filled circle), streptomycin (black filled triangle), chloramphenicol (black opened square).

2.3.4. Dopamine selective sensors

The applicability of the design strategy of modular fluorescent RNP was further tested by fluorescence titration using dopamine aptamers DH05 and DHc65 previously isolated by our group.[25] Because dopamine is a valuable heuristic bridge in defective brain chemistry study over the past medical history,[26] it is of particular interest to prepare a suitable analytical detecting tool that would accelerate clear understanding of the relationship between the function and the structure of each biologically active catecholamine.[25] Both DH05-RRE/SNARF-Rev and DHc65-RRE/SNARF-Rev complexes showed changes in the fluorescence intensity ratio R ($I_{645\ nm}/I_{600\ nm}$) upon addition of dopamine (**Figure 25A and 26A**).[19]

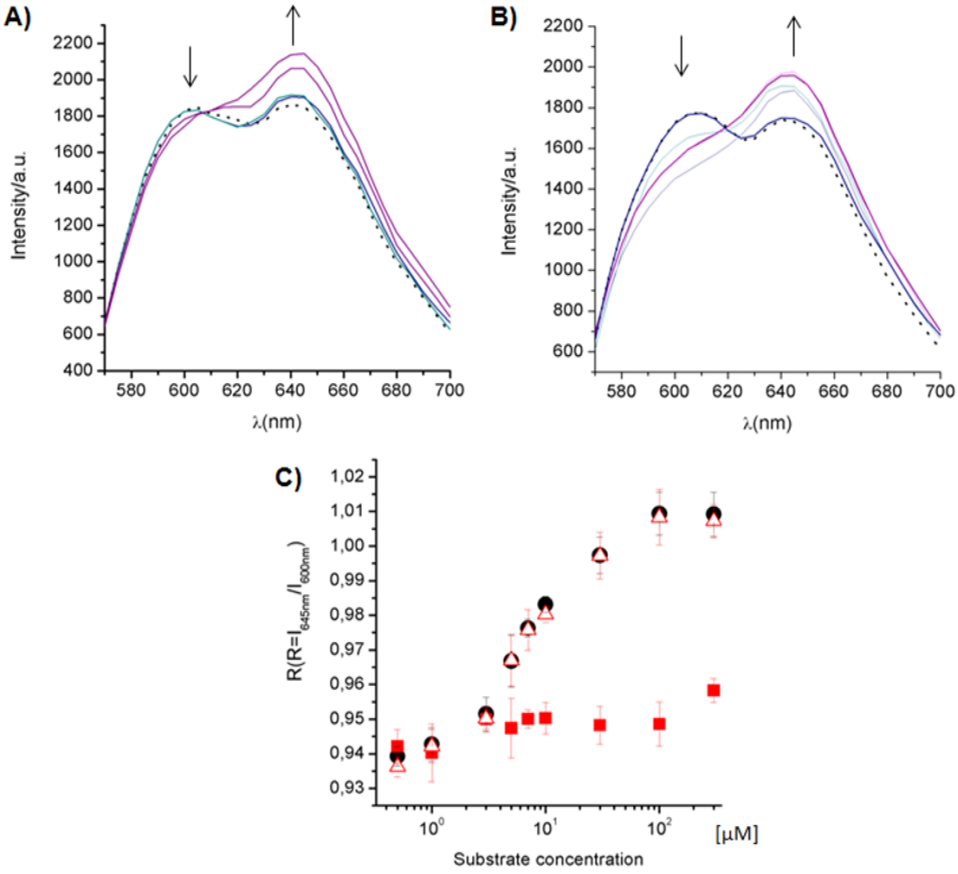


Figure 25: A) Fluorescence spectral change of DHc65-RRE/SNARF-Rev upon the addition of dopamine (0-300 μM). **B)** Fluorescence spectral change of DHc65-RRE/SNARF-abu- Rev upon the addition of dopamine (0-300 μM). **C)** Fluorescence titration plot of the emission intensity ratio (645 nm / 600 nm) of DHc65-RRE/SNARF-abu-Rev with dopamine (filled black circle), norepinephrine (open red triangle), L-tyrosine (filled red square).

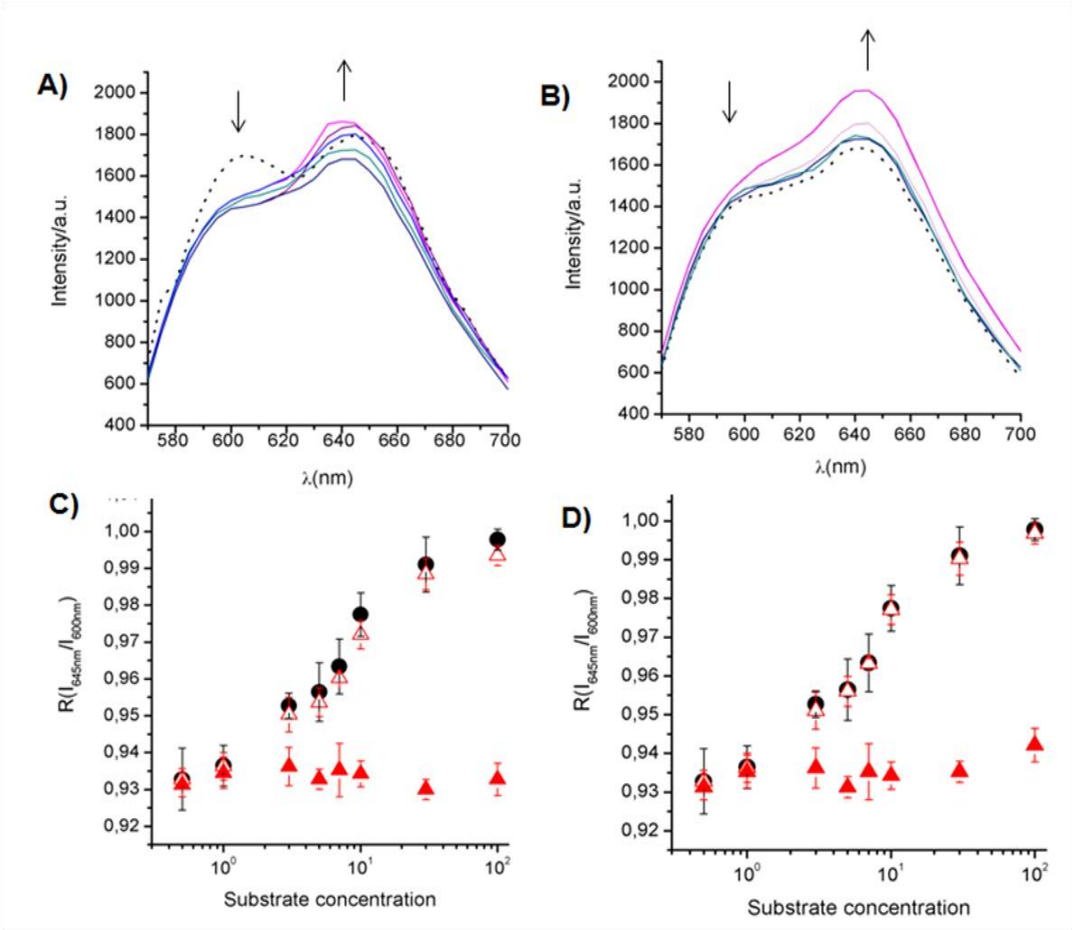


Figure 26: A) Fluorescence spectral change of DH05-RRE/SNARF-Rev upon dopamine addition (0-300 µM). **B)** Fluorescence spectral change of DH05-RRE/SNARF-abu-Rev upon dopamine addition DH05-RRE/SNARF-abu-Rev with dopamine (filled black circle), norepinephrine (open red triangle), L-tyrosine (filled red square). **C)** Fluorescence titration plot of the emission intensity ratio (645 nm/ 600 nm) of DH05-RRE/SNARF-Rev with dopamine (filled black circle), norepinephrine (open red triangle), L-tyrosine (filled red square). **D)** Fluorescence titration plot of the emission intensity ratio (645 nm/ 600 nm) of DH05-RRE/SNARF-abu-Rev with dopamine (filled black circle), norepinephrine (open red triangle), L-tyrosine (filled red square).

The increase in the emission intensity of basic phenolate form (645 nm) occurred in concomitance with a decrease in the emission intensity of the acidic phenol form at 600 nm (differential R=0.1). These results indicate that the modular RNP constructed from the induced fit-type aptamer[9] and the fluorescent RRE/Rev module successfully transduces the dopamine-binding event into measurable fluorescence signals. Each catecholamine has a distinct structure of a benzene ring with two hydroxyl groups and a terminal aminoethyl group. The selectivity of DHc65-RRE/SNARF-Rev and DH05-RRE/SNARF-Rev for dopamine against other catecholamine derivatives was also evaluated to state whether the subtle difference in their structure could be selectively recognized by SNARF-RNP sensors. The dissociation constants for the complexes of DHc65-RRE/SNARF-Rev and DH05-

RRE/SNARF-Rev with dopamine and the catecholamine derivative norepinephrine were obtained from the fluorescence titration curves (**Figures 25C** and **26C**; **Table 3**).

Ligand	K_D [µM]			
	DHc65		DH05	
	SNARF-Rev	SNARF-abu-Rev	SNARF-Rev	SNARF-abu-Rev
Dopamine	1.1	1.2	1.1	1.5
Norepinephrine	1.9	1.3	2.0	1.5
Tyrosine	>100	>100	>100	>100

Table 3: Equilibrium dissociation constants K_D (µM) of dopamine-binding RNPs.

The affinity of both RNPs to tyrosine was much lower than that to dopamine. However, both DHc65-RRE/SNARF-Rev and DH05-RRE/SNARF-Rev showed scarce efficiency in discriminating dopamine from norepinephrine while SNARF-abu-Rev containing RNPs failed in showing selectivity (**Figure 26 C**).

2.3.5. Streptomycin selective sensors

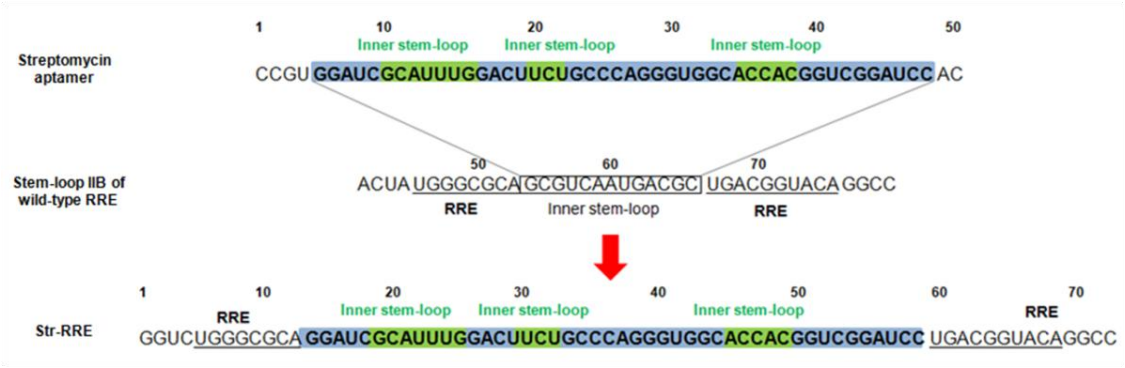


Figure 27: Nucleotide sequences of the streptomycin aptamer, RRE RNA, and modular streptomycin-binding RNP (str-RRE RNA). The streptomycin aptamer fused to the RRE sequence (str-RRE) was complexed with 6-carboxy-SNARF-modified Rev (SNARF-Rev and SNARF-abu-Rev).

Our modular RNP structure was also applied to the detection of streptomycin obtaining positive results. SNARF-Rev peptides were complexed with streptomycin aptamer[27]-modified RRE (**Figure 27**).

As observed for the case of tetracycline detection by tet-RRE/SNARF-Rev, the titration curve of str-RRE/SNARF-Rev addition shows a similar spectral change upon streptomycin that consists in a decrease of the phenol form at 600 nm and a simultaneous increase of the phenolate form at 645 nm with an isosbestic point at 625 nm (**Figure 28**).[19]

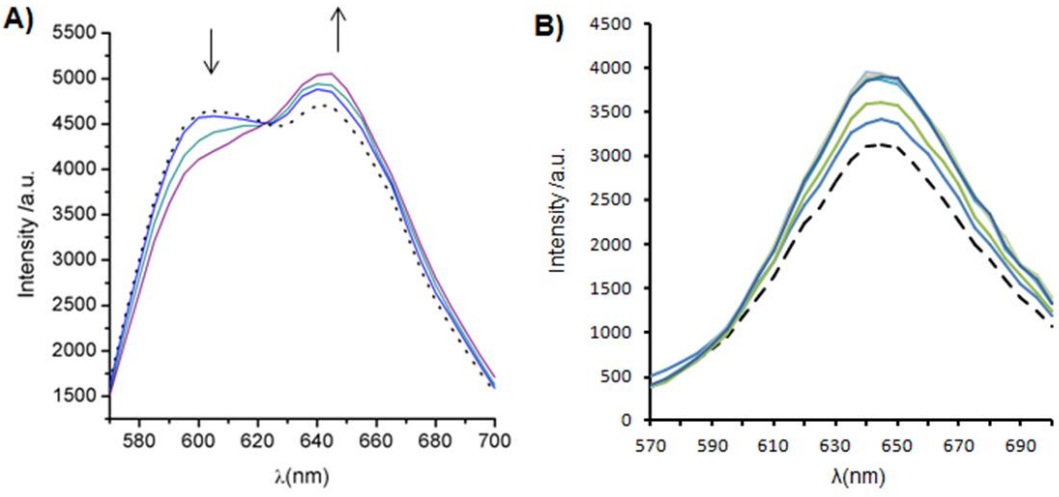


Figure 28: A) Fluorescence spectral change of str-RRE/SNARF- Rev upon the addition of streptomycin (0-300 μM). B) Fluorescence spectral change of str-RRE/SNARF-abu- Rev upon the addition of streptomycin (0-300 μM).

2.3.6. Discussion

- We have developed a versatile ratiometric fluorescence sensing system by using an RNA subunit that fuses a target specific aptamer to RRE RNA and a SNARF-modified Rev peptide. The mechanism of the dual emission sensing is ascribable to the pK_a shift of the SNARF chromophore of the reporter module that is induced by a ligand-binding event at the binding pocket of the receptor module.

- This statement is strongly suggested by the observation that the efficiency and selectivity of detection decrease when a spatial linker is inserted at the fluorophore-labeling position as in the case of SNARF-abu-Rev peptide.

- It is likely that a micro-environmental pH change near the fluorophore, which is caused by structural changes in the receptor module upon substrate binding, is a prerequisite for effective communication between the binding event at the receptor module and the fluorescence changes at the reporter module.

- Prospective applications enclose not only quantitative detection of a tagged analyte *in vitro,* but also for *in vivo* evaluation.

2.3.7. References

1) Hagihara, M., Fukuda, M., Hasegawa, T., and Morii, T., *J. Am. Chem. Soc.* **2006**, 128, 12932-12940

2) Hasegawa, T., Hagihara, M., Fukuda, M., Nakano, S., Fujieda, N., and Morii, T., *J. Am. Chem. Soc.* **2008**, 130, 8804-8812.

3) Stojanovic, M. N., and Kolpashchikov, D. M., *J. Am. Chem. Soc.* **2004**, 126, 9266-9270.
4) Hagihara, M., Fukuda, M., Hasegawa, T., and Morii, T., *J. Am. Chem. Soc.* **2006**, 128, 12932-12940.
5) Morii, T., Hagiwara, M., Sato, S., and Makino, K., *J. Am. Chem. Soc.* **2002**, 124, 4617-4622.
6) Fukuda, M., Hayashi, H., Hasegawa, T., and Morii, T., *Trans. Mat. Res. Soc. Jpn.* **2009**, 34, 525-527.
7) Hasegawa, T., Ohkubo, K., Yoshikawa, S., and Morii, T., *J. Surf. Sci. Nanotech.* **2005**, 3, 33-37.
8) Hasegawa, T., Hagihara, M., Fukuda, M., Nakano, S., Fujieda, N., and Morii, T., *J. Am. Chem. Soc.* **2008**, 130, 8804-8812.
9) R. P. Haugland in *Handbook of Fluorescent Probes and Research Products*, 9thed., Molecular Probes, Eugene, **2002**, 833-836.
10) Liew, F. F., Hayashi, H., Nakano, S., Nakata, E., and Morii, T., *Bioorg. Med. Chem.* **2011**, 19, 5771-5775.
11) Morii, T., Hagihara, M., Sato, S., and Makino, K., *J. Am. Chem. Soc.* **2002**, 124, 4617.
12) Sato, S., Fukuda, M., Hagihara, M., Tanabe, Y., Ohkubo, K., and Morii, T., *J. Am. Chem. Soc.* **2005**, 127, 30.
13) Hasegawa, T., Ohkubo, K., Yoshikawa, S., and Morii, T., *e-J. Surf. Sci. Nanotech.* **2005**, 3, 33.
14) Hagihara, M., Fukuda, M., Hasegawa, T., and Morii, T., *J. Am. Chem. Soc.* **2006**, 128, 12932.
15) Hasegawa, T., Hagihara, M., Fukuda, M., and Morii, T., *Nucleosides Nucleotides Nucleic Acids* **2007**, 26, 1277.
16) Hasegawa, T., Hagihara, M., Fukuda, M., Nakano, S., Fujieda, N., and Morii, T., *J. Am. Chem. Soc.* **2008**, 130, 8804.
17) Liew, F. F., Hasegawa, T., Fukuda, M., Nakata, E., and Morii, T., *Bioorg. Med. Chem.* **2011**, 19, 4473.
18) Nakano, S., Nakata, E., and Morii, T., *Bioorg. Med. Chem. Lett.* **2011**, 21, 4503-4506.

19) Annoni, C., Nakata, E., Tamura, T., Liew, F. F., Gelmi, M. L., Morii, T., *Org. Biomol. Chem.* **2012**, 10, 8767-8769.
20) Epe, B., Wooley, P., and Horning, H., *FEBS Lett.* **1987**, 213, 443.
21) Spahn, C. M., and Prescott, C. D., *J. Mol. Med.* **1996**, 74, 423.
22) Chopra, I., Hawkey, P. M., and Hinton, M., *J. Antimicrob. Chemother.* **1992**, 29, 245.
23) Berens, C., Thain, A., and Schroeder, R., *Bioorg. Med. Chem.* **2001**, 9, 2549-2556.
24) Liew, F. F., Hasegawa, T., Fukuda, M., Nakada, E., and Morii, T., *Bioorg. Med. Chem.* **2011**, 19, 4473-4481.
25) Liew, F. F., Hasegawa, T., Fukuda, M., Nakata, E., and Morii, T., *Bioorg. Med. Chem.* **2011**, 19, 4473-4481.
26) Björkjund, A., and Dunnett, S. B., *Trends Neurosci.* **2007**, 30, 194.
27) Wallance, S. T., and Schroeder, R., *RNA* **1998**, 4,114.

2.4. Development of covalently linked RNP sensors: a successful example of cross-linking

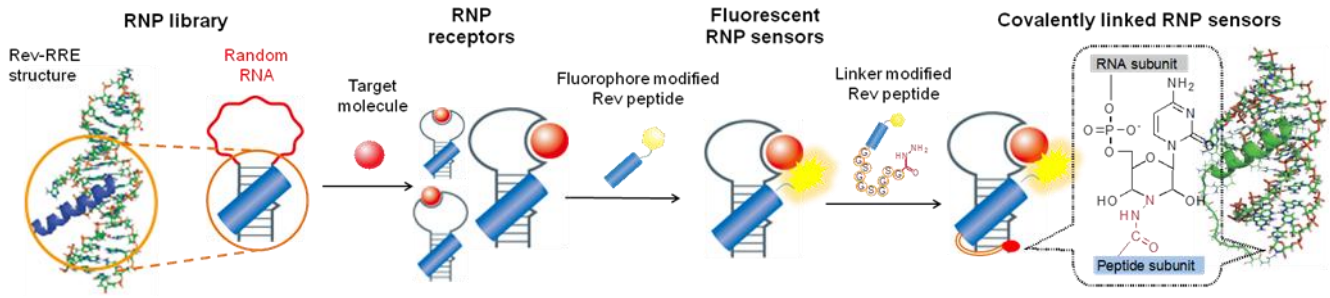


Figure 29: Schematic representation of the RNP sensors construction strategy.

2.4.1. Introduction

Almost all biologically important events inside living organism are regulated in well-coordinated manner by various molecules such as small organic molecules, ions, and proteins. Therefore, it is necessary to develop a strategy for the construction of stable fluorescent biosensors that can detect two or more cell signaling molecules at the same time in the cell. Generally, usable biosensors require optical and chemical properties, such as tunable wavelengths and concentration ranges for the ligand detection and, in some cases, stability in the cell media. In addition, the biosensor has to be easily tailored for a target and applicable to different target molecules at the same time. As a promising candidate, we have developed a strategy for constructing fluorescent sensors by utilizing a ribonucleopeptide (RNP) framework. However, the non-covalent configuration becomes a disadvantage for the practical measurements, especially when the sensor concentration is reduced. Previously reported bimolecular sensors, such as the signaling aptamer and the aptamer sensor non-covalently embedded with a fluorophore, would face the same issues. An effective method to overcome such disadvantages of the non-covalent character was developed for antibody engineering. Immunoglobuline Fv fragments, formed by the non-covalent complex of heterodimers of the heavy-chain variable domain (V_H) and the light-chain variable domain (V_L) are the smallest functional modules of antibodies required for antigen high-affinity binding. Because of their intrinsic non-

covalent nature, unmodified F_V fragments showed a limited stability at low protein concentration and under physiological conditions. Therefore, in order to improve the stability, the F_V fragments chains were covalently linked with disulfide bond or connected with a flexible peptidic linker to form a single-chain F_V molecule. The concept of stabilizing the complex formation by covalently linking can also be applied to a fluorescent RNP sensor; moreover it is not only expected to enhance the stability of the complex formation, but also to avoid exchange reactions between the subunits when multiple RNP sensors coexist. We report here the synthesis of stable fluorescent RNP sensors by the covalently linking of RNA and fluorophore-modified Rev peptide subunit and demonstrate the usefulness of the covalent RNP sensor by target sensing.

2.4.2. Covalently linked RNP assembling strategy

Non-covalent fluorescent RNP sensors for nucleotide triphosphates have been constructed as previously reported. To construct the stable fluorescent RNP sensor, the RNA subunit and peptide subunit were covalently tethered by a chemical modification (**Figure 30**).

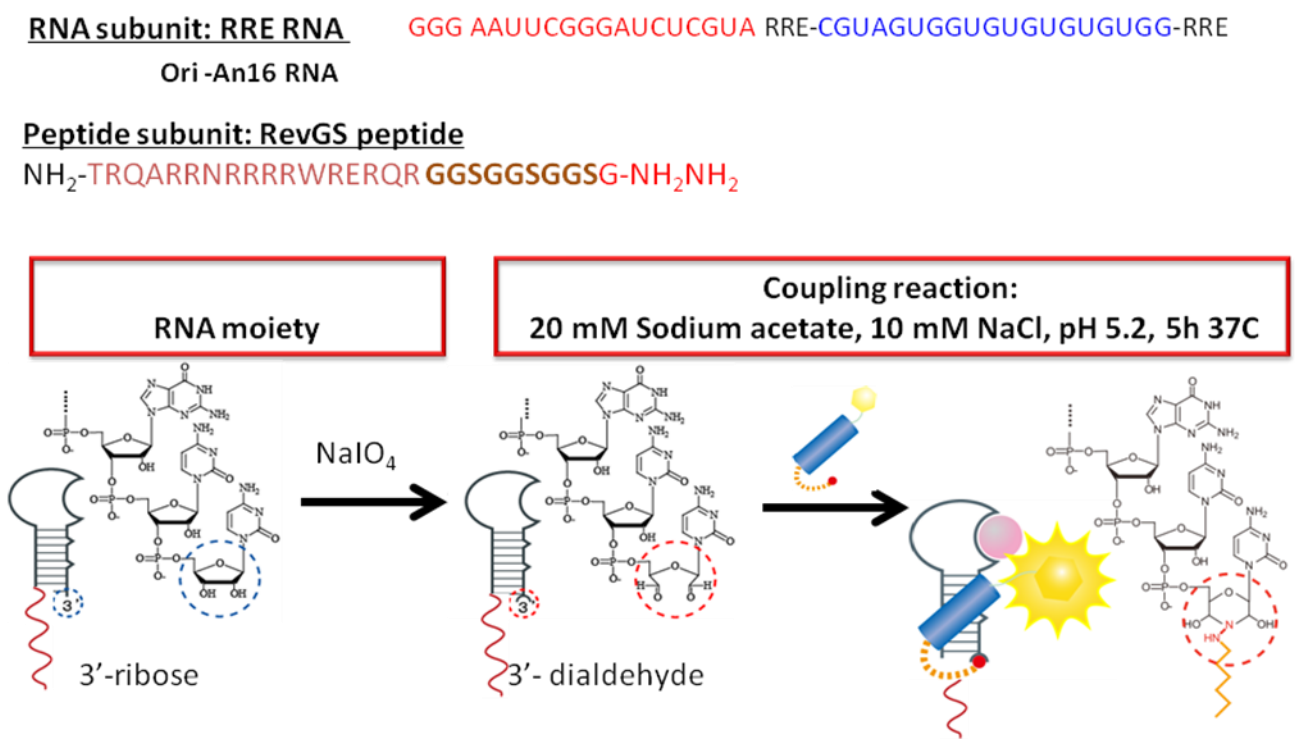


Figure 30: Schematic illustration of the covalently linking strategy for RNP sensors.

The RNA subunit of the fluorescent RNP sensor was treated with sodium periodate to form a 3'-dialdehyde group. On the other hand, a hydrazide-derivatized linker peptide, that has flexible amino acid sequences (GGSGGSGGSG), was introduced to the C-terminus of Rev peptide to be used for the covalent bond formation with RNA subunit. The length of the peptide linker was determined from the three-dimensional structure of the complex of the Rev peptide and RRE RNA. These two modified subunits were expected to form a covalent RNP complex between the 3'-terminal of RNA subunit and C-terminus of peptide subunit by a hydrazone bond. To confirm the effectiveness of the strategy, we chose the fluorescent RNP sensor for ATP and GTP as a model. At the first step, ATP-binding RNP

receptors and GTP-binding RNP receptors were selected from an RNP library as described previously. A26RNA and G23 RNA that show relatively high fluorescence intensity to each substrate were utilized. A26 RNA and G23 RNA were constructed by in vitro transcription and purified by denaturing polyacrylamide gel electrophoresis. Purified RNA was treated with freshly prepared sodium periodate to modify the 3'-terminal ribose to 3'-dialdehyde by periodate oxidation. After the reaction, the excess of sodium periodate was quenched by glycerol and the resultant RNA was purified with ethanol precipitation. The peptide subunit for the covalent linkage was obtained by introducing a fluorophore to the N-terminus and a hydrazide linker peptide at the C-terminus of Rev peptide. The Rev peptide with the linker sequence was synthesized by 4-HMBA (hydroxymethylbenzoic acid)-PEG resin according to the Fmoc chemistry protocols. 5-Carboxyfluorescein and 1-pyrenesulfonyl chloride were conjugated with N-terminal of the Rev peptide upon the cleavage from the resin. After the deprotection of the function groups on the chains, both peptides were purified by HPLC and characterized by MALDI-TOF MASS spectrometry. [5-Carboxyfluorescein hydrazine modified Rev peptide (5FAM-Rev-HZ): m/z 3469.1 (calcd. for [M+H]+ 3469.7); 1-Pyrenesulfonyl acid modified Rev peptide (Pyr-Rev-HZ): m/z 3379.2 (calcd. for [M+H]+3376.7)]. Subsequently, these two modified subunits were covalently linked. Coupling reaction of the 3'-modified A26 RNA with 5-carboxyfluorescein modified Rev peptide hydrazide (5FAM-Rev-HZ) and G23 RNA with 1-pyrenesulfonyl acid modified Rev hydrazide (Pyr-Rev-HZ) was performed in sodium acetate buffer (pH 5.2). The covalent RNP complex was purified by phenol-chloroform extraction and ethanol precipitation. Furthermore, collected RNP was purified with a denature PAGE to remove unreacted RNA and peptide. The recovery yields of RNP were calculated to be 36% (c-A26/5FAM-Rev) and 32% (c-G23/Pyr-Rev) by confirmed by MALDI-TOF mass spectroscopy, c-G23/Pyr-Rev was found; m/z 21745.3 (calcd for [M+H]+ 21588.3). These results indicate that the covalent fluorescence RNP complex was successfully formed according to our design.

2.4.3. Fluorescent response of the covalent RNP complex

Fluorescence responses of the covalently linked c-A26/5FAM-Rev, c-G23/Pyr-Rev and each non-covalent RNPs (A26/5FAM-Rev, g23/Pyr-Rev complexes) were evaluated in the presence of increasing concentration of ATP or GTP (**Figure 31**). In the case of ATP-binding RNP sensors, relative fluorescence intensity (I/I_0) of ATP sensors was calculated by dividing fluorescence intensity in the presence (I) by fluorescence intensity in the absence (I_0) of Ado (**Figure 31 a**). Relative fluorescence intensity was plotted against the Ado concentration to calculate the dissociation constant (K_D) of fluorescence RNP sensor and Ado saturation values of the changes in relative fluorescence intensity was found to be 1.8 for 0.1μM c-An16/5FAM-Rev, similar to the value of 1.8 for An16/5FAM-Rev. K_D values for the Ado complex of nc-A26/5FAM-Rev and c-A26/5FAM-Rev were 12.8 and 51.5μM, respectively (**Figure 31 b**). The binding affinity of A26/5FAM-Rev was increased by forming the covalent complex. Non-covalent complex exhibited almost no fluorescent response when the RNP concentration was decreased to 100nM, indicating that the non-covalent RNP complex was mostly dissociated into each subunit (**Figure 31 a**). As expected, the covalently linked sensor c-A26/5FAM-Rev still provided distinct fluorescent signals even at 100nM. The non-covalent RNP sensor was converted to a covalently linked assembly stabilized by an intramolecular interaction of the peptide and RNA subunits without diminishing the sensing function. On the other hand, GTP sensor cG23/Pyr-Rev shows slightly lower affinity to GTP than that of non-covalent sensor. However, the relative fluorescence intensity is significantly higher. The stability of interaction between RNA and peptide subunit clearly effect to the affinity and the fluorescence intensity change.

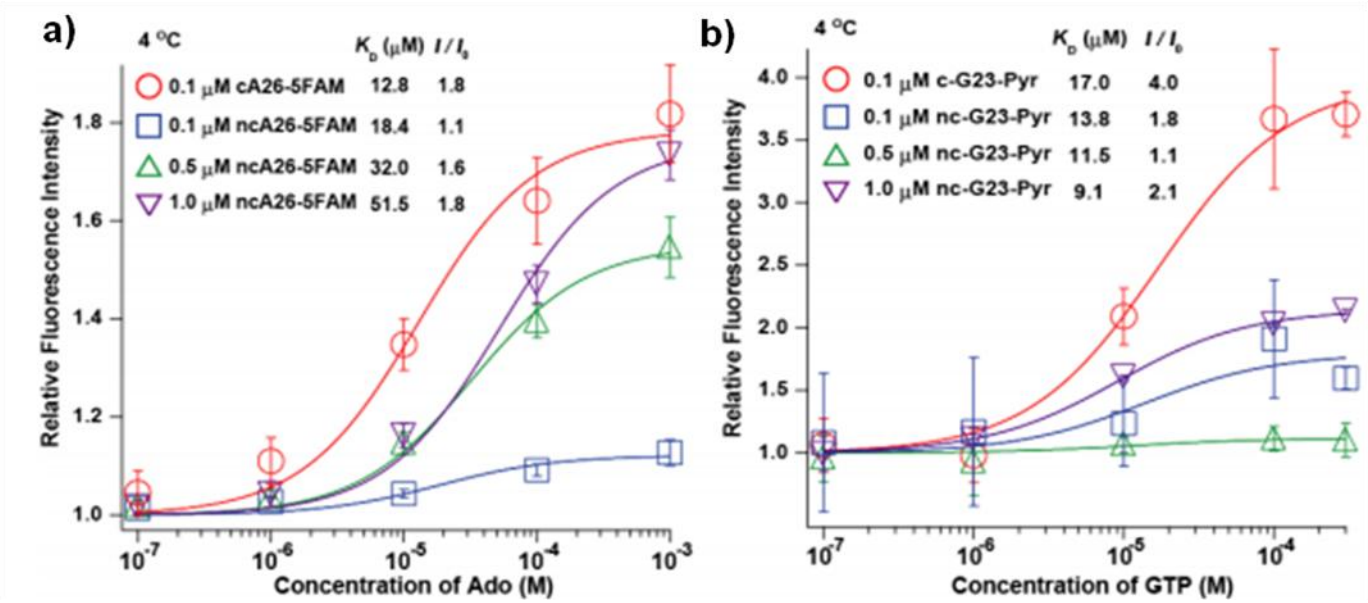


Figure 31: (a) Saturation curves for the relative fluorescence intensity changes of nc-A26/5FAM-Rev and c-A26/5FAM-Rev by titration with Ado in a buffer containing 10 mM Tris-HCl (pH 7.6), 100 mM NaCl, 10 mM MgCl2 and 0.005% Tween 20 at 4 °C. Titration of ATP-binding fluorescent RNP complex c-An16/5FAM-Rev (0.1 μM; red circle) and nc-An16/5FAM-Rev (0.1 μM; purple reverse triangle, 0.5 μM; green triangle, 1.0 μM; blue square) with Ado (0.1, 1, 10, 100, 1000 μM) were shown. **(b)** Saturation curves for the relative fluorescence intensity changes of nc-G23/Pyr-Rev and c-G23/Pyr-Rev by titration with GTP in a buffer containing 10 mM Tris-HCl (pH 7.6), 100 mM NaCl, 10 mM MgCl2 and 0.005% Tween 20 at 4 °C. Titration of GTP-binding fluorescent RNP complex c-G23/Pyr-Rev (0.1 μM; showing in red circle) and nc-G23/Pyr-Rev (0.1 μM; purple reverse triangle, 0.5 μM; green triangle, 1.0 μM; blue square) with GTP (0.1, 1, 10, 100, 300 μM) were shown. Dissociation constants and the maximum relative fluorescence intensity for the ATP and GTP complex of each RNP are shown in the figure, respectively.

2.5. Conclusions

The fluorescent RNP sensors represent a successful example of RNA aptamer-based artificial receptors for small target molecules. The non-covalent RNA complex embedding an aptamer function and a fluorophore-labeled peptide affords a fluorescent ribonucleopeptide (RNP) framework for constructing fluorescent sensors. Taking advantage of the non-covalent properties of the RNP complex, the ligand-binding and fluorescence characteristics of the fluorescent RNP can be independently tuned according to the nature of the RNA and peptide subunits, respectively. Fluorescent sensors tailored for given measurement conditions, such as a detection wavelength and a detection concentration range for a ligand of interest can be easily identified by screening of fluorescent RNP libraries. We reported in this work, an effective strategy for obtaining ratiometric fluorescent RNP sensors tethering 6-SNARF (Seminapthorhorhodafluor) fluorophore However, the non-covalent nature of RNPs could represent a disadvantage when very low concentrations are required or when multiple sensors are coexisting in the same environment. In order to overcome this drawback, we developed a strategy to convert a non-covalent fluorescent RNP sensor into a covalently linked stable fluorescent RNP sensor. The covalently linked sensors enabled ligand detection at a low sensor concentration, even in cell extracts. Furthermore, application of different target specific sensors allowed simultaneous detection of different targets monitoring each wavelength corresponding to the respective sensor. Further application of this technique will be shown in chapter 3.

2.6.Materials and methods

2.6.1. Materials

Tetracycline, dopamine, norepinephrine and tyrosine were purchased from Sigma Aldrich. Chloramphenicol, gel electrophoresis grade acrylamide and bisacrylamide was provided by Wako Chemicals GmbH. Streptomycin was obtained from Nacalai Tesque. N-Fmoc-protected amino acids, Fmoc-NH-SAL-PEG resin (0.23mmol/g), HBTU (2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate), HOBt (1-hydroxybenzotriazole), DIEA (N,N-diisopropylethylamine), TFA (trifluoroacetic acid) and distilled DMF (N,N-dimethylformamide) were purchased from Watanabe Chemical Industry. Gel electrophoresis grade acrylamide and bisacrylamide were obtained from Wako Chemicals. All other chemicals were reagent grade and used without further purification.

2.6.2. Nucleic acids synthesis

The original double-stranded DNA pools were constructed by Klenow polymerase (New England Biolabs) reaction from a synthesized oligonucleotide containing 40 random nucleotides [5'-GGAATAGGTCTGGGCGCA-N40-P(OH)-3'] and a synthesized oligonucleotide containing 7 random nucleotides (5'-GGAATAGGCCTGTACCGTCA-N7-OH-3') followed by PCR amplification to add the promoter for T7 RNA polymerase using Pyrobest DNA polymerase (TaKaRa) with 3'-DNA (5'-GGAATAGGCCTGTACCGTC-3') and 5'-DNA primer (5'-TCTAATACGACTCACTATAGGAATAGGTCTGGGCGCGCA-3': T7 RNA promoter underlined). RNA transcription was performed using an AmpliScribeT7 kit (Epicenter) for 3h at 37ºC, according to the supplier's recommended protocols. The resulting RNA was phenol/chloroform extracted, precipitated with ethanol, and pelleted by centrifugation. The RNA was purified by denaturing polyacrylamide gel electrophoresis and eluted. RNAs concentration was determined by UV spectroscopy.

2.6.3. Preparation of ribonucleopeptide receptors

RNPs that bound ATP were selected as follows: RNA was heated at 80ºC for 3 minutes and cooled to room temperature for 2h for proper secondary structure. A binding buffer (100µL, 10nM Tris-HCl pH 7.60, 100mM NaCl, 10nM $MgCl_2$) containing 1.0µM RNA, 1.5µM Rev, and a 50µL volume of ligand-conjugated agarose was incubated to allow a formation of a specific ribonucleopeptide complex for 30 minutes on ice. RNA-peptide-resin complexes were washed three time with 300µL (6-volume of resin) of binding buffer to remove unbound RNA-peptide complexes and eluted three times with 150 µL (3-volume of resin) of binding buffer containing 4mM ATP. Recovered ribonucleopeptide complexes were precipitated with ethanol and resuspended in TE buffer. After reverse transcription with AMV reverse transcriptase (Promega) of the selected RNA using the 3'-DNA primer used in PCR amplification and successive PCR amplification (RT-PCR) using 5'- and 3'- DNA primer, DNA templates were transcribed, and the resulting RNAs were subjected to the next round of selection. Selected RNA pools were converted to DNA and PCR-amplified to introduce BamHI, EcoRI restriction sites by using primers 5'-GCGGGATCCTTTCGGCCTGTACCGTCA-3' and 5'-CGGAATTCTAATACGACTCACTATAGG-3'. After enzymatic digestion (New England Biolabs), DNAs were cloned into the pUC19 vector using Ligation Kit Version 2 (TaKaRa) and sequenced using a BigDyeTerminator Cycle Sequencing Kit (Applied Biosystems) with a model 377DNA sequencer and 3130/3130x1 genetic analyzer (Applied Biosystems).

2.6.4. SNARF-Rev and SNARF-Abu-Rev preparation

The procedure for the synthesis of fluorophore-labeled Rev peptides were previously reported.[3] Briefly, the two Rev peptides were synthesized on a Shimadzu PSSM-8 peptide synthesizer according to the Fmoc chemistry protocols. In SNARF-Abu-Rev two residues of γ-aminobutyric acid were manually attached to the N-terminus of the peptide applying Fmoc-SPPS strategies (HOBt, HBTU as activators, DIEA as base). SNARF, which was synthesized according to previous report[7], was directly coupled, after activating the carboxyl function, to the N-terminal deprotected residue of Rev peptide on the resin. SNARF-Rev and SNARF-Abu-Rev peptides were then deprotected, cleaved from the resin and purified by using RP-HPLC. Characterization was achieved by MALDI-TOF MS spectrometry (AXIMA-LNR, Shimadzu Biotech) $[6\text{-carboxy-SNARF-Rev+H}]^+$ calc. m/z 3314.12 obs.m/z 3313.35; $[6\text{-carboxy-SNARF-abu-Rev+H}]^+$ calc. m/z 3729.38 obs. m/z 3728.23.

6-carboxy-SNARF-Rev

SNARF-TRQARRNRRRRWRERQRAAAAR-NH_2

The SNARF-Rev peptide purification was achieved using HPLC methodology. The gradient optimization was settled in analytical conditions using Hypersil GOLD 250x4.6 column:

- Flow rate: 1.000 ml/min;
- Sample preparation: 50% distilled H_2O+ 50% ACN;
- Wavelength detected: 548nm;
- Gradient selected (0.5%):

Time (min.)	%A	%B
3	75	25
53	50	50
58	0	100
60	75	25
65	75	25

Table 4: A= distilled H_2O + 1‰TFA; B= 50%ACN + 1‰TFA

Purification was accomplished using COSMOSIL ARII column (4.6x250mm) and the presence of the desired compound was confirmed by MALDI-TOF mass spectrum (*m/z* 3313.35; calcd for $[M+H]^+$):

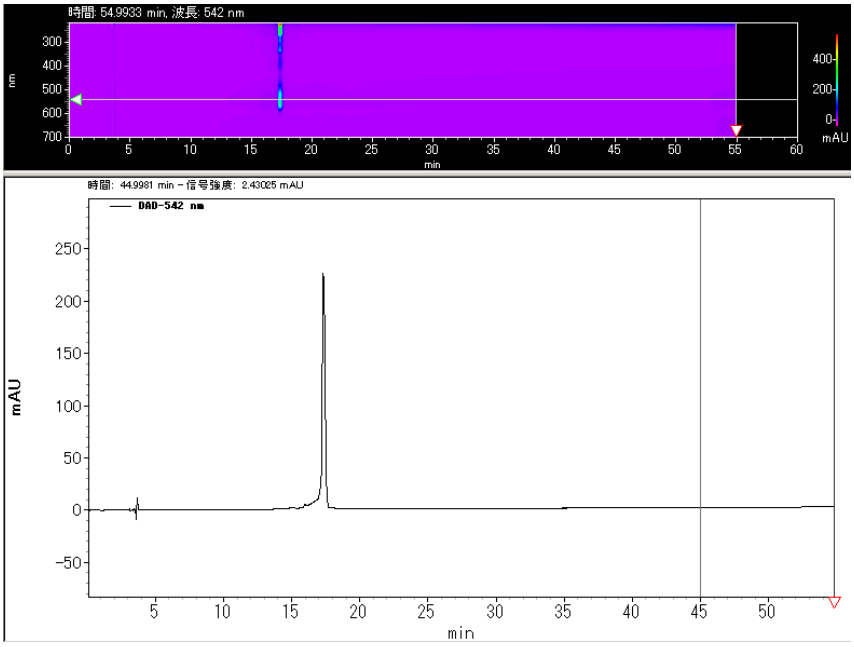


Figure 32: Hplc chromatograph of purified 6-SNARF-Rev peptide

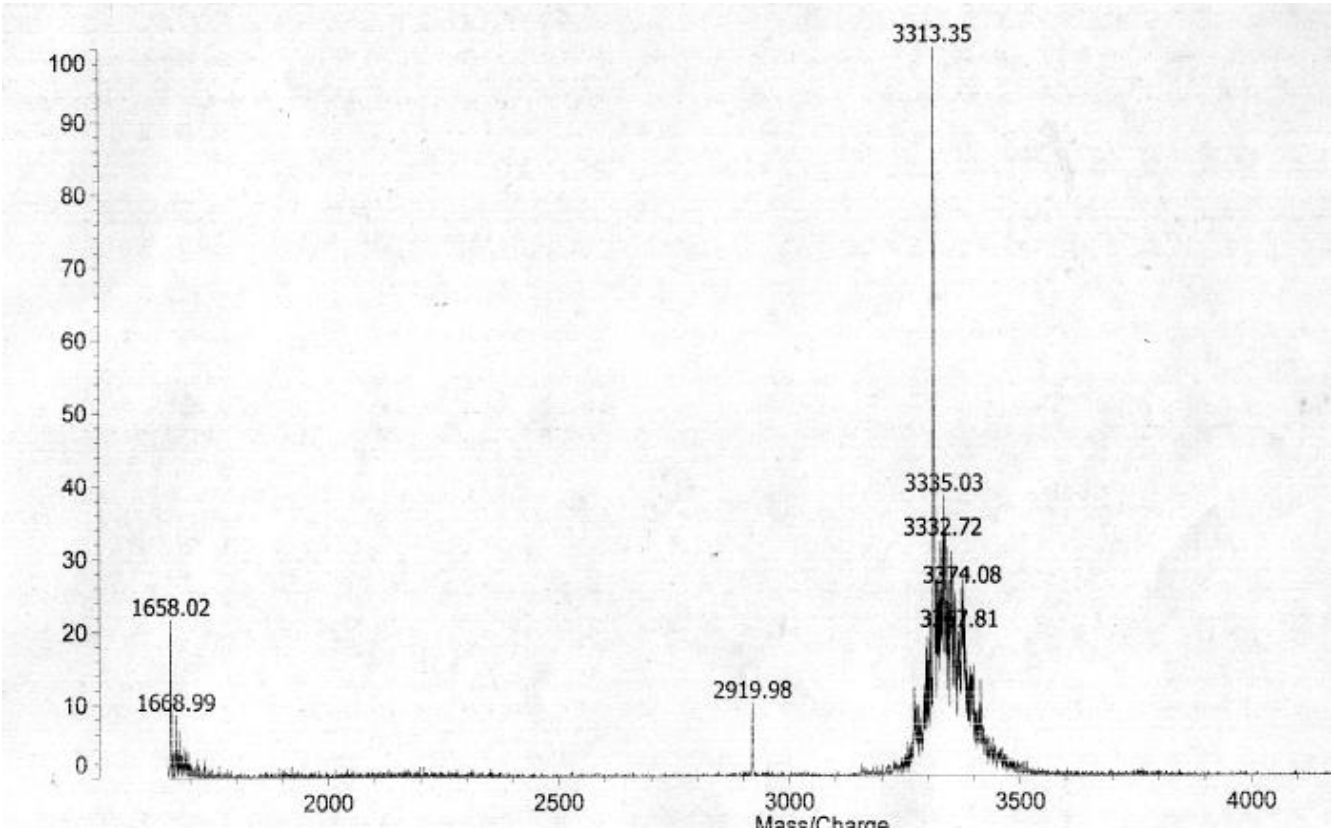
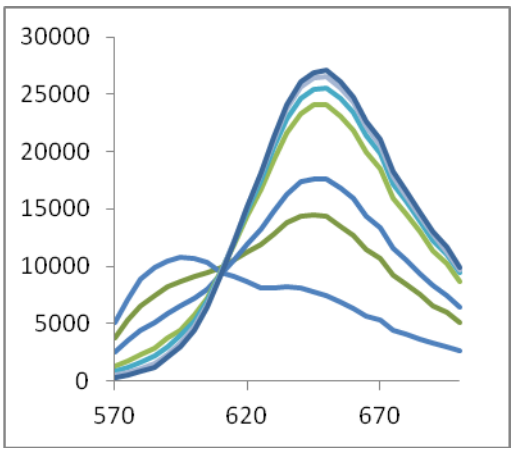
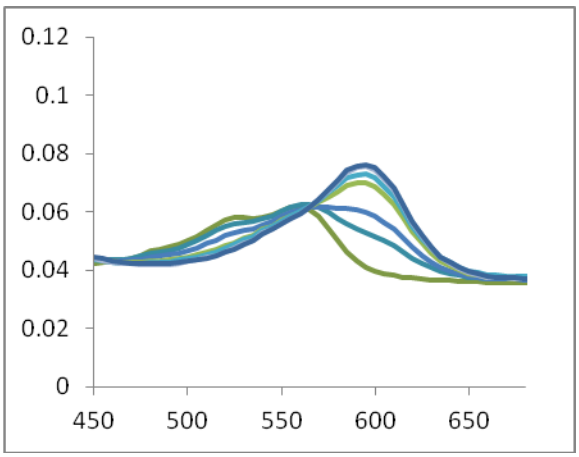


Figure 33: MALDI-TOF mass spectrum of 6-SNARF-Rev peptide

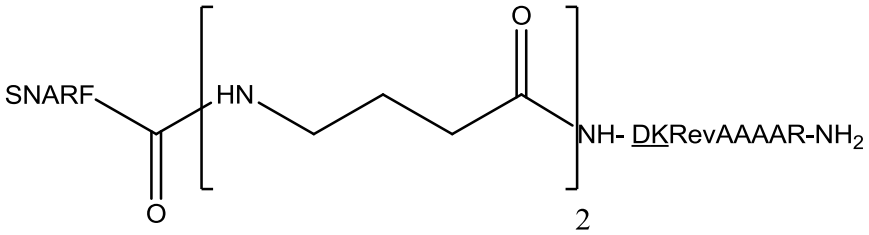
In order to verify SNARF pka change upon Rev peptide coupling, a fluorimetric assay was accomplished screening several pH conditions:

- Sample concentration: 5μM;
- pH titration: 5.13, 6.07, 6.53, 7.11, 7.54, 8.03, 9.02;
- Emission range: 570-700nm (excitation wavelength 534, 514, 488nm)
- Absorption range: 280-800nm





6-carboxy-SNARF-Abu$_2$- Rev



Purification was achieved by HPLC COSMOSIL AR II column, applying a 0.5% gradient:

Time (min.)	%A	%B
3	75	25
53	50	50
55	0	100
57	75	25
67	75	25

Table 5: Phase A: distilled H_2O +1 ‰ TFA; phase B: 50%ACN + 1‰ TFA. Flow: 1.000ml/min; 100μl injected; wavelength checked: 548nm.

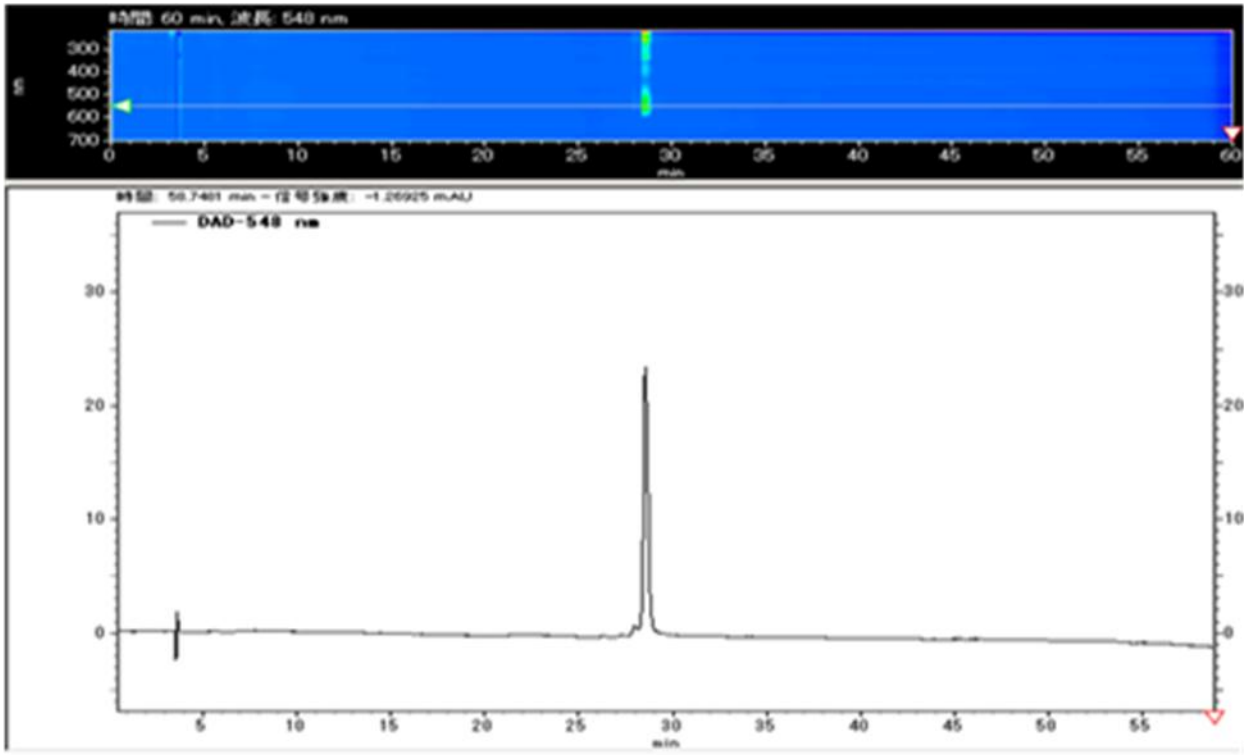


Figure 34: Hplc chromatograph of purified 6-SNARF-Abu$_2$ -Rev peptide.

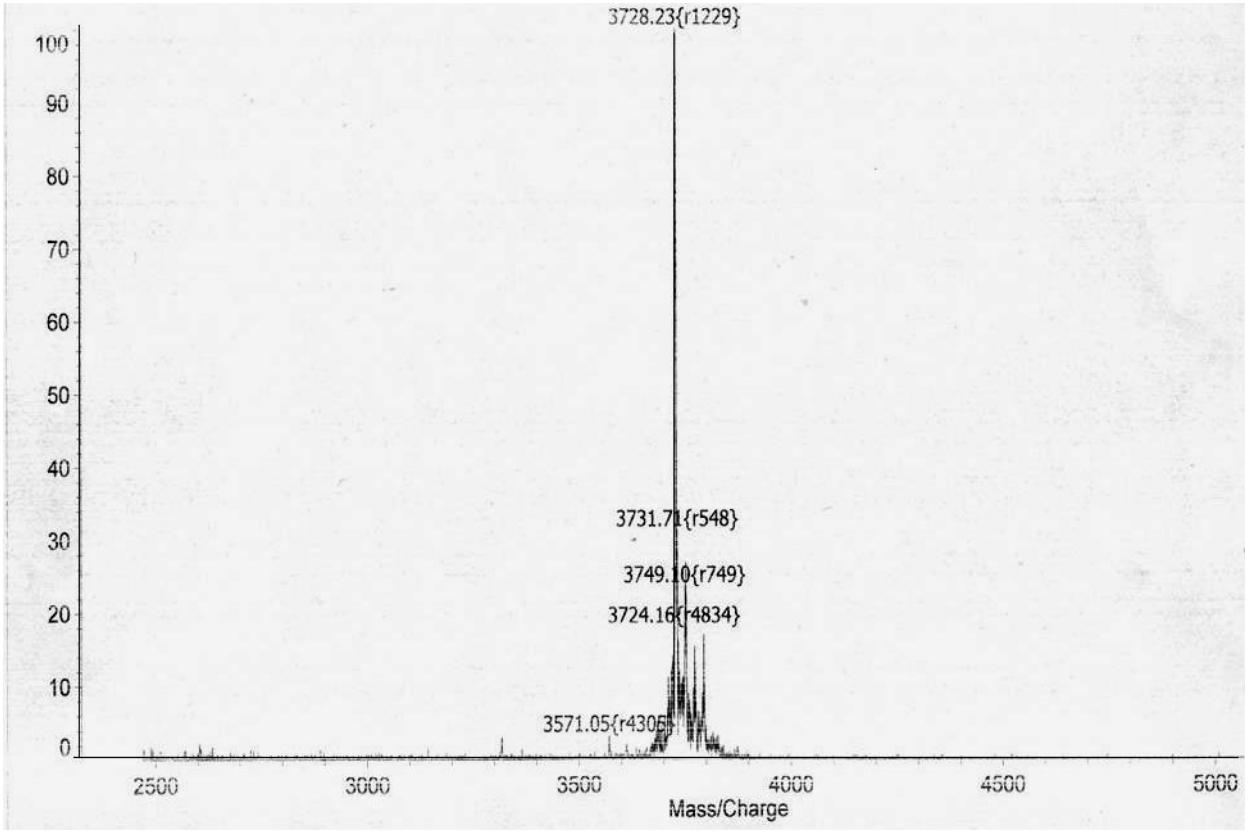
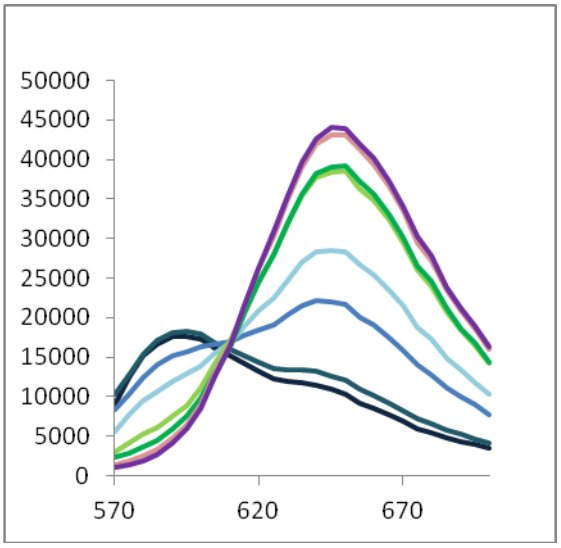
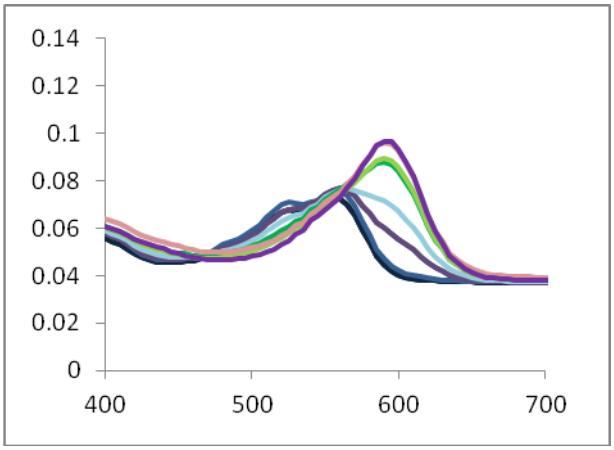


Figure 35: MALDI-TOF mass spectrum of 6-SNARF- Abu_2- Rev peptide.

The fluorescence emission spectrum of the obtained compound was determined at different pH conditions, using Infinite M200PRO instrument:

- Excitation wavelength: 534, 514, 488nm;
- Emission scan range: 570-700nm;
- Wide pH range buffer: Acetate + Mes + Tris
- pH titration: 4.02, 5.13, 6.07, 6.53, 7.11, 7.54, 8.03, 9.02;
- Samples final concentration: 5µM;
- Absorbance range: 280-800nm.





2.6.5. Fluorescent measurements on the microplate

The 96-well fluorescence measurements were performed on an Infinite M200PRO instrument (TECAN co. ltd.). For tetracycline titration, a binding solution (100µl) containing 1µM of fluorescent RNP in 10mM Tris HCl (pH 7.00), 250mM NaCl, 5mM $MgCl_2$, 0.0005% Tween 20 was prepared. For dopamine and other catecholamine derivatives, a binding solution (100 µl) containing 1µM fluorescent RNP in 10mM Tris HCl (pH 7.60), 300mM NaCl, 5mm $MgCl_2$, 0.005% Tween 20 was prepared. For streptomycin, a binding solution (100µl) of 1µM concentration of fluorescent RNP was prepared in 50mM Tris HCl (pH .60), 250mM NaCl, 10mM $MgCl_2$, 0.005% Tween 20 buffer conditions. Well-mixed samples with different concentrations of substrates were incubated at 25ºC for 30 minutes followed by the measurement of emission spectra (measures emission spectrum range: 570-700nm; excitation wavelength utilized: 534nm).

2.6.6. Determination of dissociation constants

The dissociation constant was achieved applying the following equation:

$$F_{obs} = A(([RNP]_T + [substrate]_T + K_D)-(([RNP]_T + [substrate]_T + K_D)^2 -(4[RNP]_T [substrate]_{T)}^{1/2})/2[RNP]_T)$$

where A is the increase in fluorescence at saturating substrate concentrations (R_{sat}-R_0), K_D is the equilibrium dissociation constant, $[RNP]_T$ and $[substrate]_T$ are the total concentrations of RNP and the substrate, respectively.

2.6.7. Cross linking reaction for constructing covalently linked RNP sensors

RNA subunits of RNP receptors were purified by denaturing 8% polyacrylamide gel electrophoresis. The purified RNA subunit was treated with freshly prepared sodium periodate to convert cis-diol of the 3'-terminal ribose to 3'-dialdehyde by periodate oxidation. Freshly prepared 50eq. of 0.01M sodium periodate (10µL; 100nmol) was added to about 100µM RNA (20µL; 2nmol) in 32µL of 0.02M sodium acetate (pH 5.20) and the reaction mixture was incubated for 30minutes 37ºC in the dark. After the reaction, the excess of sodium periodate was reduced by adding 1M glycerol and the resultant oxidized RNA was purified with ethanol precipitation. In order to assess the oxidation reaction success, 5 µl sample was mixed with 100µl of a BODIPY-$NHNH_2$ solution (1mg/ml in TE buffer pH 7.60) and let reacting for 1hour at room temperature in dark conditions. Sample was then purified by EtOH precipitation and checked by gel image (Pharos imager) (**Figure 36**).

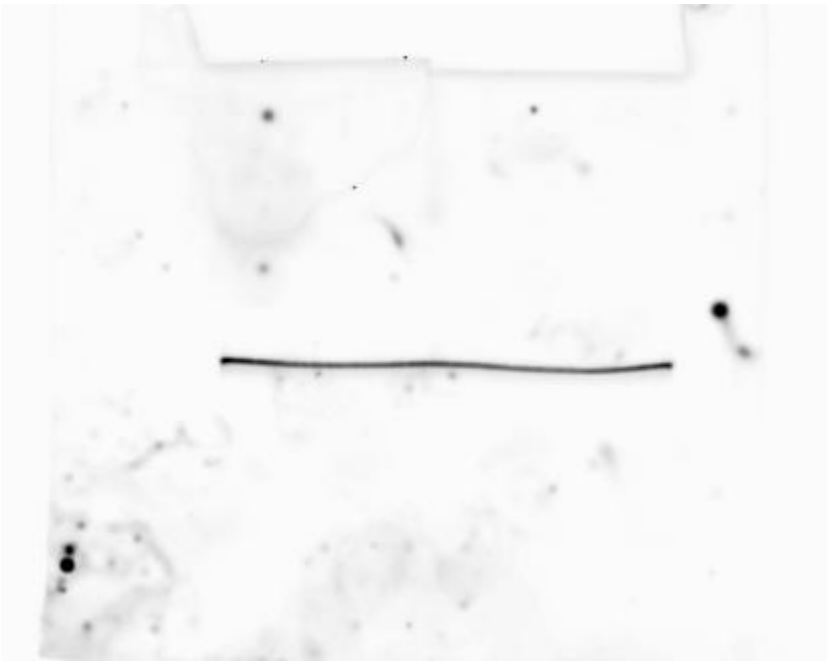


Figure 36: Oxidation step check on A26 by BODIPY-$NHNH_2$ (Pharos Imager, Exc. Wv. 488nm)

The peptide subunit was synthesized by introducing a fluorophore to the N-terminus and the peptide linker (GGSGGSGGSG) with a hydrazide group synthesized as follows, 4-hydroxymethylbenzoic acid PEG resin (HMBA-PEG resin, 0.22mmol/g) was placed in a dry flask and was added sufficient amount of DMF to soak the resin and allowed to swell for 30 minutes. N-α-Fmoc-glycine (10eq. relative to resin loading) was dissolved in DMF, and then mixed DIC (5eq. relative to resin loading). After incubating for 20minutes, the solution is mixed to the resin prepared above and a catalytic amount of DMAP is added. The resin/amino acid mixture was incubated at room temperature for 1 hour with occasional swirling. The first coupling was repeated 2 times. The resin with the first amino acid was loaded onto an automated peptide synthesizer PSSM-8 (Shimadzu) and the subsequent synthesis was performed according to the Fmoc chemistry protocols by using protected Fmoc-amino acids, and HBTU. A fluorophore with an activated group (5-carboxyfluorescein N-succinimidyl ester or Cy-5 mono NHS ester) was directly coupled to the Fmoc-deprotected synthetic peptide on the resin. The fluorophore-labeled peptide was cleaved from the resin with 0.1M hydrazine hydrate in DMF for 1 hour at room temperature, washed with DMF and then deprotected by a solution containing phenol (0.75g), distilled water (0.5mL), thioanisol (0.5mL), 1,2-ethandiol (0.25mL), and TFA (10mL) for 3

hours. The modified peptide was purified by reverse phase HPLC (**Figure 37**), and characterized by MALDI-TOF mass spectrometry (AXIMA-LNR, Shimadzu Biotech); 5-carboxyfluorescein modified Rev peptide hydrazine (5FAM-Rev-$NHNH_2$), *m/z* 3485.87 (calcd for $[M+H]^+$ Na^+3484.7) (**Figure 38**).

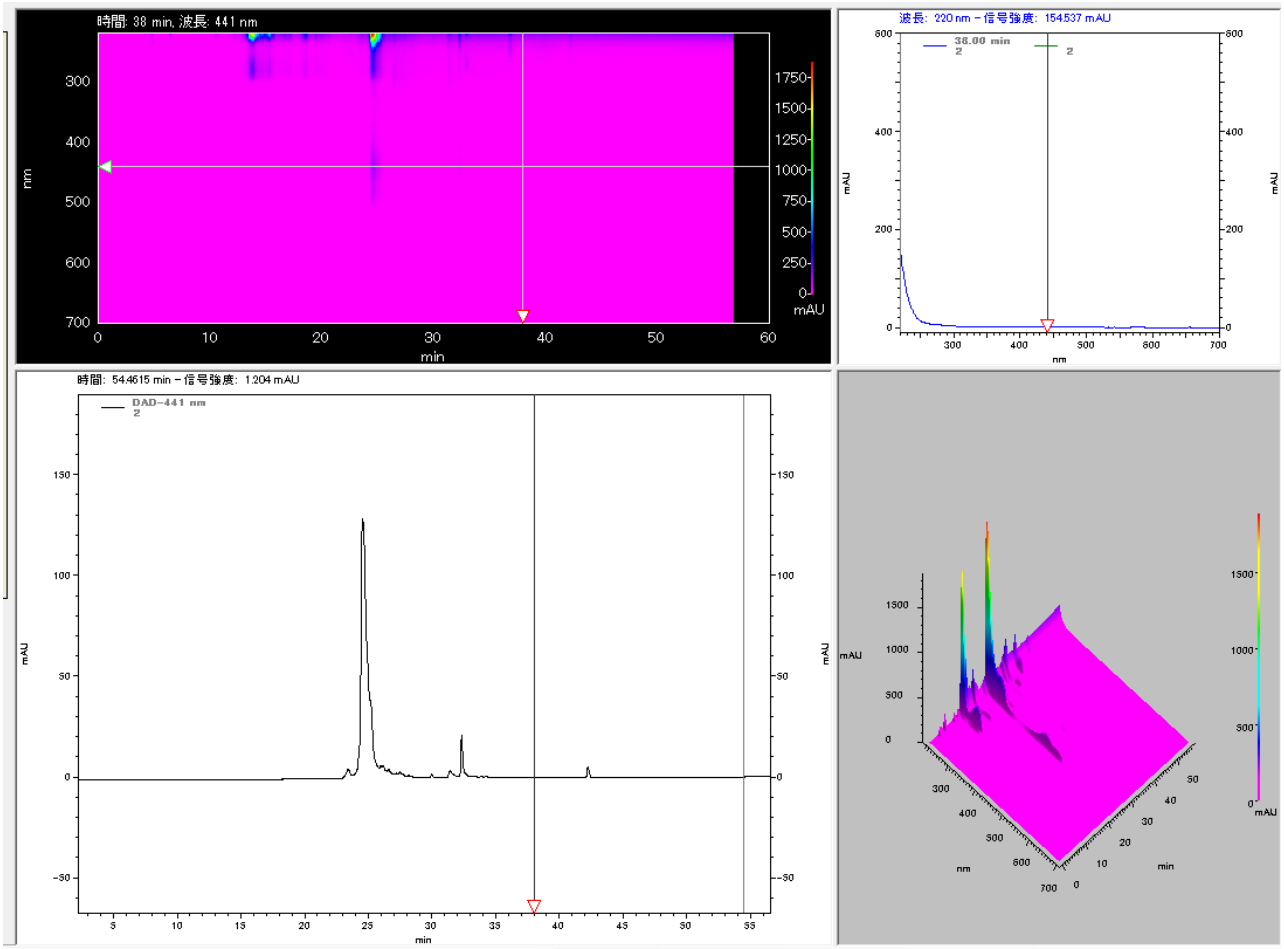


Figure 37: Hplc chromatogram of 5-FAM-Rev-GGSGGSGGSG-$NHNH_2$ peptide

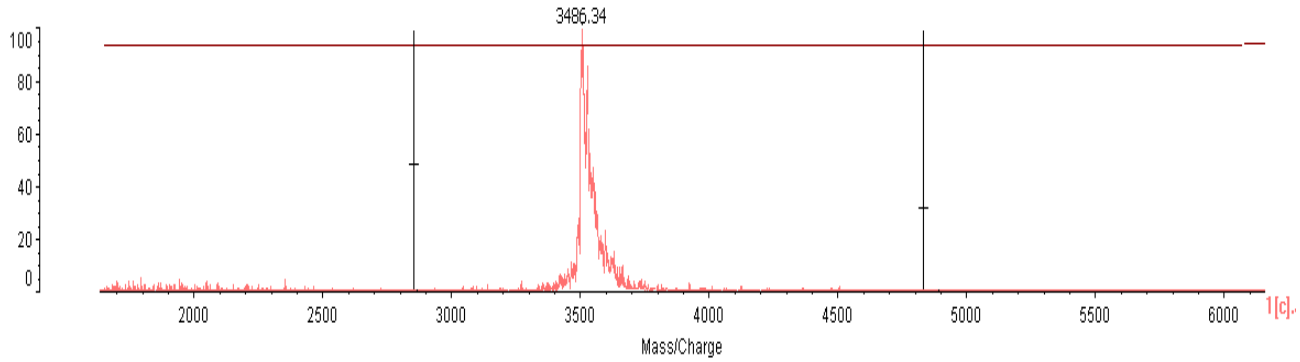


Figure 38: MALDI-TOF mass spectrum of 5-FAM-Rev-GGSGGSGGSG-$NHNH_2$ peptide

The coupling reaction between the 3'oxidized RNA and fluorophore labeled Rev-$NHNH_2$ (1.5eq.) was carried out in 0.02M sodium acetate buffer pH 5.20 containing 0.01M NaCl (total 50µL) at 37ºC in the dark. After 8 hours, the reaction mixture was extracted by phenol-chloroform and then purified by ethanol precipitation. The precipitate was dissolved in TE buffer. The sample solution was purified by 6M urea 15% denature PAGE, and was subsequently quantified by the absorption at 260nm (A26 ε= 738400 $M^{-1}cm^{-1}$, A17 ε= 695700 M^{-1} cm^{-1}, An15 ε= 612200 M^{-1} cm^{-1}, An16 ε=635500 M^{-1} cm^{-1}); 80% yield was recovered (**Figure 39** and **40**).

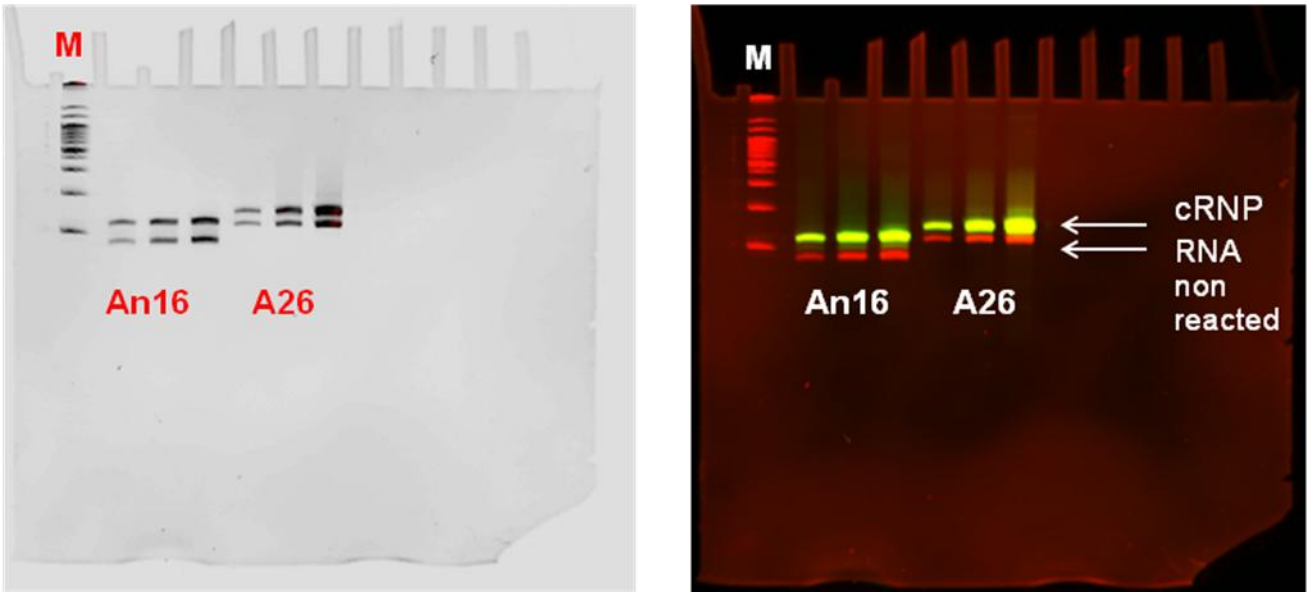


Figure 39: 15% urea acrylamide gel image of An16 and A26 cRNPs before purification. On the left side EtBr staining; on the right side fluorescence image (exc. Wv. 488nm) and EtBr staining overlaid image.

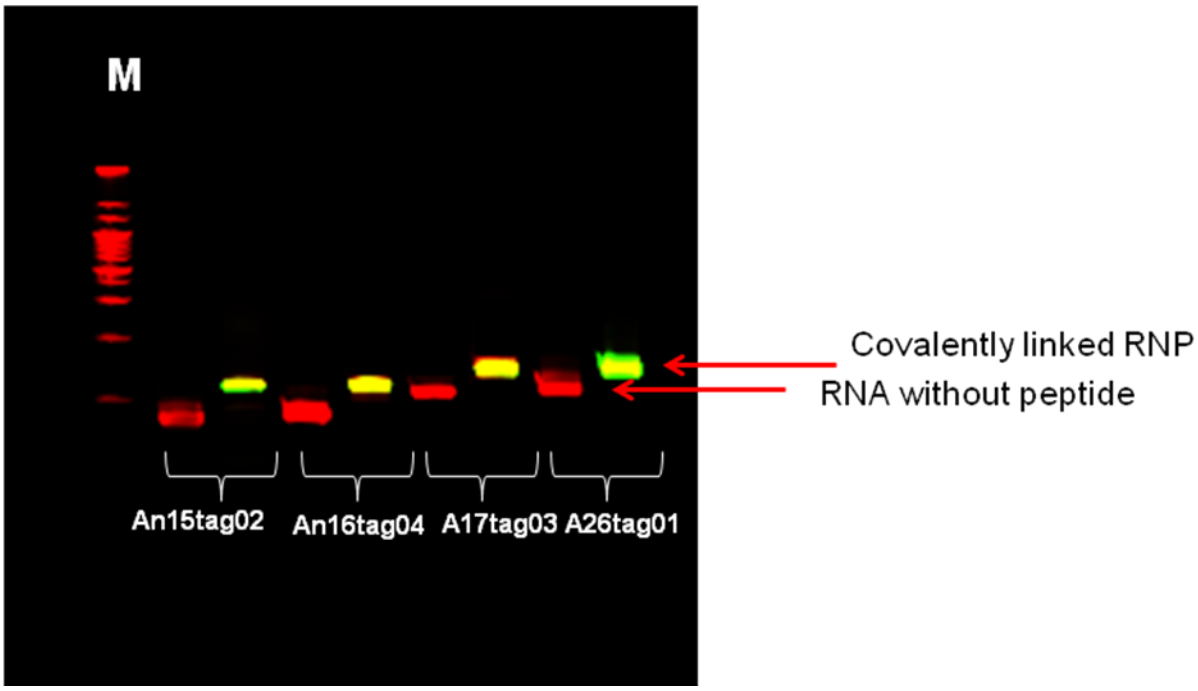
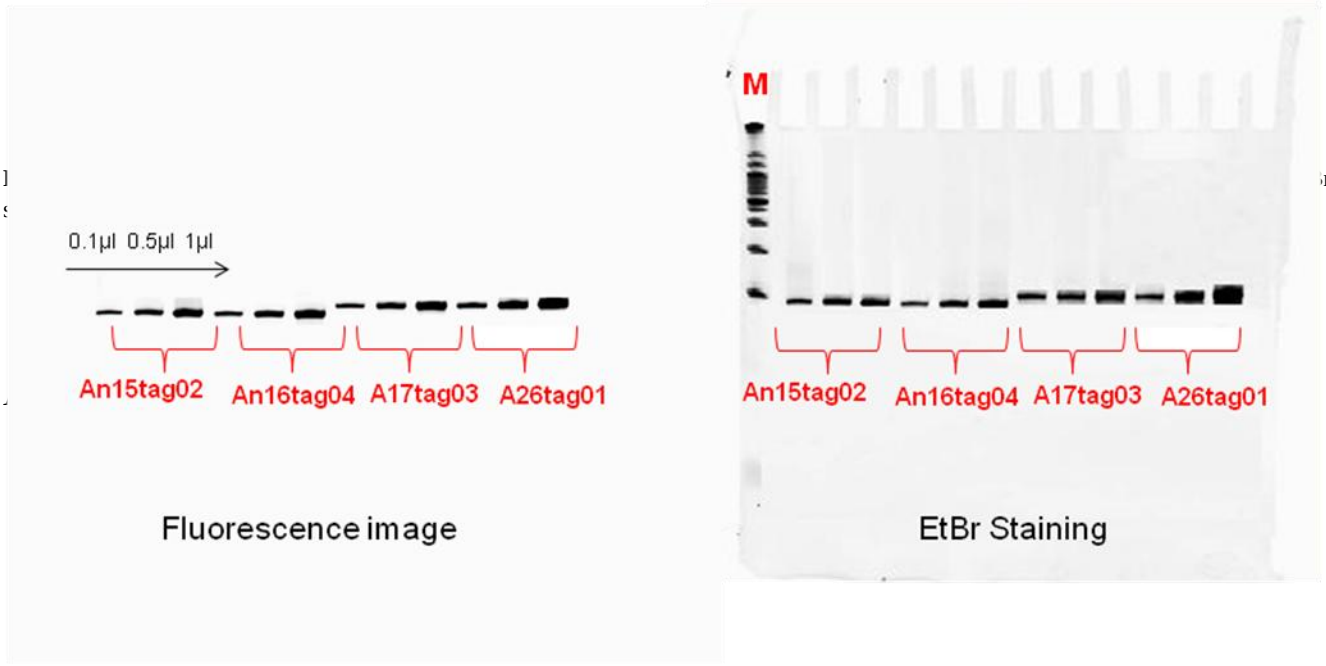


Figure 40: Mobility shift variation after coupling reaction; 15% urea acrylamide gel image of covalently liked RNPs, respectively from left to right, An15tag02, an16tag04, A17tag03 and A26tag01.

The purity was confirmed by gel image (Pharos Imager, Exc. Wv. 488nm/EtBr staining) as shown in **Figure 41 (**Yield after purification: .A26 82%; An16 75%; An15 67%; A17 66%). A mass spectrum was also obtained using MALDI-TOF sprectometer (**Figure 42** and **43**).



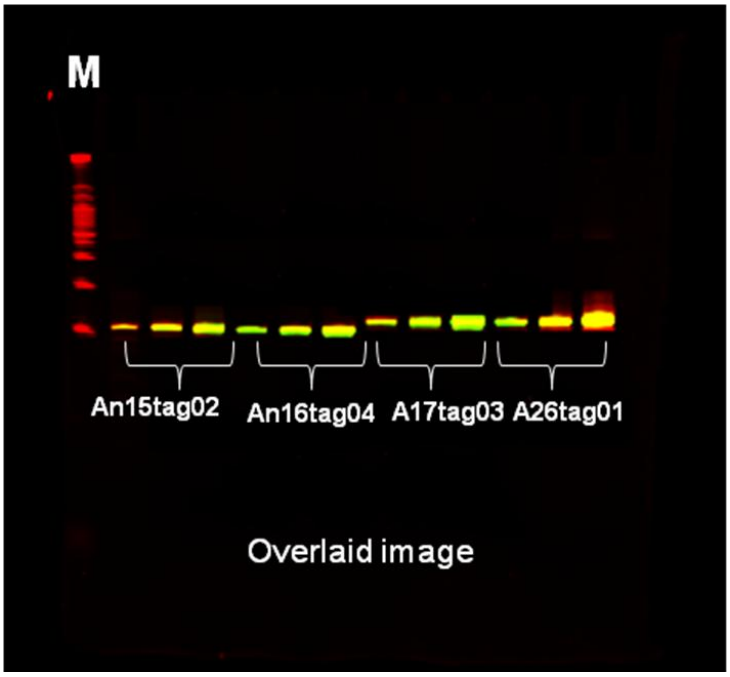


Figure 41: Covalently linked RNPs purity check. From left to right: An15tag02, An16tag04, A17tag03, A26tag04.

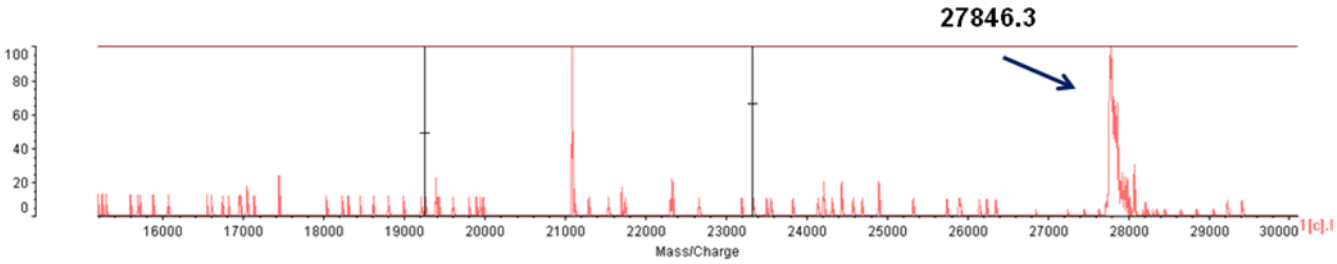


Figure 42: MALDI-TOF mass spectrum of A26tag01 cRNP (Mass founded: *m/z* 27846.3).

An16 cRNP calcd. mass *m/z* 24538.6

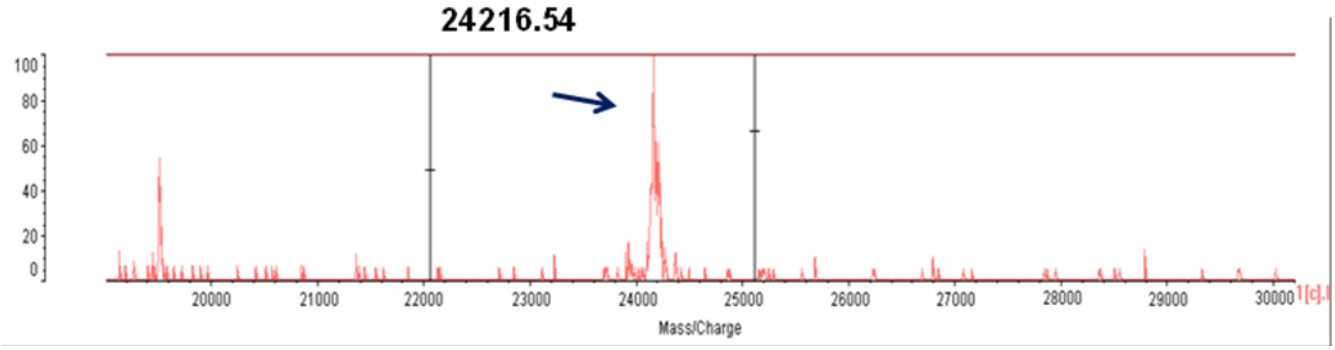


Figure 43: MALDI-TOF mass spectrum of A26tag01 cRNP (Mass founded: *m/z* 24216.

2.6.8. Fluorescence spectral measurements of RNPs

Fluorescence spectra were recorded on a Hitachi F-4500 fluorescence spectrofluorophotometer with excitation and emission bandwidth of 5nm. The measurements were performed in a buffer (100µL) containing 10nM Tris-HCl (pH 7.60), 100mM NaCl, 10mM $MgCl_2$, 0.005% Tween 20, non-covalent or covalent RNP and an indicated amount of substrate. Excitation wavelength was 485nm for 5FAM-Rev and 350nm for Pyr-Rev. The emission wavelength was 535nm for 5FAM-Rev and 390nm for Pyr-Rev were used for the determination of affinity and relative fluorescence intensity.

Chapter 3:

Construction of a molecular switchboard by assembling ribonucleopeptide biosensors

3.1.Introduction

3.1.1. Nanomaterial based on DNA

DNA is one of the most important components of all living organisms. Genomic DNA is the archive of genetic information, which comprehend fundamental information for many species including humans, even though the sequence has always been continuously changing. This instability in terms of over millions of years has been changing the organisms, and some of those changes have allowed them to adapt to new environments and some others have resulted in their disappearance. For example, humans, other primates and guinea pigs have lost the ability to produce ascorbic acid, or vitamin C,[1] even though they require the vitamin for healthy living. Genetic mutations sometimes cause diseases, even when only one base is deleted, or replaced by a different base if the change is critical to the expression of a functional protein at a normal level.[2-5] Most of mutations in genomic DNA are corrected by enzymatic systems to keep the information unchanged.[6,7] The double stranded structure of DNA makes this correction process precise, because the strand without mutations is used as a template for the correction. From the structural point of view , there are many advantages in using DNA as a building block of nano-sized material with precisely designed architecture.[8-13] In particular, the specificity of the hydrogen interactions between adenine-thymine and guanine-cytosine allows us to design the whole structure of DNA-based materials. In addition, two strands of double-stranded DNA can be covalently bound to each other using DNA ligase when the 5'-end of one strand is complimentary to the 3'-end of another strand. Oligomers or polymers of DNA can be chemically synthesized by automated machines in a large amount, so that DNA with a proper length and sequence can be obtained with high purity in amounts required for clinical application. Under conditions where there are no enzymes degrading DNA, DNA is very stable compared with proteins, another component candidate for nano-sized materials with designed architecture. Moreover, restriction enzymes, ligases and other enzymes are available to design and modify the properties of DNA-based materials. DNA materials science related to structural DNA nanotechnology has grown rapidly as a research area in the past two decades. The technology allows the construction of various self-assembled scaffolds, which can be used for the placement and arrangement of functional molecules and nanomaterials and for the production of complex molecular devices. The field of DNA nontechnology can be traced back to the words written by Nadrian Seeman in 1982: " It is possible to generate sequences of oligomeric nucleic acids which will preferentially associate to form migrationally immobile junctions, rather than linear duplexes, as

they usually do."[14] Seeman had wanted to organize proteins in three-dimensional (3D) crystals so that he could study their structure with X-ray crystallography. Three decades later the field has outgrown its roots in protein crystallography and delivered numerous advances in the control of matter on the nanoscale. [15]

3.1.2. References

1) Roy, R. N., Guha, B. C., *Nature* **1985**, 182, 319-320.
2) Roberts, R. G., Bobrow, M., Bentley, D. R., *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 2331-2335.
3) Sutherland, G. R., Richards, R. I., *Proc. Natl. Sci. U.S.A.* **1995**, 3636-4361.
4) Weatherall, D. J., *Nat. Rev. Genet.* **2001**, 2, 245-255.
5) Jones, P. A., Baylin, S. B., *Cell* **2007**, 128, 683-692.
6) Modrich, P., Lahue, R., *Annu. Rev. Biochem.* **1996**, 65, 101-133.
7) Kunkel, T. A., Erie, D. A., *Annu. Rev. Biochem.* **2005**, 74, 681-710.
8) Seeman, N. C., *Trends Biotechnol.* **1999**, 17, 437-443.
9) Seeman, N. C., *Chem. Biol.* **2003**, 10, 1151-1159.
10) Seeman, N. C., *Nature* **2003**, 421, 427-431.
11) Feldkamp, U., Niemeyer, C. M., *Angew. Chem. Int. Ed. Engl.* **2006**, 45, 1856-1876.
12) Condon, A., *Nat. Rev. Genet.* **2006**, 7, 565-575.
13) Simmel, F. C., *Angew. Chem. Int. Ed. Engl.* **2008**, 47, 5884-5887.
14) Seeman, N. C., *J. Theor. Biol.* **1982**, 99, 237-247.
15) Endo, M., Yang, Y., Sugiyama, H., *Biomater. Sci.* **2013**, 1, 347-360.

3.2.DNA origami technology development

3.2.1. 2D DNA origami

DNA origami, developed by Rothemund in 2006, has enabled the construction of a wide variety of 2D structures of around 100nm in size, including rectangles, triangles, and even a smiley face and five-pointed star (**Figure 1**).[1]

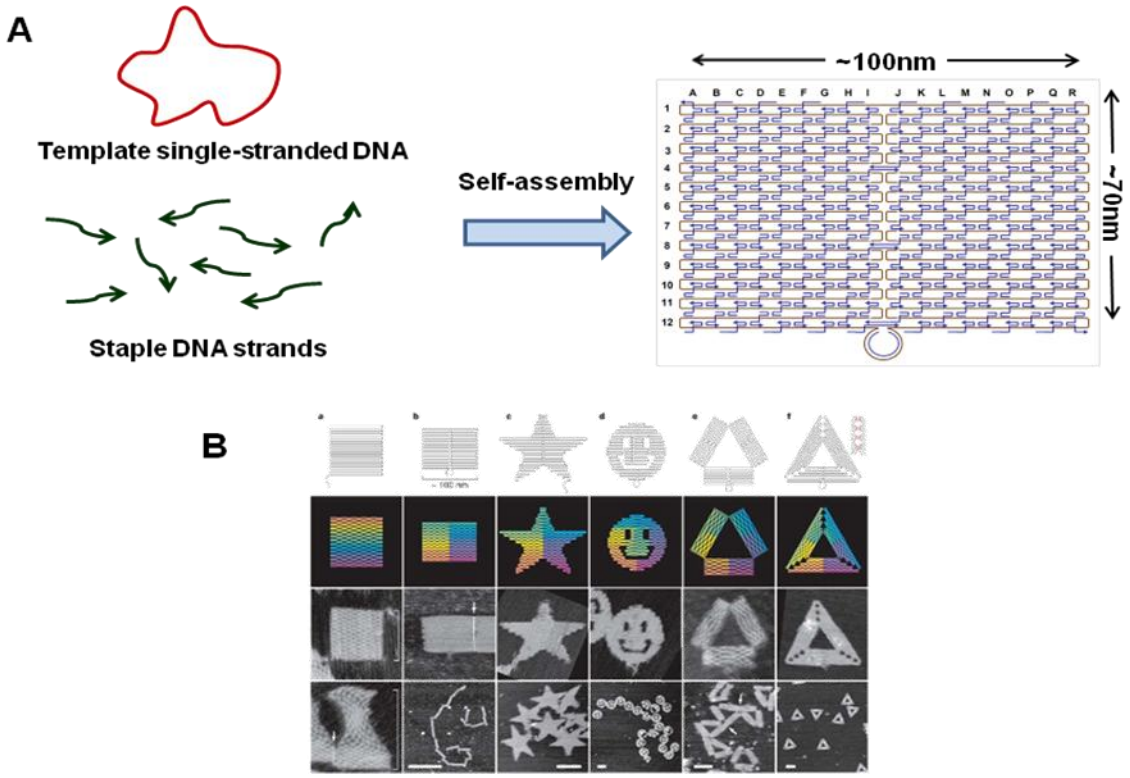


Figure 1: DNA origami structure. A) The method employed to prepare DNA origami structure from the template single-stranded DNA and staple strands. B) Design and AFM images of self-assembled DNA origami structures.

In this method, a long single-stranded DNA (M13mp18; 7249 nucleotides) and the sequences-designed complementary strands (called "staple strands"; most are 32-mer) are mixed and then annealed from 95ºC to room temperature over 2h, resulting in the formation of target structures by self-assembly (**Figure 1**). The structure can be imagined using atomic force microscopy (AFM), and the assembled structure can be formed according to a design. To create 2D DNA origami structures, adjacent double-stranded DNAs (dsDNAs) should be connected to each other *via* a cross-over. In this design, the geometry of the double helices involved has three helical rotations fro 32 base pairs (**Figure 2**). For

example, two neighboring crossovers of the central dsDNA in an arrangement of three adjacent dsDNAs should be located at the opposite sites (180º rotated, 0.5turns).

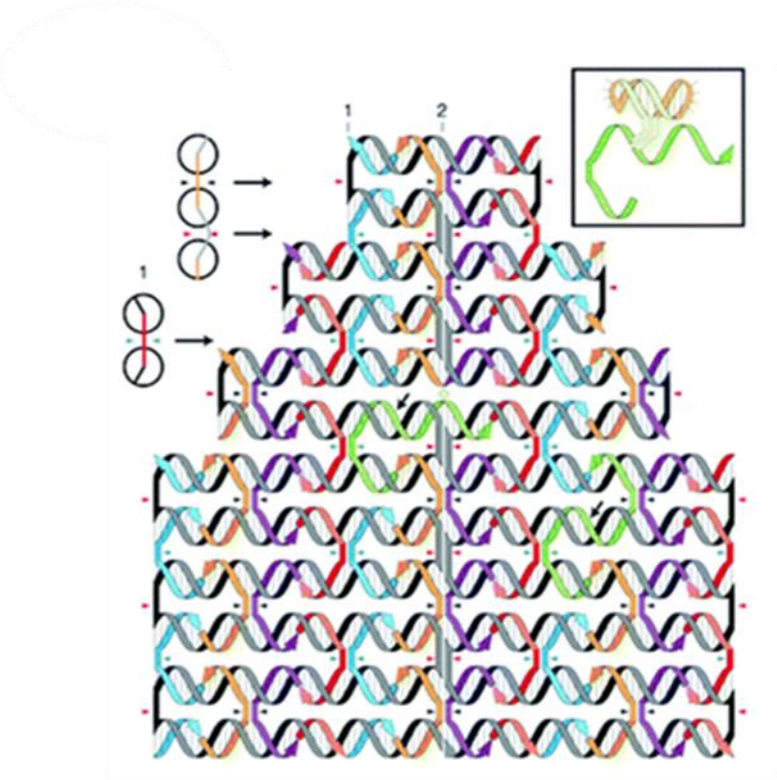


Figure 2: Design of a self-assembled DNA origami structure and geometry of the incorporated dsDNAs. Colored strands and a gray/black strand represent staple strands and template single-stranded DNA, respectively. Staple strands connect the adjacent duplexes with crossover. Inset: structure of harpin DNA for a topological marker.

Therefore, the crossovers should be separated by 16 base pairs (1.5 turns). To maintain a stable planar structure, this rule should be followed when placing multiple staple strands. DNA origami structures are formed using many different staple strands, so that hairpin DNA markers can be placed at any position on the surface of the DNA structure. Hairpin DNA used as a topological marker is observed as a dot under AFM imaging. In this case, hairpins are placed perpendicular to the surface of DNA origami, therefore each hairpin should be placed at a position eight base pairs from the crossover (270° rotation). The distance between the centers of the adjacent staples is about 6 nm, so the adjacent hairpins can be observed as distinct spots according to the spatial resolution of the AFM. Using the hairpin markers, patterns, such as the map of a hemisphere (**Figure 3**) can be drawn precisely on the DNA origami surface.

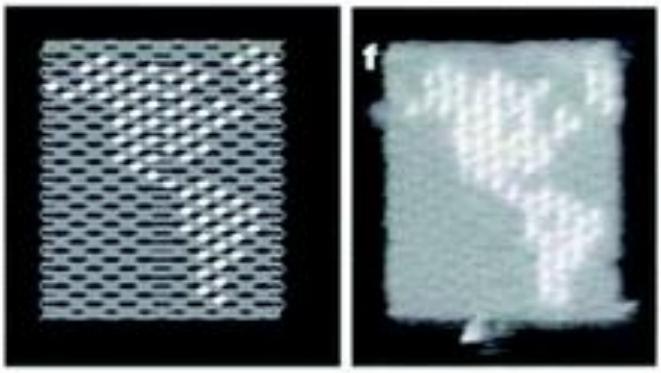


Figure 3: Drawing of a hemisphere on the DNA origami with hairpin DNAs (with dots) and an AFM image of the assembled DNA origami.

In addition, when functional molecules and nanoparticles are conjugated to specific staples, they can be placed on the origami surface at selected sites. A remarkable property of DNA origami is that all the positions of the structure have DNA sequence information (an address). Before the emergence of DNA origami technology, it was difficult to create ∼100 nm-sized structures through the self-assembly of small DNA components. DNA origami technology solved this problem and provided the breakthrough that enabled expansion of the shape design and the creation of addressable structures. Since the DNA origami system uses a long single-stranded template DNA, the size of the 2D structure is determined by the length of the template strand. Various single-stranded DNAs were isolated and used as a template in the preparation of DNA origami.[2,3] A strategy of using DNA tiles (17 × 16 nm) instead of staples has also been developed, which allows size expansion by the introduction of 25 to 56 DNA tiles.[4]

3.2.2. Programmed arrangement of multiple DNA origami components

The programmed arrangement of DNA origami is an important technique for preparing the desired large structures, particularly in terms of integrating complicated functions. Sugiyama et al. explored techniques for arranging multiple DNA origami components, and developed methods to arrange rectangular DNA origami tiles horizontally in a programmed fashion.[5] In addition, as the ends of the helical axes align at both edges of the DNA origami rectangles, a rectangular DNA origami tiles can be assembled *via* π-interaction at the edges. Specific concave and convex connectors were introduced into the origami rectangles to align these rectangles precisely with neighboring origami tiles. DNA rectangles should correctly assemble by shape and sequence complementarity, where complementary strands are introduced into the concavity and the convex connectors. After self-assembly, it was observed that the DNA tiles were aligned and oriented in the same direction. Furthermore, to align origami tiles accurately, the positions of the connectors and the concavity were changed to connect two specific tiles. Five tiles were designed to align horizontally. In this system, a two-step self-assembly was adopted: first, individual origami tiles were prepared, and then the multiple tiles were assembled in the second stage by slower annealing from 50 °C. The DNA origami is stable enough to heat at 50 °C in secondary annealing. To allow the identification of the DNA tiles, hairpin markers were introduced onto the individual tiles. After self-assembly, judging by the order of the markers, the five tiles were aligned correctly. In addition, hairpin markers were used to display letters of the alphabet on the origami surface. The letters D, N, A, N, and O were each introduced onto one of the five tiles. After self- assembly using the first three tiles and last four tiles, the words "DNA" and "NANO" were displayed, respectively. Using five tiles with the letters K, Y, O, T, and O, the five-letter word "KYOTO" was formed (**Figure 4**).

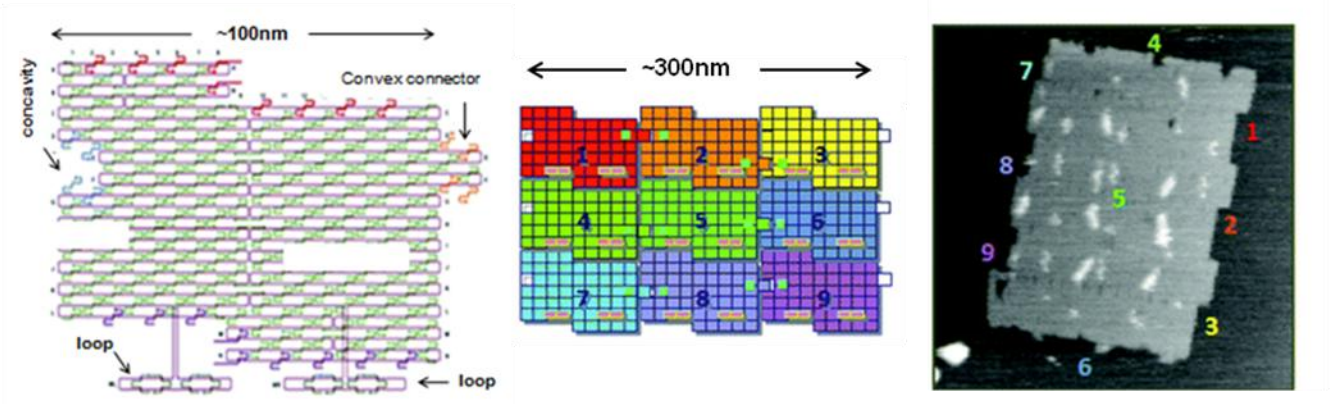


Figure 4: Programmed self-assembly of DNA origami. Structure of DNA origami having a concavity and a convex connector; the structure is called a "DNA jigsaw piece" for 2D assembly. A 3x3 assembly of nine tiles and the AFM image of the assembly.

The method described above was applied to prepare the 2D assembly system.[6] The shape and sequence selectivity were introduced to both lateral edges for extension in the vertical direction. Nine DNA tiles were designed and prepared. Three tiles were then programmed to be connected vertically or horizontally, and three sets of vertical or horizontal trimers were finally assembled into a 3 × 3 assembly with ~30% yield; this assembly was confirmed by hairpin markers introduced onto the individual origami tiles.

3.2.3. 3D DNA origami structures

Given the geometry of the periodic double-helical DNA structure, 3D structures can be designed by extending the 2D DNA origami system. Two strategies for preparing 3D DNA origami structures have been developed: one is the bundling of dsDNAs, where the relative positioning of adjacent dsDNAs is controlled by crossovers, and the other is the folding of 2D origami domains into 3D structures using interconnection strands. In the former method, developed by Shih and co-workers, relative positioning of adjacent dsDNAs is geometrically controlled by the crossovers. By arranging the positions of the crossovers, tubular and multilayered structures can be constructed (**Figure 5**).[7]

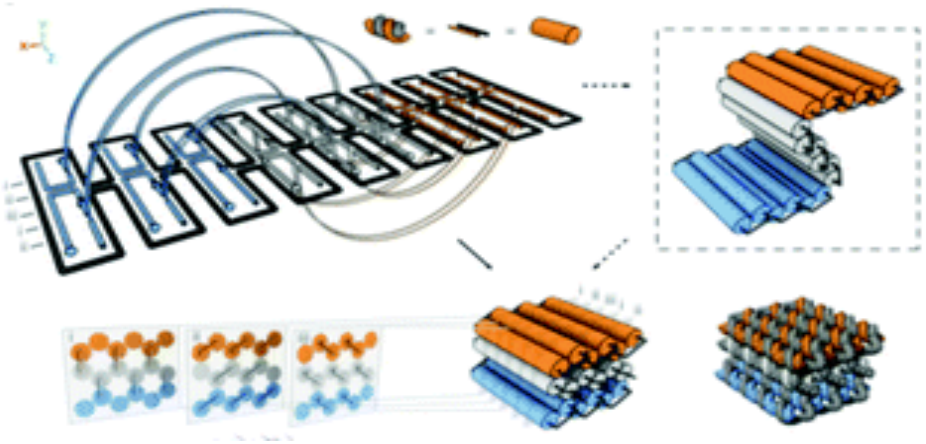


Figure 5: Design and construction of 3D DNA origami structures. Scheme for folding the 2D pleated structure into a 3D multilayered structure using staple strands connecting adjacent layers. Sectional views of the positions of the cross-overs in the multilayered structure sliced at seven-base-pair intervals.

By increasing or decreasing the number of base pairs between crossovers (in this case, two helical rotations for 21 base pairs), the relative positional relationship between adjacent dsDNAs is controlled.

Using the rotational angle of 240° for seven base pairs, three adjacent dsDNAs can be placed at a relative angle of ±120° with cross-overs every 7 or 14 base pairs. By alternating this relative positioning between adjacent dsDNAs, the duplexes form a pleated structure. When adjacent dsDNAs are placed to rotate in one direction, the contiguous duplexes finally form a six-helix bundled tubular structure. Thus, when some parts of the pleated structures are turned backward by the introduction of one-directional rotation of adjacent dsDNAs, the structures fold to become a stacked layer structure. In this case, to stabilize the 3D structures, adjacent layers of dsDNAs should be further connected by crossovers. Due to the complexity and high density of the introduced crossovers, an accurate folding into the target 3D structure requires a week-long folding time. When the pleated structures were integrated as multilayered structures, the repeating units of the six-helix bundled tubular structures formed a honeycomb lattice, viewed from the axial direction of the double helices. It was also possible to create more complex structures by perpendicularly joining these 3D structures. In addition, a wireframe icosahedron structure was assembled from three double-triangle monomers made of a six-helix bundled tubular structure with connections. caDNAno software, which is publicly available, has been developed to support the design of these 3D structures.[8] Furthermore, using the layered structures described above, new 3D structures were built by changing the helical twist from the average helical pitch of 10.5 base pairs per turn to 10 or 11 base pairs per turn.[9] When dsDNAs having different helical pitches were bundled together, torque and repulsion between base pairs caused overall structural changes including twisting or 30–180° bending (**Figure 6**).

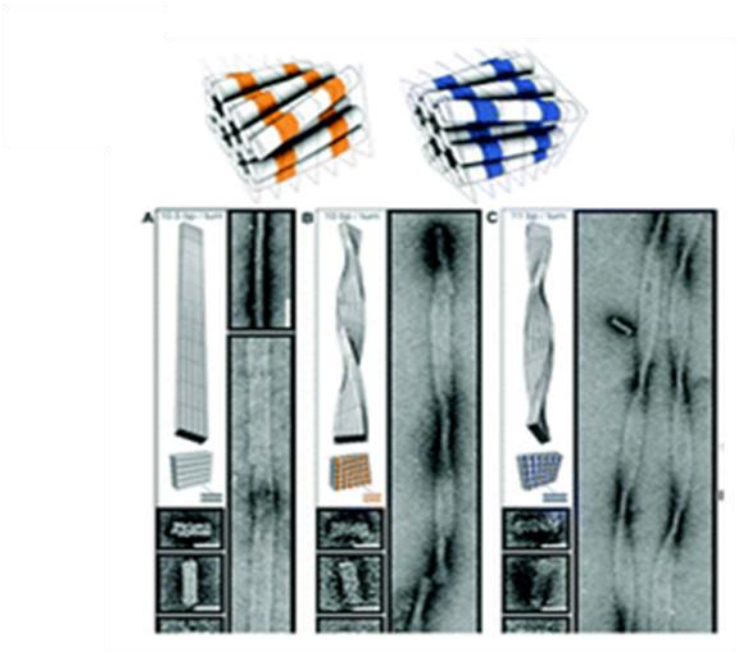


Figure 6: Global twisted structures of six-helix DNA bundles obtained by the selective deletion or insertion of nucleotides to change the helical turns from the normal 10.5 base pairs to 10 or 11 base pairs. TEM images of the polymerized ribbons containing 10.5-base-pair, 10-base-pair and 11-base-pair helical pitches.

Using these structures as building blocks, left-handed or right-handed helical ribbon structures were prepared. In addition, when angle-controlled duplex bundles were connected to each other, a six-tooth gear and a spherical wireframe capsule were created (**Figure 7**).

Figure 7: Global bending of six-helix DNA bundles by the deletion and insertion of nucleotides in the adjacent duplexes. Assembly of four components of a quarter circle with three teeth (50nm radius) and TEM images of the 12-tooth gear.

3.2.4. Construction of 3D structures from sequentially connected subunits

Using a different strategy, a DNA box structure was created by folding multiple 2D origami domains with interconnecting strands.[10] Six independent rectangles were sequentially linked, and were designed to be folded using interconnection strands in a programmed fashion (**Figure 8**).

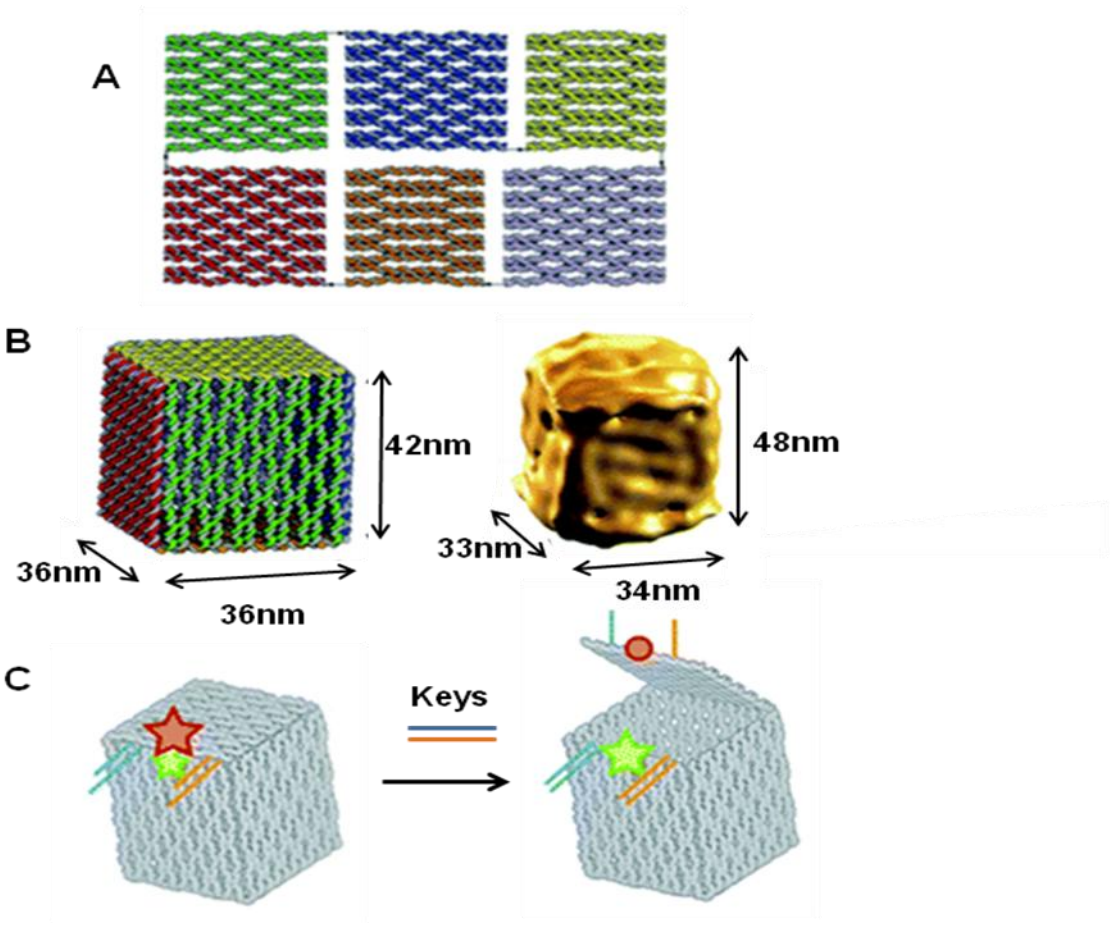


Figure 8: Design and construction of 3D structure from sequentially connected multiple rectangular plates. **A)** DNA box structure by folding of six DNA origami rectangles using interconnection strands introduced at the edges of rectangles. **B)** The DNA box model reconstructed from cryo-EM images. **C)** Controlled opening of the box lid using selective DNA strands (key). Lid opening event was monitored by FRET.

Analyses of the assembled structure by AFM, cryo-electron microscopy (cryo-EM), dynamic light scattering, and small-angle X-ray scattering indicated that the size was close to the original design. The lid of the box could be opened using a specific DNA strand to release the closing duplex by strand displacement, and the opening event was monitored by fluorescence resonance energy transfer (FRET) (**Figure 8 C**).

3.2.5. Applications of DNA origami structures

A wide variety of designs of 2D and 3D structures of around 100nm in size have been realized by constructing DNA origami. Compared with the use of small DNA assemblies, the DNA origami method reduces experimental labor and the need for strict stoichiometry and eliminates uncertainties. The DNA origami method was first reported in 2006. In the relatively short time multidimensional structures, functionalization, single-molecule imaging, and the construction of molecular of molecular machines have been realized. Functionalized DNA origami has already been combined with top-down nanotechnology, including semiconductor processing techniques. As the size of DNA origami is compatible with cellular uptake, it is expected that cell-targeting applications will benefit from the design of various shapes and effective functionalizations. Here I reported examples of state-of-art effective applications of this technique.

3.2.5.1. Single molecule chemical reactions

Selective bond cleavage and bond formation reactions were performed on a DNA origami surface. Target organic molecules with specific reactivity were introduced into specific positions on the DNA origami. Reductive cleavage of disulfide bonds and oxidative cleavage of olefin by singlet oxygen were carried out on the DNA origami surface, and the reactions proceeded quantitatively at the single-molecule level.[11] In addition, amide bond formation and click reactions on alkine were performed with 80–90% yield, sequencing reactions were also performed (**Figure 9**). These chemical reactions were monitored by the cleavage of biotin-attached chemical linkers and bond formation with biotin-tethered functional groups, which can be labeled with streptavidin for visualization by AFM.

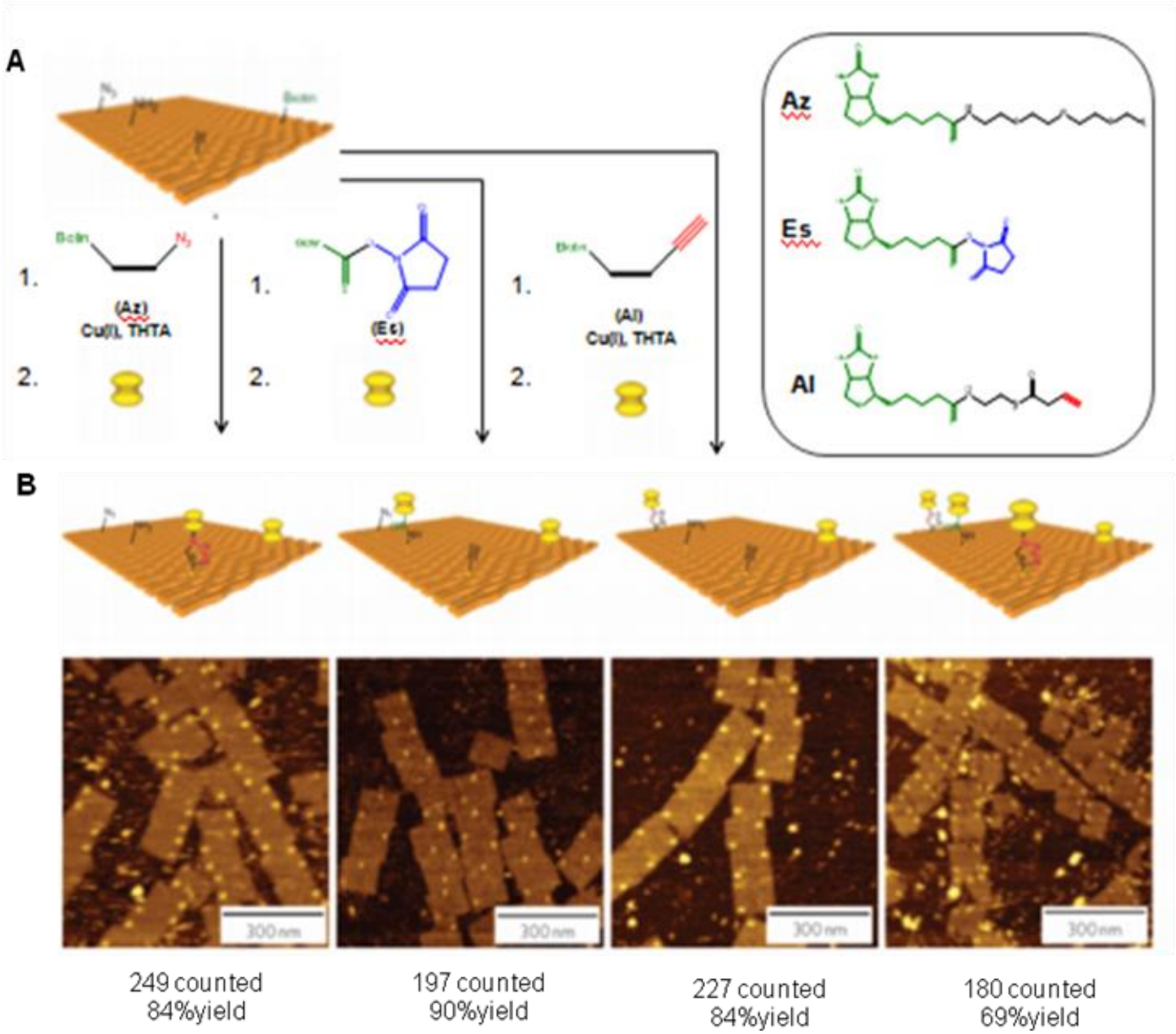


Figure 9: Single chemical reaction on DNA origami. **A)** Reactive groups (azido, amino, and alkyne groups) were incorporated into the DNA origami by conjugation with staple DNA strands. The coupling reactions were then performed using the biotin-attached functional groups. The completion of the reactions was visualized by the binding of streptavidin. **B)** AFM images of the three individual reactions and three successive reactions by the treatment of three biotin-attached functional groups. Yields are presented below.

3.2.5.2. Single biomolecule imaging

Direct observation of enzymes interacting with DNA is expected to be one of the most significant technologies for investigating the mechanical behavior of enzymes. Because DNA origami is used as a scaffold for AFM observation, the movement of biomolecules, including proteins and enzymes, and the DNA structural change itself can be visualized and analyzed if the substrate dsDNAs and the target DNA structures are attached onto the origami scaffold. DNA modification using enzymes often requires bending specific DNA strands to facilitate the reaction. The DNA methylation enzyme EcoRI methyltransferase (M.EcoRI) bends dsDNA by 55–59° during the methyl-transfer reaction.[12]

To control the methyl-transfer reaction of M.EcoRI and examine the structural effect on methylation, 2D DNA scaffold was designed and prepared, named a "DNA frame", which accommodates two different lengths of dsDNA fragments: a tense 64-mer dsDNA and a relaxed 74-mer dsDNA (**Figure 10**).[13]

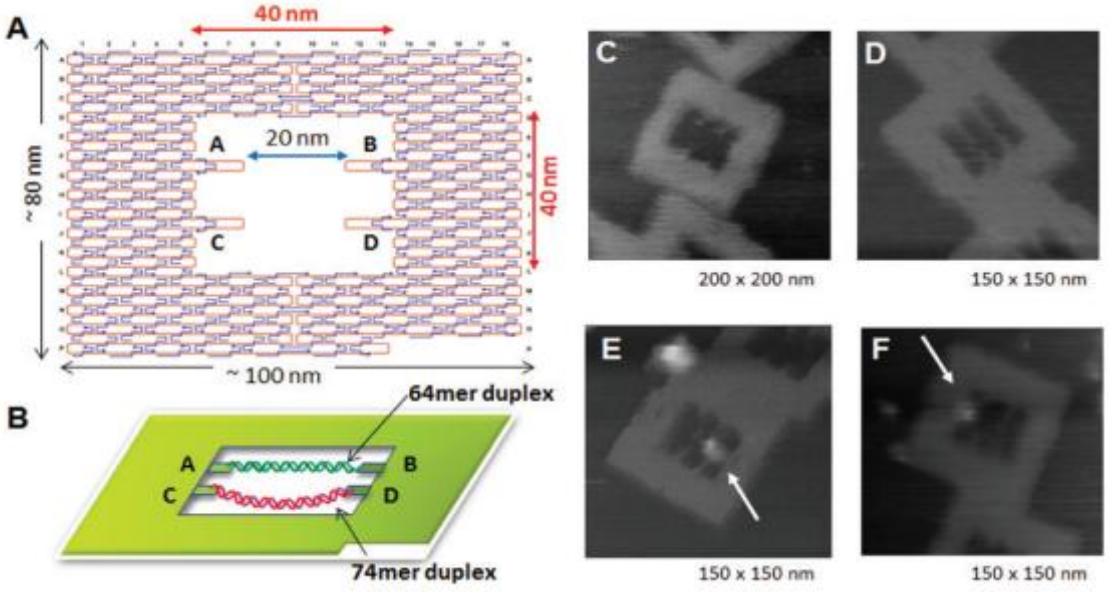


Figure 10: Control of the enzyme reactions in the DNA origami scaffold. **A), B)** DNA frame structure for incorporation of two different dsDNAs; tense 64-mer dsDNA and relaxed 74-mer dsDNA having the specific sequence for M, EcoRI at the centre. AFM images of DNA frame **C)** two-dsDNA attached DNA frame **D),** and M. EcoRI bound to the 64-mer dsDNA **E)** and the 74-mer **F)** dsDNA.

High-speed AFM revealed the different dynamic movements of the dsDNAs and complexes of M.EcoRI with 64-mer and 74-mer dsDNAs. After treatment of the dsDNA in the DNA frame with M.EcoRI and subsequent digestion by the restriction enzyme EcoRI, AFM analysis revealed that, compared with the 64-mer dsDNA, the 74-mer dsDNA was less effectively cleaved, indicating that the methylation preferentially occurred in the relaxed 74-mer dsDNA, rather than in the tense 64-mer dsDNA. Biochemical analysis of the methylation and specific digestion using real-time PCR supported the above results. These results indicate the importance of structural flexibility in the bending of dsDNA during the methyl-transfer reaction with M.EcoRI. Therefore, DNA methylation can be regulated using the tension-controlled dsDNAs incorporated in the DNA frame nanostructure. We next developed a novel method for the analysis of DNA repair by employing a DNA frame containing various dsDNAs and using high-speed AFM.[14] Sugiyama et al. employed DNA base excision repair enzymes, 8-oxoguanine glycosylase[15] and T4 pyrimidine dimer glycosylase,[16] for the analysis of a reaction on the defined DNA nanostructure. These enzymes have glycosylase/ AP lyase activity for removing damaged nucleobases and cleaving the DNA strand.[17] Various dsDNAs with a damaged base were placed onto a DNA nanochip as a dsDNA cassette and analyzed the repair reaction at the single-molecule level. We placed two different lengths of substrate dsDNAs, tense 64-mer and relaxed 74-mer dsDNAs, onto a DNA frame to examine the structural effect on the glycosylase/AP lyase activity, including cleavage of the DNA strand and trapping of reaction intermediates. The relaxed 74-mer dsDNA trapped the enzymes with $NaBH_4$ reduction and was cleaved more effectively compared with the 64-mer dsDNA. In addition, dynamic movement of the enzymes and the single DNA repair reaction were directly observed in the DNA frame using a high-speed AFM. The DNA frame system serves to elucidate the detailed properties of the repair enzymes by direct observation of the events involved in DNA repair. This method can be used for other DNA-modifying and repair enzymes that bend the double helix during the enzymatic reaction. The method can be extended to the direct observation of various enzymatic phenomena in the designed nanoscale space.

3.2.5.3. DNA nanostructures for cellular application

The various applications of DNA origami nanostructures described above have great biological potential and have already been extended to cellular studies. A few examples of DNA nanostructures being resistant to various types of endo- and exonucleases have been reported.[18] DNA origami constructs were able to maintain their integrity without degradation or damage in cell lysate of a series

of cell lines.[19] The high stability of DNA nanostructures in a biological system and their favorable compatibility with functional biomolecules, such as proteins and aptamers, demonstrate that the nanostructures are promising biomaterials for living-cell analysis and as platforms for safe drug delivery.

Besides intracellular signal triggering, Douglas and co-workers created a hexagonal barrel called a DNA nanorobot (**Figure 11 A**) for transporting molecular payloads to target cells and subsequent multiple interactions with the cellular system.[20] By employing an aptamer-based locking mechanism, the reconfiguration of the nanorobot structure for payload delivery was triggered using sensing of molecules on the cell surface as signal inputs. Aptamer-complement duplexes were introduced on the left and right sides of the front of the barrel. This hexagonal barrel nanodevice can be unlocked in response to the protein keys between an aptamer-complement duplex and aptamer–target complex (**Figure 11 B**). The hollow inside can be loaded with different types of payloads, such as proteins and nanoparticles (**Figure 11 C, D**), in a highly organized fashion for the delivery of various payloads, and the innate conformational regulation of the robot can be controlled selectively. To obtain a high yield of the nanorobot in a closed state, two "guide" staples were incorporated close to the locking site (**Figure 11 E**).

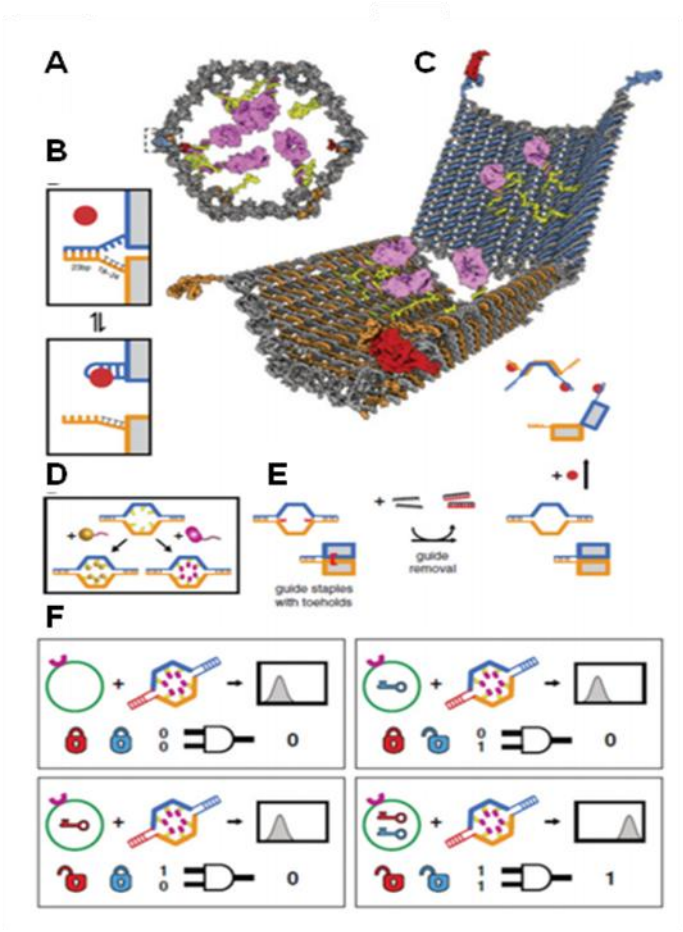


Figure 11: Design of aptamer-gated DNA nanorobot. **A)** Schematic drawings of the closed nanorobot in hexagonal barrel shape loaded with protein payloads inside in front orthographic view. The nanodevice was fastened by two DNA aptamer based locks on the left (dashed box) and right side. **B)** The nanodevice is unlocked when a DNA aptamer (orange) and the complementary strand (orange) dissociate in the presence of antigen key (red). **C)** Schematic view of the open state of nanorobot by protein key displacement of aptamer locks. **D)** Gold nanoparticles (gold) and Fab' antibody fragments are employed as payloads inside of the nanorobot after modifications. **E)** Guide staples with toe-holds are incorporated to obtain the high yield of nanorobot in the closed state. **F)** AND-gated nanorobot is activated (aptamer-encoded unlock) by molecular inputs expressed by target cells.

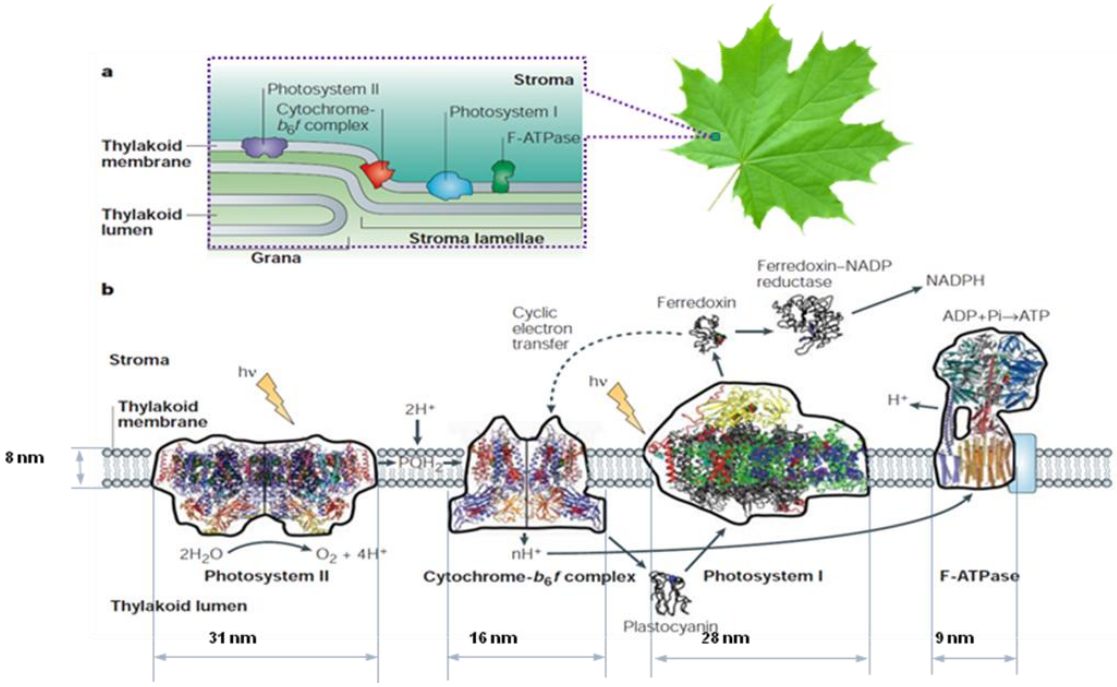
The aptamer locking mechanism was designed to play the role of a logic gate by binding to the target cells. When the same aptamer sequences are used at two locking sites, the robot can be activated in response to only one type of key, whereas two types of inputs (cell surface antigens) as keys are required at the same time if two different aptamer sequences are employed in order to activate the robot's function and expose the payload for further interactions with target cells (**Figure 11 F**). Furthermore, this robot can be used to interface with cells and stimulate their signaling in an inhibition

or activation manner by selective regulation of the nanorobot function. Therefore, DNA origami affords a new strategy for applications in cellular studies.[21]

3.2.6. References

1) Rothemund, P.W., *Nature*, **2006**, 440, 237-302.
2) Hogberg, B., Liedl, T., Shih, W. M., *J. Am. Chem. Soc.*, **2009**, 131, 9154-9155.
3) Pound, E., Ashton, J. R., Becerril, H. A., Woolley, A. T., *Nano Lett.*, **2009**, 9, 4302-4305.
4) Zhao, Z., Yan, H., Liu, Y., *Angew. Chem. Int. Ed.*, **2010**, 49, 1414-1417.
5) Endo, M., Sugita, T., Katsuda, Y., Hidaka, K., Sugiyama, H., *Chem-Eur. J.*, **2010**, 16, 5228.
6) Rajendran, A., Endo, M., Katsuda, Y., Hidaka, K., Sugiyama, H., *ACS Nano*, **2011**, 5, 665-671.
7) Douglas, S. M., Dietz, H., Liedl, T., Hogberg, B., Graf, F, Shih, W. M., *Nature*, **2009**, 459, 414-418.
8) Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M., Shih, W. M., *Nucleic Acids Res.*, **2009**, 37, 5001-5006.
9) Dietz, H., Douglas, S. M., Shih, W. M., *Science*, **2009**, 325, 725-730.
10) Andersen, E. S., Dong, M., Nielsen, M. N., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. N., Sander, B., Stark, H., Oliviera, C. L., Pedersen, J. S., Birkedal, V., Besenbacher, F., Gothelf, K. V., Kjems, J., *Nature*, **2009**, 459, 73-76.
11) Voigt, N. V., Torring, T., Rotaru, A., Jacobsen, M. F., Ravnsbaek, J. B., Subramani, R., Mamdouh, W., Kjems, J., Mokhir, A., Besenbacher, F., Gothelf, K. V., *Nat. Nano-technol.*, **2010**, 5, 200-203.
12) Youngblood, B, Reich, N. O., *J. Biol. Chem.*, **2006**, 281, 26821-26831.
13) Endo, M., Katsuda, Y., Hidaka, K., Sugiyama, H., *J. Am. Chem. Soc.*, **2010**, 132, 1592-1597.
14) Endo, M., Yang, Y., Emura, T., Hidaka, K., Sugiyama, H., *Chem. Commun.*, **2011**, 47, 10743-10745.
15) Bruner, S. D., Norman, D. P., Verdine, G. L., *Nature* **2000**, 403, 859-866.
16) Morikawa, K., Matsumoto, O., Tsujimoto, M., Katayanagi, K., Ariyoshi, M., Doi, T., Ikehara, M., Inaoka, T., Ohtsuka, E., *Science* **1992**, 256, 523-526.
17) Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S., Verdine, G. L., *Curr. Biol.* **1996**, 6, 968-980.
18) Castro, C. E., Kilchherr, F., Kim, D. N., Shiao, E. L., Wauer, T., Wortmann, P., Bathe, M., Dietz, H., *Nat.Methods* **2011**, 8, 221-229.
19) Mei, Q., Wei, X., Su, F., Liu, Y., Youngbull, C., Johnson, R., Lindsay, S., Yan, H., Meldrum, D., *Nano Lett.* **2011**, 11, 1477-1482.
20) Douglas, S. M., Bachelet, I., Church, G. M., *Science*, **2012**, 335, 831-834.
21) Endo, M., Yang, Y., Sugiyama, H., *Biomater. Sci.*, **2013**, 1, 347-360.

3.3.Networks of functional modules on DNA origami

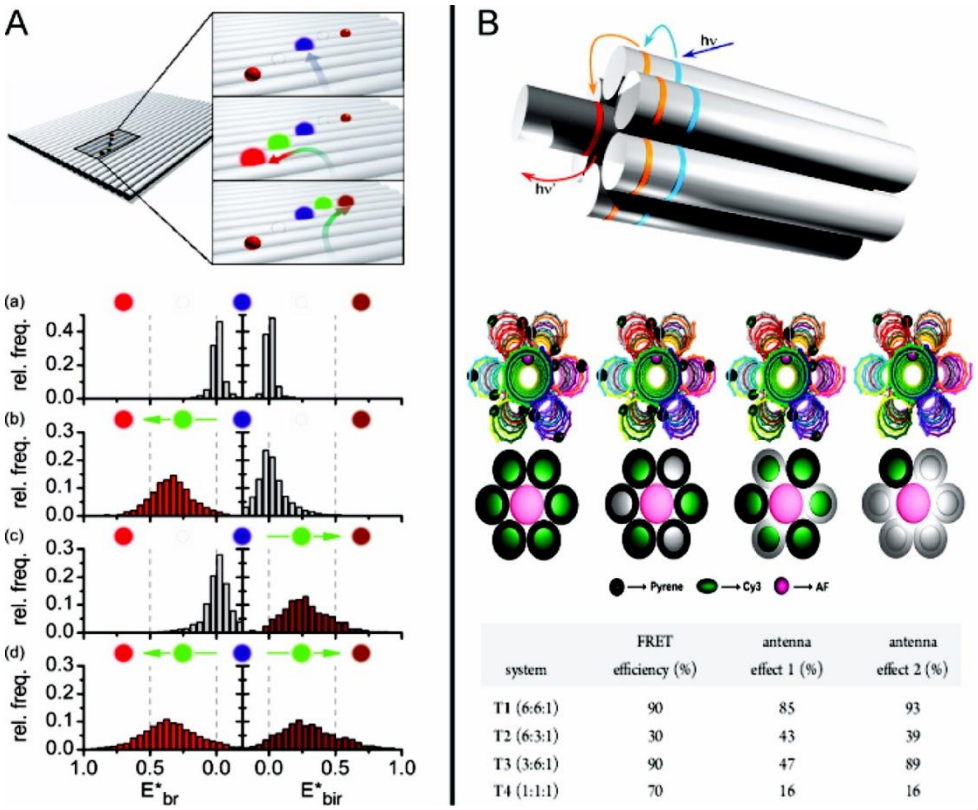


Biological systems use complex macromolecular nanostructure networks to mediate a range of cellular functions, such as biomolecular synthesis, signal transduction, and gene expression and regulation, all with high efficiency and specificity. Many of these macromolecular systems have evolved through the spontaneous self-assembly of components into highly organized spatial structures, where the position and orientation of molecules are precisely controlled to facilitate functionality. For example, the multienzyme cascades[1] found in biochemical synthesis pathways and the light harvesting system in photosynthetic reaction centers[2] both rely on very specific arrangements of components. Over the past few decades, molecular self-assembly processes have been exploited to construct various nanostructures including vescicles, nanofibers, and nanotubes from self-assembling lipids, peptides, nucleic acids, and polysaccharides.[3] However, it remains a challenge to accurately arrange multiple heterogeneous components into geometric patterns with nanometer precision, as in natural systems. Additional challenges include the development of novel assembly algorithms to increase structural complexity and improve the fidelity and yield of the assembly process. DNA is among the most promising biomolecules for the construction of complex biomolecular networks.[4] One of the most attractive properties of DNA origami structures is the addressability of the surface, the result of the unique sequence at each oligonucleotide staple position. Thus, various patterns can be displayed by selectively modifying staple strands at desired locations with single-stranded probe extensions. The

DNA origami method has several advantages over "tile-based" assembly approaches: 1) scaffolded DNA can be folded into nearly any symmetric or asymmetric structure; 2) well-formed nanostructures are generated with high yield using unpurified oligonucleotides, because the scaffold imposes the correct stoichiometry between strands; 3) spatially addressable assembly is achieved with a resolution of ~6nm. Moreover, DNA origami nanostructures are reliable directors in the organization of heterogeneous nanoscale entities such as peptides,[5] proteins[6] and nanoparticles.[7] Supermolecular networks of molecules that are scaffolded by DNA nanostructures exhibit well-controlled intercomponent distance and relative numbers. This characteristic presents exciting opportunities for fundamental studies of distance-dependent molecular interactions and for practical applications including biosensing, molecular biophysics, biocatalysis, drug delivery, and responsive nanodevices.

3.3.1. Light-harvesting networks

In natural photosynthesis, light is harvested by antenna systems that consist of networks of spatially organized chromophores to facilitate unidirectional energy transfer to a red-ox center.[8] In artificial systems, DNA nanostructures can be used to arrange multiple pairs of fluorescence donors and acceptors into precise geometric patterns to achieve efficient energy transfer. In **Figure 12 A**, a DNA origami tile was used to organize several distinct fluorophores into closely packed linear arrangements to achieve multicolor energy transfer, observable at the single-molecule level.[9] Energy transfer was directed along a path from a blue to red dye or from a blue to IR dye by placing a "jumper dye" between the primary donor and the final acceptor. As shown in **Figure 12 B**, an artificial light-harvesting antenna was constructed by assembling multiple donor–acceptor pairs on a seven-helix DNA bundle.[10] Steady-state and time-resolved fluorescence spectroscopy was used to measure the efficiency of energy transfer for networks with various ratios of donor to acceptor dyes.



system	FRET efficiency (%)	antenna effect 1 (%)	antenna effect 2 (%)
T1 (6:6:1)	90	85	93
T2 (6:3:1)	30	43	39
T3 (3:6:1)	90	47	89
T4 (1:1:1)	70	16	16

Figure 12: Energy-transfer within DNA nanostructures. (**A**) Four-color FRET and (**B**) artificial light-harvesting network.

3.3.2. Organization of multienzyme reaction pathways

The metabolism of living systems involves complex synthetic pathways with numerous multistep reactions that possess extraordinary yields and specificities. Many enzyme systems carrying out these reaction pathways are highly organized complexes with precisely controlled enzyme positions and orientations, facilitating efficient diffusion of substrates between the enzymes.[1] Artificial synthesis of these multienzyme systems is generally achieved by genetic fusion,[11] chemical cross-linking and coimmobilization;[12] however, precise control over spatial organization of components is lacking for these methods. With DNA nanostructures as assembly scaffolds, it has become feasible to organize multiple enzymes with controlled spacing in linear as well as 2D or 3D geometric patterns, which enables the study of cascade activity.[13] One of the first demonstration was the assembly of a bioenzymatic NAD(P)H:FMN oxidoriductase and luciferase cascade on a double-stranded DNA

scaffold with an observed ~3fold increase in activity compared with the corresponding unassembled enzyme pair (**Figure 13 A**).[14]

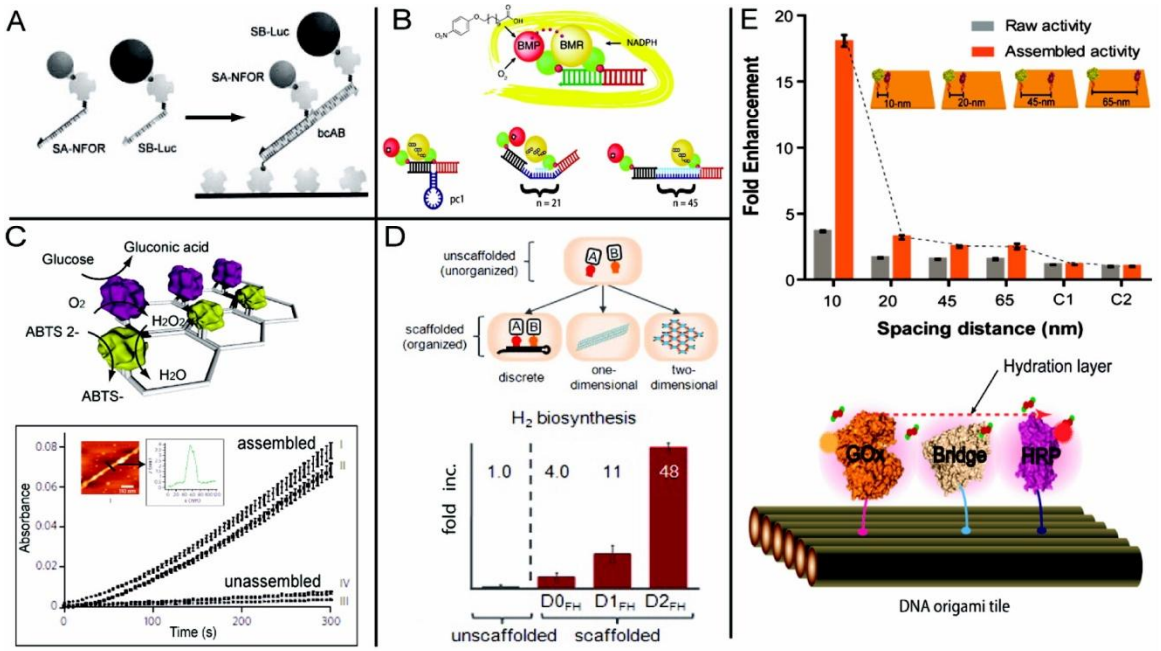


Figure 13: DNA/RNA nanostructures for engineering multienzyme systems. A linear double-stranded DNA scaffold for (**A**) assembling an enzyme cascade, NAD(P)H:FMN (NFOR) oxidoreductase and luciferase (Luc), and (**B**) evaluating the distance-dependent activity of cytochrome P450 BM3 by varying the spacing between the BMR reductase domain and the BMP porphyrin domain. (**C**) Two-dimensional DNA strip for organizing GOx/HRP cascades. (**D**) *In vivo* assembly of RNA nanostructures to organize the [FeFe]-hydrogenase and ferredoxin enzyme pathway for improved hydrogen production. (**E**) Organization of a GOx/HRP cascade on DNA origami tiles with controlled spatial positions (top), and a protein bridge for facilitating surface-limited intermediate diffusion between enzymes (bottom).

This strategy was later applied to probing the distance-dependent activity of multi-domain complexes of cytochrome P450 BM3 by varying the length of spacing scaffolds between the BMR reductase domain and the BMP porphyrin domain (**Figure 13 B**).[15] Two-dimensional DNA nanostructures provide an even greater opportunity to organize multienzyme systems into more complicated geometric patterns. There was a report of the self-assembly of a glucose oxidase (GOx) and horseradish peroxidase (HRP) enzyme cascade on 2D hexagonal DNA strips, with the distance between the two enzymes controlled by the underlying nanostructure (**Figure 13 C**).[16] A greater than 10-fold activity enhancement was observed compared with the corresponding unstructured enzymes. In addition to *in vitro* assembly, multienzyme pathways can also be organized by introducing nucleic acid

nanostructures as assembly scaffolds *in vivo*, an approach facilitated by recent advances in RNA nanotechnology.[17] This idea was demonstrated by the assembly of an intracellular reaction pathway ([FeFe]-hydrogenase and ferrexodin) for enhancing bacterial hydrogen production.[18] In **Figure 13 D**, discrete, 1D and 2D RNA scaffolds were assembled *in vivo* through the incorporation of aptamers for capturing the target enzyme cascade. Remarkably, a 48-fold enhancement of hydrogen production was observed for the RNA-templated [FeFe]-hydrogenase and ferroxidin network. This work suggests that a metabolic engineering approach can be used to introduce structural nucleic acid nanostructures inside cells for the organization of multienzyme reaction pathways. Recently, a GOx/HRP cascade was organized on DNA origami tiles with precisely controlled spatial positions, which was applied to investigating the distance-dependent interenzyme substrate diffusion (**Figure 13 E**).[19] The study revealed that substrate transfer between enzymes might occur at the connected hydration shells for closely paced enzymes and demonstrated this idea by constructing a protein bridge to facilitate the intermediate transfer across protein surfaces.

3.3.3. Orthogonal adaptors for site-specific protein positioning on DNA origami

Methods to attach proteins at specific locations on DNA scaffolds have been reported and include those based on antibody–antigen interactions,[20-23] aptameric binding,[24,25] Ni-NTA–hexahistidine interactions,[26-28] and biotin–avidin interactions.[29-35] Orthogonal targeting of specific locations can also be achieved by hybridization with DNA-tethered proteins,[36-39] sequence-specific DNA binding of pyrrole-imidazole polyamides,[40-42] and self-ligating protein tags.[43,44] Many of these methods require modification of the protein. Therefore, a method that is fully based on protein components would accelerate the specific assembly of proteins on the DNA nanoarchitecture. Morii et al. reported that different locations within DNA-origami structures could be site-specifically and orthogonally targeted by using sequence-specific DNA-binding proteins as an adaptor, and demonstrated that adaptor-fused functional proteins are assembled at specific locations within DNA-origami structures. Zinc-finger proteins (ZFPs) are one of the best characterized classes of DNA-binding proteins;[45,46] designed, artificial ZFPs bind to a wide variety of DNA sequences.[47-52] Each zinc-finger domain is capable of recognizing a tract of four base pairs in the major groove of a DNA duplex. A three-fingered protein recognizes a tract of ten base pairs with nanomolar affinity. Two well-characterized ZFPs, zif268[53] and AZP4,[54] each with an affinity for a unique sequence of ten base pairs in the low nanomolar range, were chosen as the orthogonal adaptors for specific locations in the DNA-origami structures. Each ZFP was

engineered to possess an N-terminal cysteine residue (C-zif268 and C-AZP4) as a selective chemical modification site (**Figure 14**).

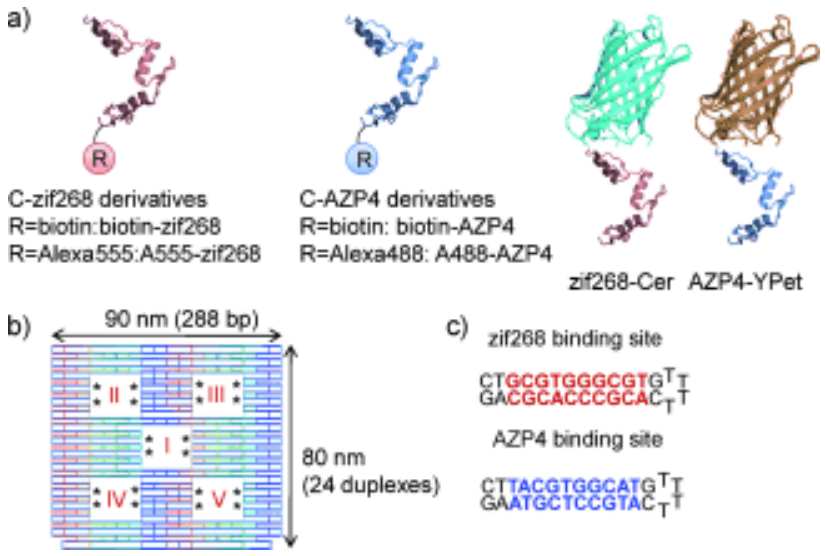


Figure 14 a) Illustrations showing the structure of the zinc-finger adaptors and adaptor-fused proteins. **b**) A structural image of the DNA-origami structures. The addressable cavities are indicated by the numbers I, II, III, IV, and V. The positions of the addresses (binding sites) of zinc-finger adaptors are indicated by asterisks. Bp=base pairs. c) Nucleotide sequences for the specific binding site of the zinc-finger adaptors.

ZFPs turned out to be convenient and site-selective adaptors for targeting specific locations within DNA-origami structures. The diversity of target DNA sequences and the semi-programmable design of ZFPs offers orthogonal adaptors, thereby enabling the placement of multiple engineered proteins at different locations onto DNA-origami structures. Nature uses multiple proteins and/or enzymes in close proximity to efficiently carry out chemical reactions and signal transductions. Such assemblies of multiple proteins may be realized in vitro by using DNA-origami structures that have defined binding sites and various kinds of ZFP adaptor-fused proteins.[55]

3.3.4. References

1) Savage, D. F., Afonso, B., Chen, A. H., Silver, P. A., *Science*, **2010**, 327, 1258-1261.
2) Cogdell, R. J., Gall, A., Köhler, J., *Q. Rev. Biophys.*, **2006**, 39, 227-324.
3) Stupp, S. I., *Nano Lett.*, **2010**, 10, 4783-4786.
4) Lin, C., Liu, Y., Yan, H., *Biochemistry*, **2009**, 48, 1663-1674.
5) Williams, B. A. R., Lund, K., Liu, Y., Yan, H., Chaput, J. C., *Angew. Chem. Int. Ed.*, **2007**, 46, 3051-3054.
6) Kuzyk, A., Laitinen, K. T., Törmä, P., *Nanotechnology*, **2009**, 20, 235305.
7) Tan, S. J., Campolongo, M. J., Luo, D., Cheng, W., *Nat. Nanotechnol.*, **2011**, 6, 268-276.
8) Gust, D., Moore, T. A., Moore, A. L., *Acc. Chem. Res.*, **2000**, 34, 40-48.
9) Stein, I. H., Steinhhauer, C., Tinnerfeld, P., *J. Am. Chem. Soc.*, **2011**, 133, 4193-4195.
10) Dutta, P. K., Varghese, R., Nangreave, J., Lin, S., Yan, H., Liu, Y., *J. Am. Chem. Soc.*, **2011**, 133, 4193-4195.
11) Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., Prather, K. L. J., Keasling, J. D., *Nat. Biotechnol.*, **2009**, 27, 753-759.
12) Sheldon, R. A., *Adv. Synth. Catal.*, **2007**, 349, 1289-1307.
13) Teller, C., Willner, I., *Trends Biotechnol.*, **2010**, 28, 619-628.
14) Niemeyer, C. M., Koehler, J., Wuerdemann, C., *ChemBioChem*, **2002**, 3, 242-245.
15) Erkelenz, M., Kuo, C. H., Niemeyer, C. M., *J. Am. Chem. Soc.*, **2011**, 133, 16111-16118.
16) Wilner, O. L., Weizmann, Y., Gill, R,, Lioubashevski, O., Freeman, R., Willner, I., *Nat. Nanotechnol.*, **2009**, 4, 249-254.
17) Guo, P., *Nat. Nanotechnol.*,**2010**, 5, 833-842.
18) Delebecque, C. J., Linder, A. B., Silver, P. A., Aldaye, F. A., *Science*, **2011**, 333, 470-474.
19) Fu, J., Liu, Y., Woodbury, N. W., Yan, H., *J. Am. Chem. Soc.*, **2012**, 134, 5516-5519.
20) He, Y., Tian, Y., Ribbe, A. E., Mao, C., *J. Am. Chem. Soc.*, **2006**, 128, 12664-12665.
21) Williams, B. A. R., Lund, K., Liu, Y., Yan, H., Chaput, J. C., *Angew. Chem.*, **2007**, 119, 3111-3114.
22) Williams, B. A. R., Lund, K., Liu, Y., Yan, H., Chaput, J. C., *Angew. Chem. Int. Ed.*, **2007**, 46, 3051-3054.
23) Kuzuya, A., Sakai, Y., Yamazaki, Xu, Y., Komiyama, M., *Nat. Commun.*, **2011**, 2, 449.
24) Rinker, S., Ke, Y., Liu, Y., Chhabra, R., Yan, H., *Nat. Nanotechnol.*, **2008**, 3, 418-422.

25) Chhabra, R., Sharma, J., Ke, Y., Liu, Y., Rinker, S., Lindsay, S., Yan, H., *J. Am. Chem. Soc.*, **2007**, 129, 10304-10305.

26) Shen, W., Zhong, H., Neff, D., Norton, M. L., *J. Am. Chem. Soc.*, **2009**, 131, 6660-6661.

27) Goodman, R. P., Erben, C. M., Malo, J., Ho, W. M., McKee, M. L., Kapanidis, A. N., Tuberfield, A. J., *ChemBioChem*, **2009**, 10, 1551-1557.

28) Selmi, D. N., Adamson, R. J., Attrill, H., Goddard, A. D., Gilbert, R. J. C., Watts, A., Tuberfield, A. J., *Nano Lett.*, **2011**, 11, 657-660.

29) Yan, H., Park, S. H., Finkelstein, G., Reif, J. H., LaBean, T. H., *Science*, **2003**, 301, 1882-1884.

30) Li, H., Park, S. H., Reif, J. H., LaBean, T. H., Yan, H., *J. Am. Chem. Soc.*, **2004**, 126, 418-419.

31) Lund, K., Liu, Y., Lindsay, S., Yan, H., *J. Am. Chem. Soc.*, **2005**, 127, 17606-17607.

32) Kuzuya, A., Kimura, M., Numajiri, K., Koshi, N., Ohnishi, Okada, F., Komiyama, M., *ChemBioChem*, **2009**, 10, 1811-1815.

33) Voigt, N. V., Torring, T., Rotaru, A., Jacobsen, M. F., Ravnsbaek, J. B., Subramani, R., Mamdouh, W., Kjems, J., Mokhir, F., Besenbacher, F., Gothelf, K. V., *Nat. Nanotechnol.*, **2010**, 5, 200-203.

34) Numajiri, K., Yamazaki, T., Kimura, M., Kuzuya, A., Komiyama, M., *J. Am. Chem. Soc.*, **2010**, 132, 9937-9939.

35) Aldaye, F. A., Senapedis, W. T., Silver, P. A., Way, J. C., *J. Am. Chem. Soc.*, **2010**, 132, 14727-14729.

36) Niemeyer, C. M., Bürger, W., Peplies, J., *Angew. Chem.*, **1998**, 110, 2391-2395.

37) Niemeyer, C. M., Bürger, W., Peplies, J., *Angew. Chem. Int. Ed.*, **1998**, 37, 2265-2268.

38) Wilner, O. I., Weizmann, Y., Gill, R., Lioubashevski, O., Freeman, R., Willner, I., *Nat. Nanotechnol.*, **2009**, 4, 249-254.

39) Jahn, K., Torring, T., Voigt, N. V., Sorensen, R. S., Kodal, A. L. B., Andersen, E. S., Gothelf, K. V., Kjems, J., *Bioconjugate Chem.*, **2011**, 22, 819-823.

40) Cohen, J. D., Sadowski, J. P., Dervan, P. B., *Angew. Chem*, **2007**, 119, 8102-8105.

41) Cohen, J. D., Sadowski, J. P., Dervan, P. B., *Angew. Chem. Int. Ed.*, **2007**, 46, 7956-7959.

42) Cohen, J. D., Sadowski, J. P., Dervan, P. B., *J. Am. Chem. Soc.*, **2008**, 130, 402-403.

43) Saccà, B., Meyer, R., Erkelenz, M., Kiko, K., Arndt, A., Schroeder, H., Rabe, K. S., Niemeyer, C. M., *Angew. Chem.*, **2010**, 122, 9568-9573.

44) Saccà, B., Meyer, R., Erkelenz, M., Kiko, K., Arndt, A., Schroeder, H., Rabe, K. S., Niemeyer, C. M., *Angew. Chem. Int. Ed.*, **2010**, 49, 9378-9383.

45) Klug, A., Schwabe, J. W. R., *FASEB J.*, **1995**, 9, 597-604.

46) Brayer, K. J., Segal, D., *Cell Biochem. Biophys.*, **2008**, 50, 111-131.
47) Pabo, C. O., Peisach, E., Grant, R. A., *Annu. Rev. Biochem.*, **2001**, 70, 313-340.
48) Blancafort, P., Segal, D. J., Barbas III, C. F., *Mol. Pharmacol.*, **2004**, 66, 1361-1371.
49) Papworth, M., Kolasinska, P., Minczuk, M., *Gene*, **2006**, 366, 27-38
50) Negi, S., Imanishi, M., Matsumoto, M., Sugiura, Y., *Chem. Eur. J.*, **2008**, 14, 3236-3249.
51) Klug, A., *Annu. Rev. Biochem.*, **2010**, 79, 213-231.
52) Pavletich, N. P., Pabo, C. O., *Science*, **1991**, 252, 809-817.
53) Sera, T., Uranga, C., *Biochemistry*, **2002**, 41, 7074-7081.
54) Smith, J. J., Conrad, D. W., Cuneo, M. J., Hellington, H. W., *Protein Sci.*, **2005**, 14, 64-73.
55) Nakata, E., Liew, F. F., Uwatoko, C., Kiyonaka, S., Mori, Y., Katsuda, Y., Endo, M., Sugiyama, H., Morii, T., *Angew. Chem.*, **2012**, 51, 2421-2424.

3.4. Objectives and challenges

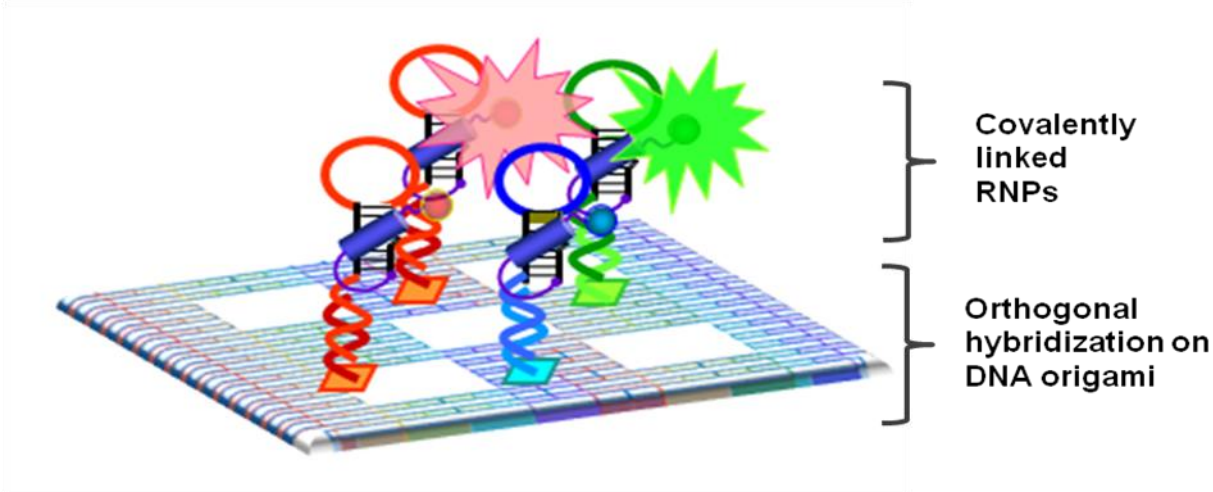


Figure 15: Schematic illustration of covalently linked RNPs arranged on DNA origami

Receptor clusters are important sources of information related to biological processes and complex metabolic functions of the cell. The presence of biologically relevant functional modules in these networks has been theorized but up-to-date the availability of *in vitro* reproducible models to study this phenomenon is limited. In this project we approached the issue of constructing reproducible clusters of ATP-binding RNP receptors with varied distances and combinations outside the cell, using DNA origami technology. As previously mentioned, DNA origami technology[1] presents the unique feature of fully addressable nanoarchitectures that are easily designed and constructed. By taking advantage of their addressable nature, DNA nanostructures have been used as one of the ideal scaffolds for the site-directed assembly of multiple proteins at specific locations on a DNA origami scaffold, termed as molecular switchboard.[2,3] The molecular switchboard strategy is also applicable to construct nano-assembled clusters of fluorescent sensors on DNA origami. ATP selective biosensors based on ribonuceopeptide (RNP) that consist of a noncovalent complex of ATP-binding RNA and a fluorophore-labeled peptide[4] were described in chapter 2. Covalently linked fluorescent RNP sensors enabled ligand detection at a low sensor concentration.[4] We have constructed molecular switchboards with programmable placement of multiple ATP-binding biosensors at defined locations on DNA origami.

This work was presented in the form of a poster in the 40th International Symposium on Nucleic Acids Chemistry (ISNAC 2013) at Kanagawa University, Yokohama and it is material of an article on the way of being published.

3.5. Strategy for assembling covalently linked RNP receptors on DNA origami

3.5.1. Tag design

The versatility of biomolecular recognition supports the high-affinity, high-specificity recognition of an enormous range of molecular targets. Despite these positive attributes, biological recognition exhibits a potentially significant limitation: the physics of single-site binding produces a hyperbolic dose-response curve for which the useful dynamic range spans a fixed change in target concentration. Specifically, the transition from 10% to 90% site occupancy requires a fixed 81-fold span of target concentration (**Figure 16**).

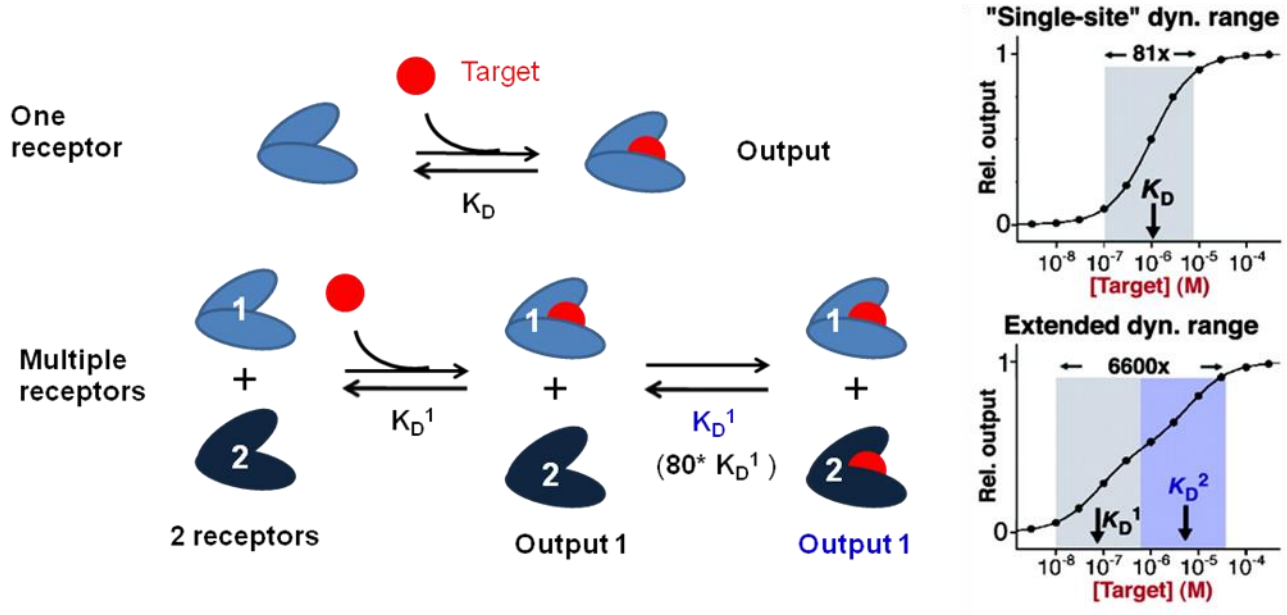


Figure 16: Using nature's tricks to extend or narrow the fixed dynamic ranges of single-site receptors. Top: The dynamic range of sensors with a single-site receptor spans an 81-fold range of target concentration over which the sensor response transits from 10% to 90% of its signal output. Bottom: This useful dynamic range can be extended by combining multiple receptors differing in their affinity for the same target.

This fixed dynamic range complicates the use of biosensors in many applications, for instance they are poorly suited for applications requiring the very precise measurement of target concentration or the therapeutic indices of many drug, that are often less than an order of magnitude.[6,7] Faced with the above limitations, evolution has invented a number of simple mechanisms by which the normally fixed dynamic range of single-site binding can be extended, narrowed or otherwise "edited" to better ensure the survival of an organism. For example, in order to create extended or even more complex three-state

dose-response sensing systems, evolution often employs pairs of closely related receptors.[8,9] This project has the aim to expand the potential applicability of a system of affine receptors arranging a detection field on DNA origami surface.

Four ATP specific RNA receptors, previously isolated by *in vitro* selection by Morii et al., presenting a wide range spectrum of affinity towards the target, were selected. Respectively An16, A26, An15 and A17 (K_D values reported in the box) which nucleobase sequences are reported in **Figure 17**.

Kd	TAG	RRE	V	C	NC	C	V	RRE
An15 tag02 (0.8μM)	GGAGGC UUCAGC UUC GAACU	GGUCUGGGCGCA	AU	GUAGUGG	UU	UGUGUGUG	AU	UGACGGUACAGGCC
An16 tag04 (2.3μM)	GGU AACUAC UUCC CUGGGAA	GGUCUGGGCGCA	C	GUAGUGG	UG	UGUGUGUG	G	UGACGGUACAGGCC
A17 tag03 (13.3μM)	GGA UUCGGCUCUUCAGCGAA	GGUCUGGGCGCA	UUAGAU	GUAGUGG	GUAGUG	UGUG	UUUUCUG	UGACGGUACAGGCC
A26 tag01 (142μM)	GGA UAA GGCUUC GGCCUAUG	GGUCUGGGCGCA	UUCCG	GUAGUGG	UUGUGUG	UGUG	CGGUUUU	UGACGGUACAGGCC

Figure 17: ATP specific RNP receptors nucleic base sequences. K_D values are reported in the box.

In order to locate the RNP receptors with precise positioning on DNA origami, the RNA modules were modified by a DNA-complementary tag at the 5' terminal. With the aim of avoiding interferences with the target binding region and to guarantee the proper folding of the RNA structure, the orthogonal 4 types of tag sequences were designed as follows.

A26

Aptamer sequence:

GGUCUGGGCGCAUUCCG**GUAGUGG**UUGUG**UGUGUG**CGGUUUU**UGACGGUACAGGCC**

Complementary sequence:

GGC CUG UAC CGU CAA AAA CCG CAC ACA CAC AAC CAC UAC CGG AAU GCG CCC AGA CC

Complementary Reverse sequences:

CCA GAC CCG CGU AAG GCC AUC ACC AAC ACA CAC ACG CCA AAA ACU GCC AUG UCC GG **01**

CC AGA CCC GCG UAA GGC CAU CAC CAA CAC ACA C AC GCC A AA A AC UGC CAU GUC CGG **02**

C CAG ACC CGC GUA AGG CCA UCA CCA ACA CAC ACA CGC CAA AA A CUG CCA UGU CCG G **03**

Total				
	A	U	C	G
A	AAA (3) AAC (2)	AUC AUG	ACC (2) ACA (4) ACG ACU	AGG AGA
U	UAA	UUG UUA	UCA UCG UCC	UGU UGC
C	CAG CAU (2) CAC CAA (2)	CUG CUA	CCA (4) CCG (2) CCC	CGC (2) CGG CGU
G	GAC (5) GAU	GUA GUC	GCG GCC (3)	GGC

Start sequence **GGA or GGU**

Aversive Sequence **CAN, ACN, CGN, GAC, CCR, GCS, AAM**

Potential Aversive Sequence

UAN, UGN, GCN, GUN, AUN, GAU, UCR, CUR, UUR, GGN, AAN, AGN, GAN

W: A,T (U)R:A,G M: A, C K: T(U), G Y: T(U), C S: G, C

H: A, C, T (U) B: G, C, T(U) V: A, G, C D: A, G, T(U)

N: A, G, C, T (U)

Blue: aversive sequence

Frequency-indicated Aversive Sequence	⇒	Frequency-indicated Potential **Aversive Sequence**
5 CAC		YRY
4 CCA, ACA		YYR, RYR
3 AAA, GCC		RRR, RYY
2 ACC, CAU, CAA, CCG, CGC		YRU, YRR, YYG, YGY

Sequence Component **GGA UUN UAB GGD CUH**

Add Frequency Aversion **GGA UUN UAR GGY CUH (green: deliberate permission)**

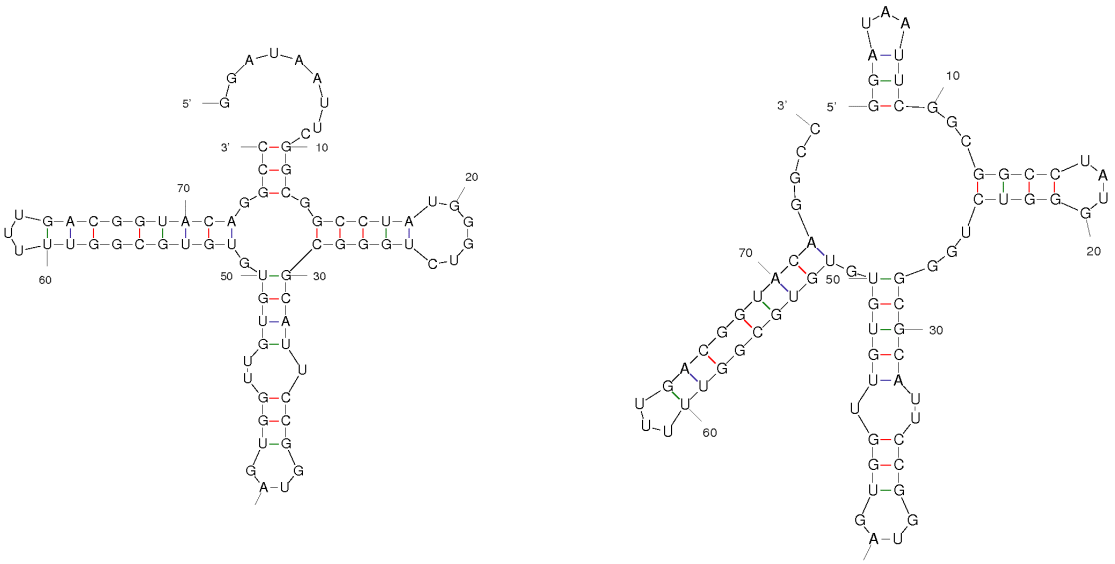
Example1 **GGA UUC GGC UAA CUAGGCUUG4A 6U :4C 7G (10:11):total 21**

GGA UAAUUC GGCGGCCUAUG4A 5U :4C 7G (9:11):total 20

A nucleobase sequence analysis was carried out evaluating the composition and the frequency with which the base combinations (considered as codons pairings) appear in the aptamer sequence. The less frequent combinations were selected as potential candidates for the tag framework design. As second selection criterion, the melting temperature, the process by which the interactions between the strands of the double helix are broken, separating the two nucleic acid strands, was taken in consideration. T and A rich sequences are more easily melted than C and G rich regions stabilized by the formation of three hydrogen bonds. These mechanical features are reflected by the use of sequences such

as TATAA at the start of many genes to assist RNA polymerase in melting the DNA for transcription. It was set to keep the melting temperature above 55ºC in order to obtain a RNA sequence sufficiently stable (See material and methods). Successively, in order to verifying that the tag moiety has no interference with the correct folding of the RNA module, a software simulation was performed (**Figure 18**).

Tag: GGAUAAUUCGGCGGCCUAUG
GGUCUGGGCGCAUUCCGGUAGUGGUUGUGUGUGUGCGGUUUUUGACGGUACAGGCC



Furthermore, the necessity of handling nanomolar order concentrations required the creation of covalently linked RNPs sensors (cRNPs).[5] As previously reported in chapter 2, an hydrazine functionalized poly glycine chain was inserted at the C-terminus of the 5-carboxyfluorescein (5FAM) modified-Rev peptide and the coupling reaction occurred through a six member ring formation on the 3'-terminal ribose on the RRE RNA subunit.

3.5.2. DNA origami scaffold design

A rectangular DNA-origami structure (100nm x 96nm) that has five addressable cavities, has been proven to be a useful scaffold for locating RNP biosensors at specific addresses and modifying sensors population distribution: the circles represent the distance distribution from the center, varying he tag disposition is possible to increase or diminish the receptors density around the central cavity (**Figure 19**). On the left side above an hairpin motif was introduced as topological marker.

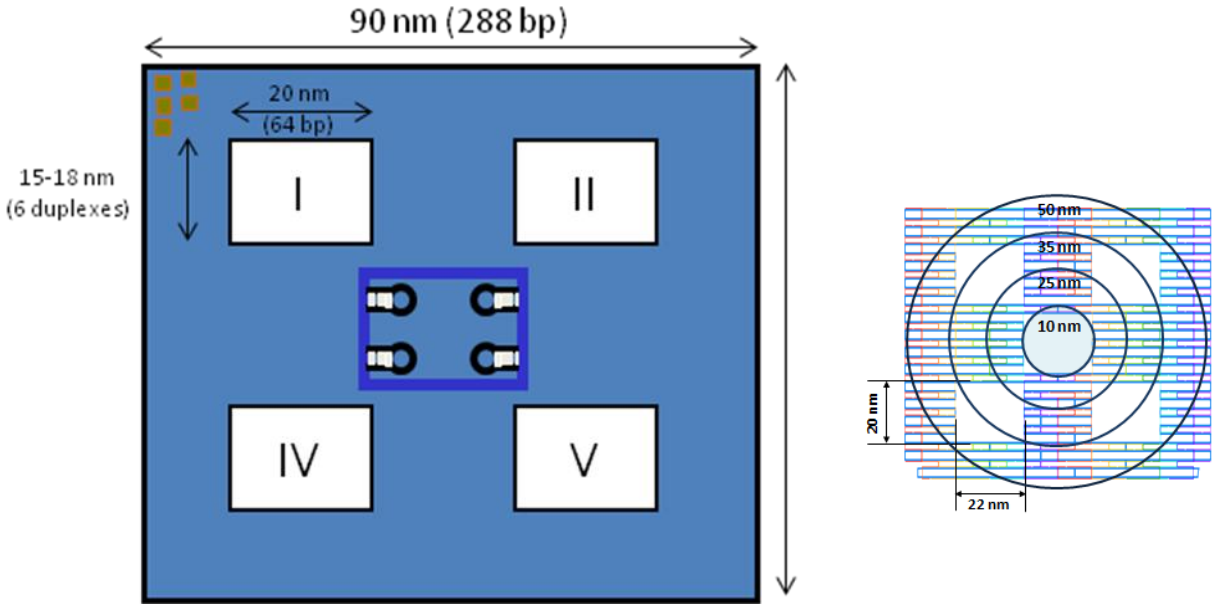


Figure 19: 5well-frame rectangular DNA origami scaffold design

With the aim of analyzing the clustering effect of ATP receptors on their affinity towards the target, different DNA origami switchboards were designed varying the number of binding sites located around the central cavity. Respectively, 4, 6, 12 and 18 binding sites (**Figure 20**).

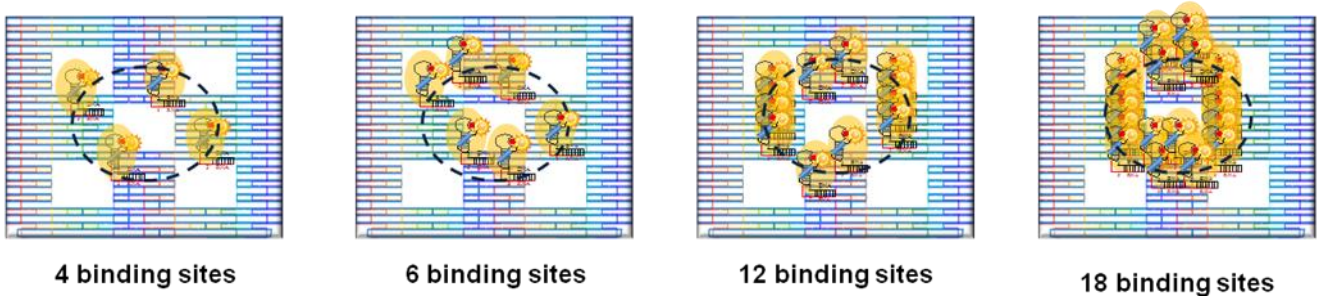


Figure 20: Density variation of cRNP sensors population around the central cavity on DNA origami

3.5.3. Atomic force microscope (AFM) analysis

3.5.3.1. Theory

Atomic force microscopy (AFM) was invented by Binnig et al. in 1986.[10] AFM is being used increasingly for the study of biomacromolecules at the single-molecule level.[10-12] It is a force-based technique that works by scanning the surface of the sample using the sharp tip of a cantilever. For imaging, the sample must be immobilized on a flat surface such as mica, glass, highly oriented polygraphite, or gold-coated surfaces. Topographical images are produced by monitoring the deflection of the cantilever caused by the tip-sample interactions during raster scanning of the tip over the surface. AFM is operated usually in three different modes: contact, tapping, and noncontact. Originally, samples were imaged in contact mode.[10] However, the contact mode is unsuitable for analysis of less robust biological samples, such as proteins, because the continuous contact between the tip and the sample leads to damage of the sample, which also often detaches from the surface on which it is adsorbed. In 1993, the tapping mode was invented,[13] in which the cantilever oscillates at a constant frequency and the tip-sample interaction is considerably reduced. Studies using tapping mode have demonstrated that it is highly effective for nondestructive imaging of biological samples. A schematic diagram of the instrumentation of a typical tapping-mode AFM is given in **Figure 21**. At the early stage of AFM, analysis was performed only in air. Observation in liquid was first reported in 1987, when researchers imaged the surface of a sodium chloride crystal in a paraffin oil environment.[14] Later developments produced a method suitable for single-molecule analysis of biomolecules in physiologically relevant solutions.

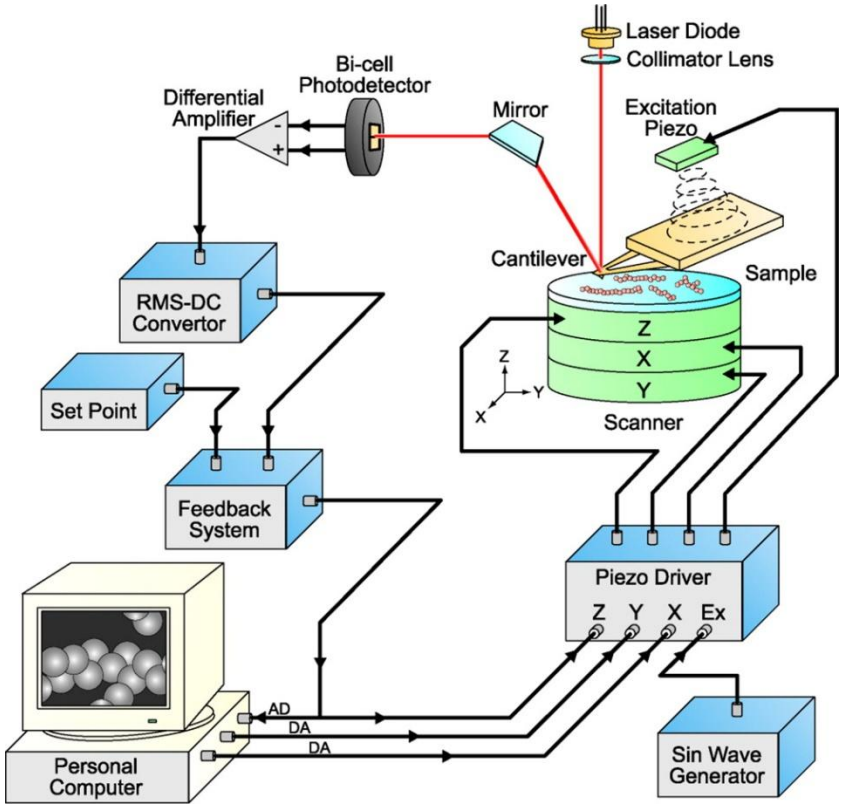


Figure 21: Graphical representation of the components and a typical instrumental setup of AFM.

Despite the success of AFM in imaging protein structure, detailed investigations of the structural dynamics of proteins have been hampered by the intrinsically slow scanning speed of conventional AFM. The scanning rates of commercially available AFM instruments usually range from several seconds to several minutes per frame, whereas biological reactions occur on the sub-second time scale or even faster. Several improvements have been implemented to achieve an AFM instrument that is much faster and compatible with the study of biological samples.[15-20] The early stage of AFM development was carried out by Hansma et al.,[16,21] who published the first report on application of AFM with improved scanning speed to study biomolecular activity in real-time.[17] Developments of Quate et al. led to imaging of a large area in a single scan.[22,23] Developments of the components for high-speed (HS) AFM were also reported by Fantner et al.[19] and Kindt et al. However, these improvements alone were insufficient to enable a HS-AFM that would revolutionize biomolecular analysis. One of the key breakthroughs for enabling HS-AFM was the use of small cantilevers.[24] Also, the design of stiff and compact piezo scanners in combination with methods from control theory have

improved the imaging speed.[25] Work by Ando et al. improved the performance of several components, including the cantilever, scanning stage, amplitude-to-dc converter, and dynamic PID controller. These developments culminated in realization of a first-generation HS-AFM instrument with an image acquisition speed of 12.5 frames/s.[26,27] Further developments enhanced the image acquisition speed and led to construction of a second-generation HS-AFM instrument with a scan speed of 20 frames/s.[28] Developments continued, and current laboratory-built HS-AFM instruments can record a movie with an imaging speed of ~33 frames/s and temporal and spatial resolutions compatible with the dynamic analysis of biological samples.[29,30]

3.5.3.2. cRNP sensors binding ability confirmation on DNA origami

The general approach for DNA origami annealing procedure requires preparing a solution of M13 (10nM), the staple strands mixture (5eq.) in DNA origami buffer (40mM Tris-HCl, 20mM acetic acid, 12.5mM $MgCl_2$) at pH 8.30. The solution so obtained is heat-shocked at 95ºC for 1minute and successively cooled down at 53ºC, the temperature is maintained stable for 30 minutes-1 hour in order to permit the annealing process and then reduced at 4ºC. Samples are purified by size-exclusion chromatography to get rid of excess amount of staple strands and analyzed by AFM image (**Figure 22**).

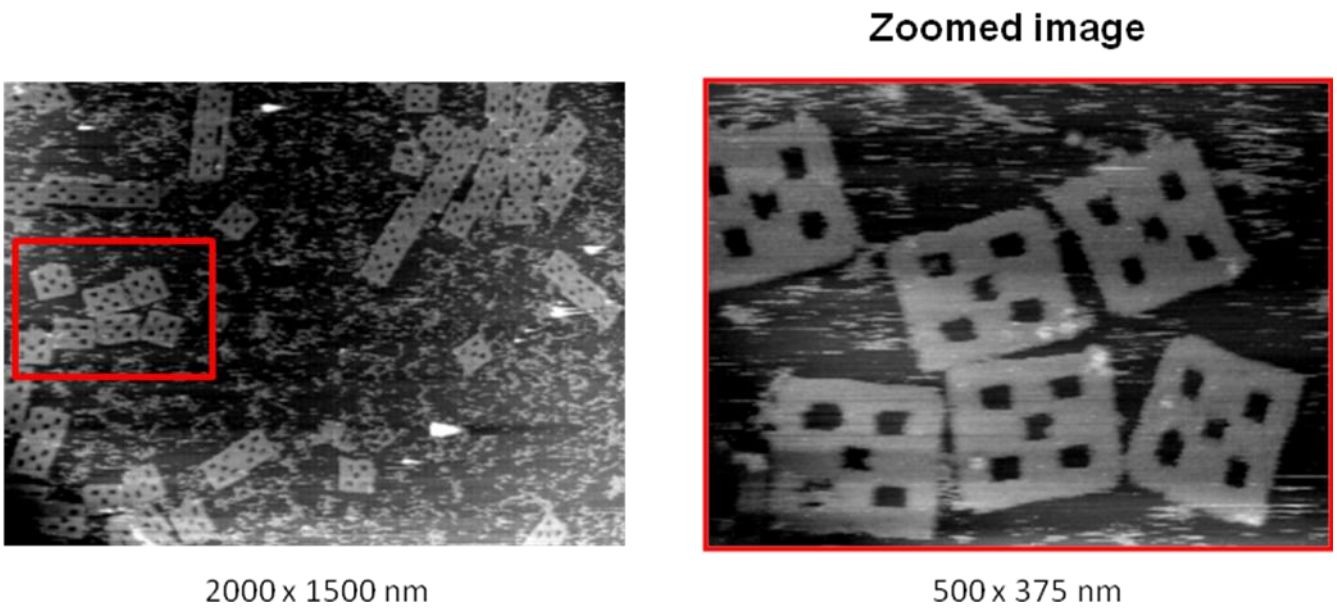


Figure 22: High speed AFM image of rectangular 5-well-frame DNA origami after purification

In order to hybridize cRNPs sensors on DNA origami, two protocols were adopted. The first one consists into two separated steps: first DNA origami was annealed (at pH 7.60) and purified by gel size-exclusion chromatography; in a second time the hybridization process was carried out incubating DNA origami with 3 equivalent of RNP sensors starting from 40^0C and decreasing the temperature with a ramp rate of 0.1^0C/3 secs. DNA origami hybridized was purified a second time by gel size-exclusion chromatography in order to remove the excess amount of non-hybridized cRNPs. An optimized protocol consists in a one-pot reaction in which DNA origami annealing and cRNPs hybridization are carried out in a time-product saving manner (**Figure 23**).

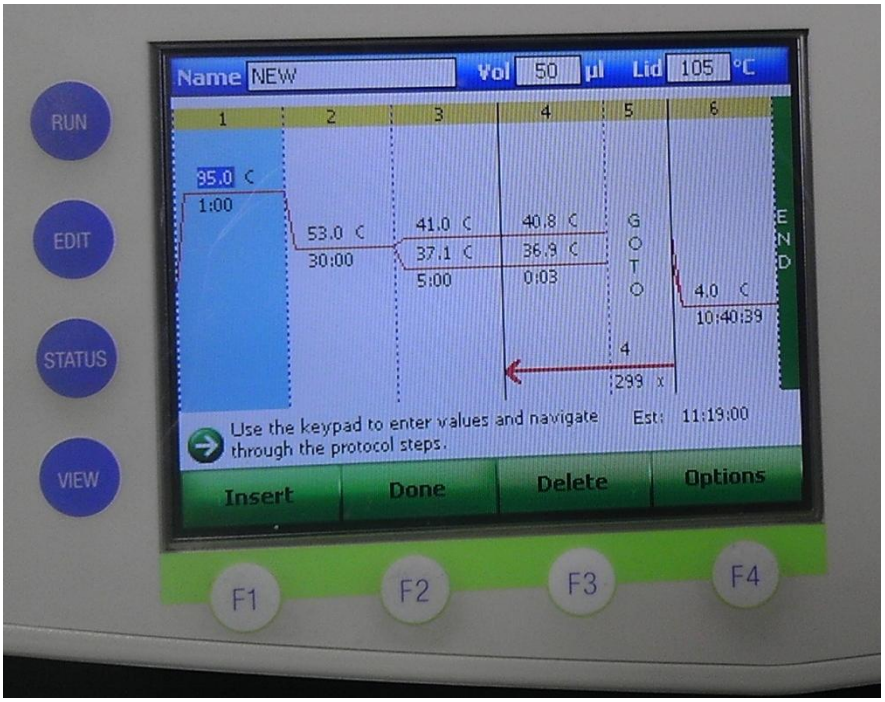


Figure 23: PCR machine set up optimized for the one-pot reaction

The cRNPs sensors (3eq.) are directly added to the DNA origami mixture after the heat-shock process; the annealing and hybridizations events occurs consecutively without need of intermediate purification steps.

The cRNPs sensors binding ability on DNA origami was confirmed both by gel image and AFM measurement as shown in **Figure 24** and **Figure 25.**

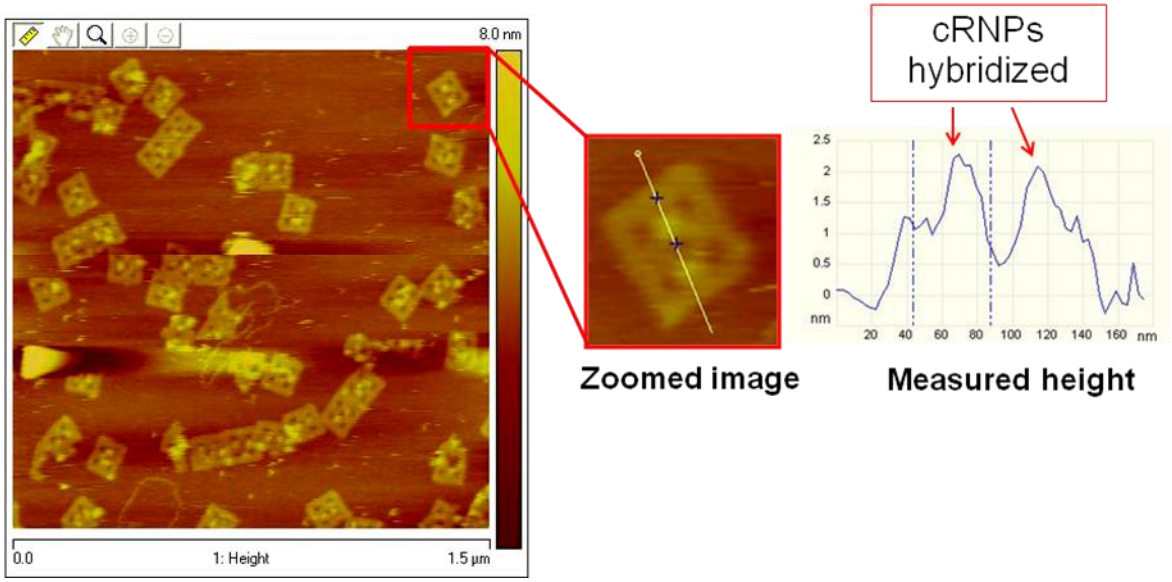


Figure 24: AFM image of rectangular 5-well-frame DNA origami hybridized with cRNPs (12 binding sites)

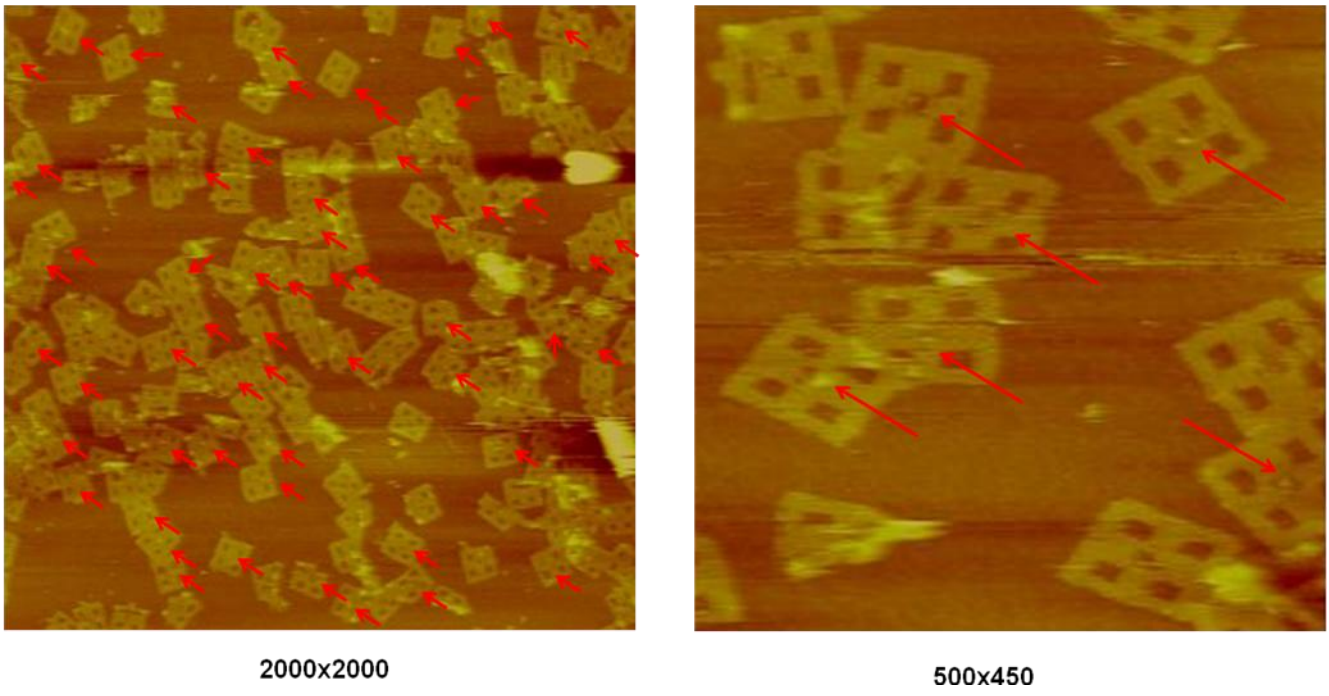


Figure 25: AFM image of rectangular 5-well-frame DNA origami hybridized with cRNPs (4 binding sites at the central cavity)

The AFM measurements, performed at pH 7.60, show the presence of clusters of cRNP sensors around the central cavity on DNA origami as expected; an increase of height around 1nm is also observed. Even though it was possible to discriminate between 12, 18 and four binding sites analyzing AFM images (**Figure 26**), the image resolution doesn't permit to evaluate the exact number of cRNPs bound. For this reason the binding yield was calculated by volume analyses on gel image (**Figure 27**). This technique also allowed us to evaluate the specificity of the binding event and its covalently nature not provable from AFM analyses. It was observed that the one pot reaction method produced DNA origami well formed and a binding yield ability around 85%. The non specific biding percentage, around 4%, can be considered negligible.

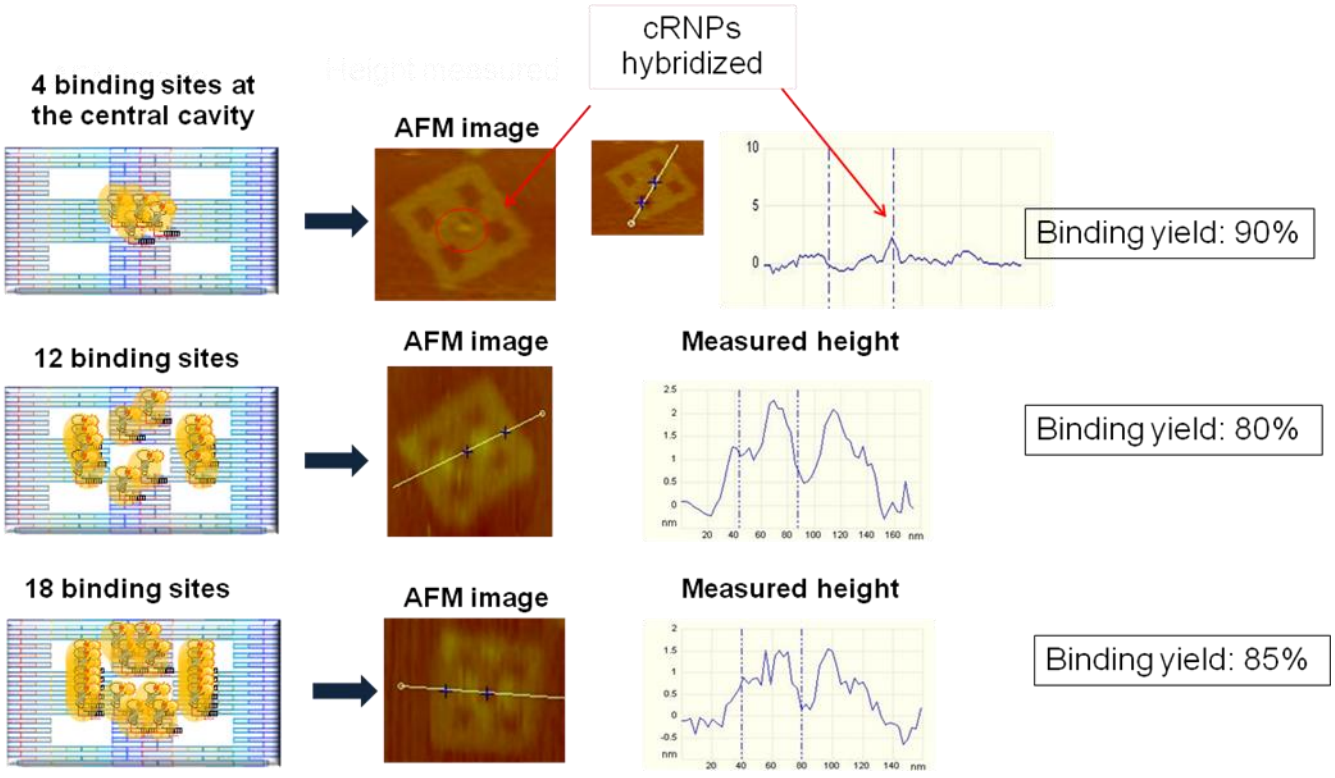
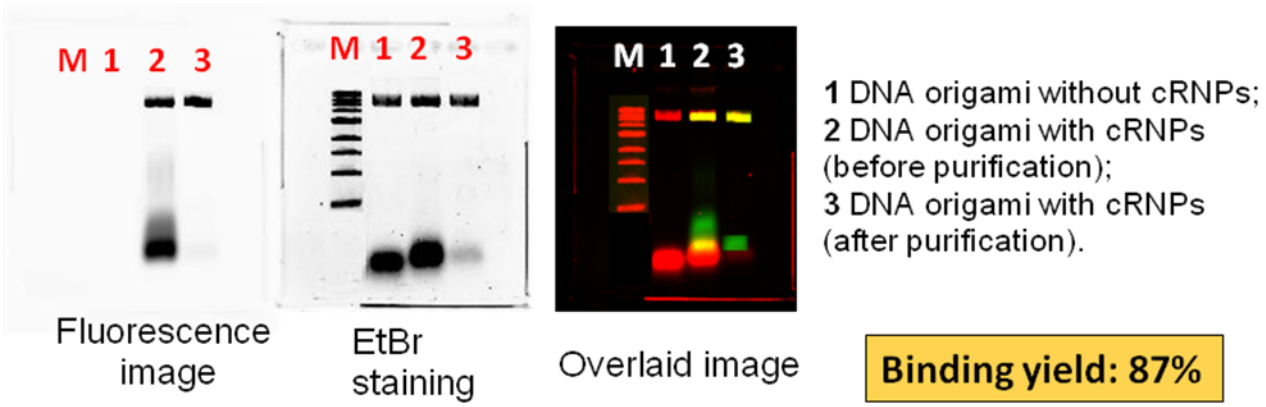
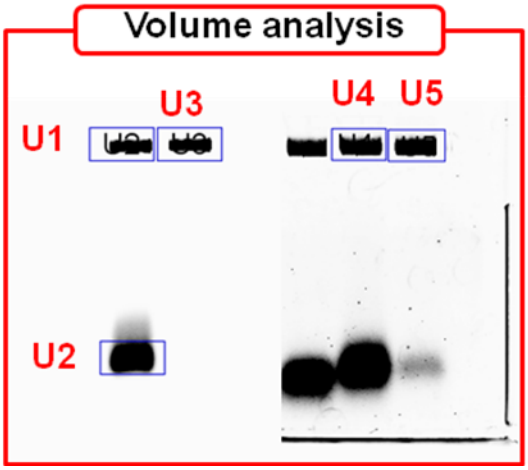


Figure 26: On the left side, schematic illustration of the binding sites distribution on DNA origami. In the centre, AFM image of hybridized cRNPs on DNA origami and measured height.





cRNP	F.I. on fluorescence	Concentration	F.I. on EtBr	Origami concentration	Yield
Total F. I. on fluorescence	**U1+U2**	**3eq.to DNA origami (900nM)**			
F. I. on DNA origami (without purif.)	**U1**	**Proportional to 3eq. (174uM)**	**U4**	**10nM**	**96.8%** **(174.4/180)**
F. I. on DNA origami (with purif.)	**U3**	**Proportional to U1**	**U5**	**Proportional to origami concentration**	**87%**

Figure 27: Yield evaluation on gel image: The Fluorescence image (Exc. Wv. 488nm) revealed the presence of 5-FAM tethering-cRNPs, the ethidium bromide (EtBr) image corresponds to the DNA origami staining. The overlaid image shows that the sample contains both

DNA origami and cRNPs giving confirmation of the effective hybridization event (yellow color). Moreover, calculating the fluorescence intensity of each band before and after purification, it is possible to evaluate the binding yield on DNA origami.

3.5.4. Orthogonal binding ability confirmation

With the aim of varying not only the sensors density but also combining different sensors on the same DNA origami platform, four different tags were designed. To confirm the orthogonal binding ability of our designed tag a gel image study was carried out. In the following example, An16 tag04 cRNP was incubated with both, tag01-modified DNA origami and tag04-modified DNA origami. Only in the presence of the complementary tag04 a successful hybridization event was observed (**Figure 28**).

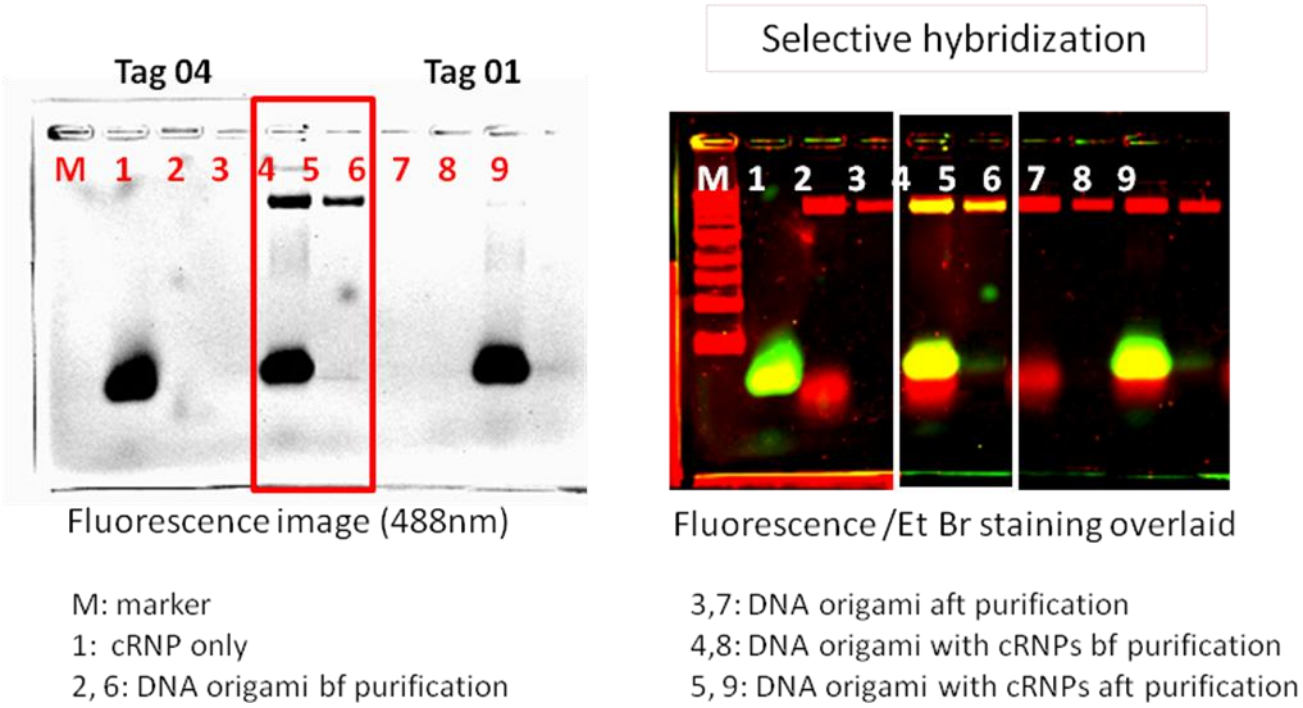


Figure 28: Orthogonal binding ability check on gel image. On the left side, the fluorescence image (Exc. Wv. 488nm) shows the presence of An16tag04 cRNP receptors. On the right side, a fluorescence/EtBr staining overlaid image confirms the presence of both DNA origami and cRNPs (yellow color) only in the presence of the complementary tag04 on DNA origami. In the case of tag01-modified DNA origami, only a red band, corresponding to the DNA alone, is observed indicating that no binding event occurred.

From this study it was evicted that the four designed tag are able to bind specifically their DNA-complementary target in an orthogonal manner.

3.5.5. Dissociation constant evaluation on DNA origami

In order to explore the behavior of cRNP receptors on DNA origami, a dissociation constant study was performed. The K_D was evaluated from ATP titration experiments on a microplate reader. DNA origami hybridized with cRNPs was prepared as previously reported; after gel size-exclusion chromatography purification, the samples were incubated 30minutes at 4ºC in the presence of increasing amount of ATP. Successively, they were analyzed on a plate reader (Exc. Wv. 470nm) and the fluorescence intensity change plotted against ATP concentration (**Figure 29-32**). cRNP concentration was assessed by volume analysis on gel image.

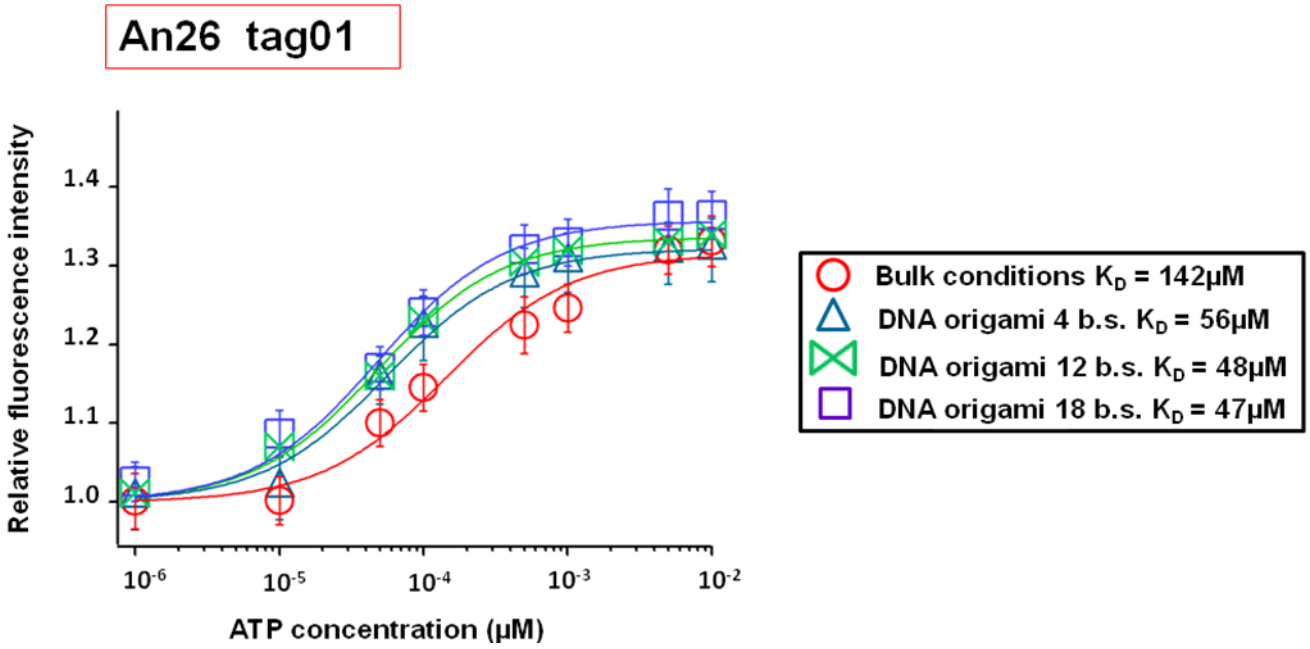


Figure 29: K_D evaluation on A26tag01 cRNP. Respectively, in bulk conditions (red circle); 4 binding sites (blue triangle); 12 binding sites (green ribbon); 18 binding sites (purple square).

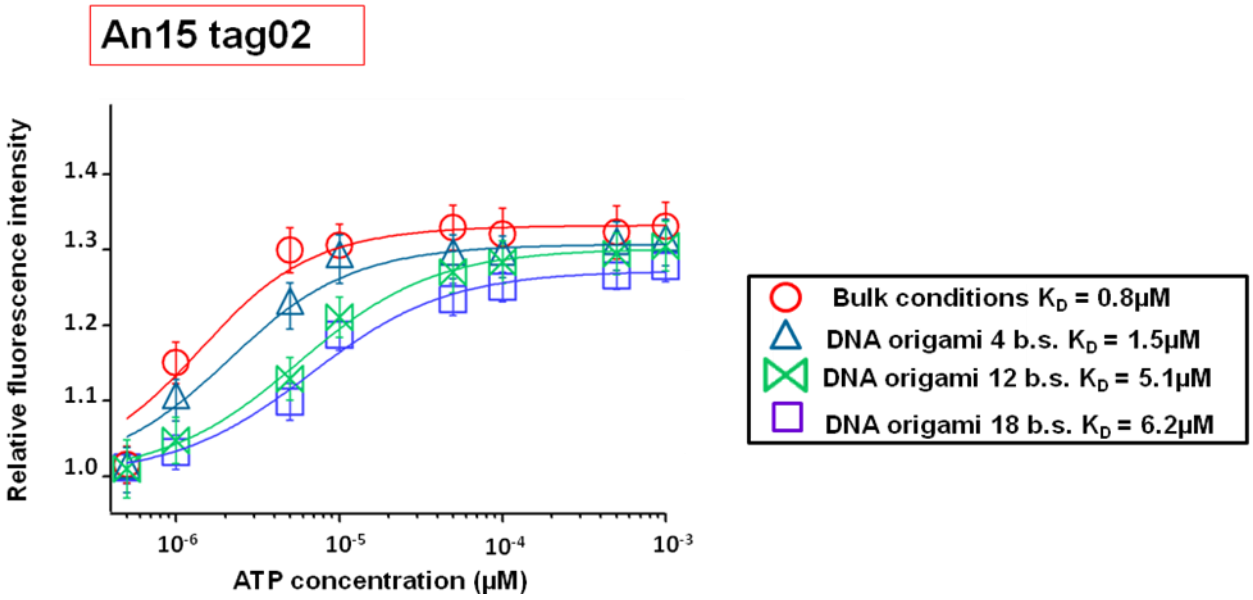


Figure 30: K_D evaluation on An15tag02 cRNP. Respectively, in bulk conditions (red circle); 4 binding sites (blue triangle); 12 binding sites (green ribbon); 18 binding sites (purple square).

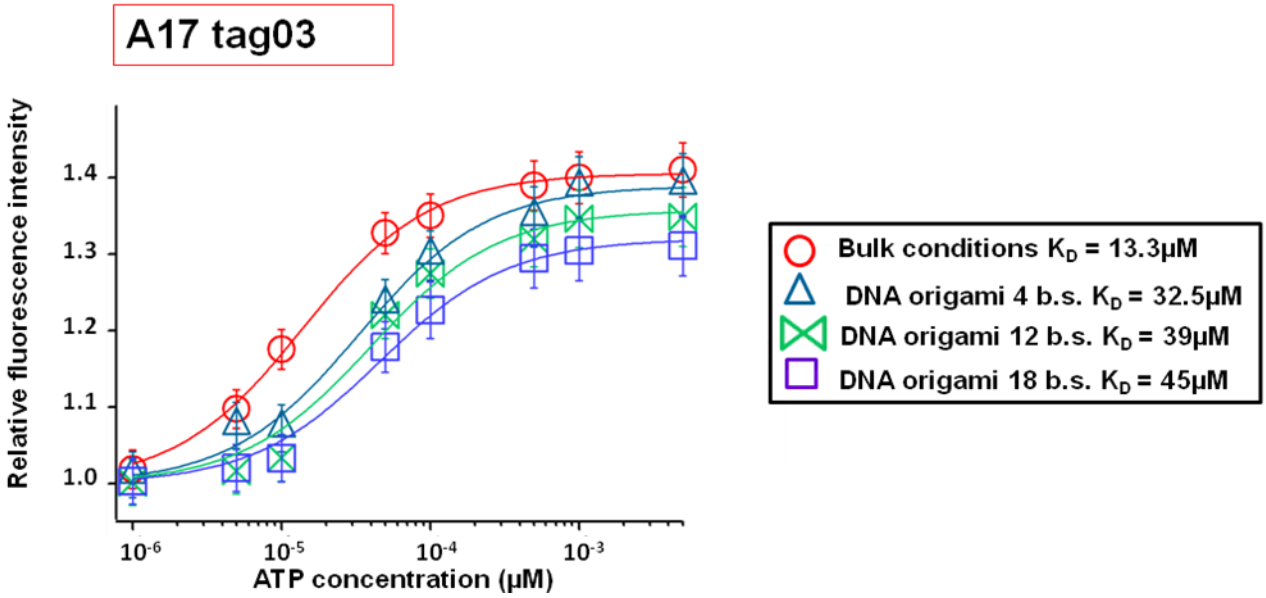


Figure 31: K_D evaluation on A17tag03 cRNP. Respectively, in bulk conditions (red circle); 4 binding sites (blue triangle); 12 binding sites (green ribbon); 18 binding sites (purple square).

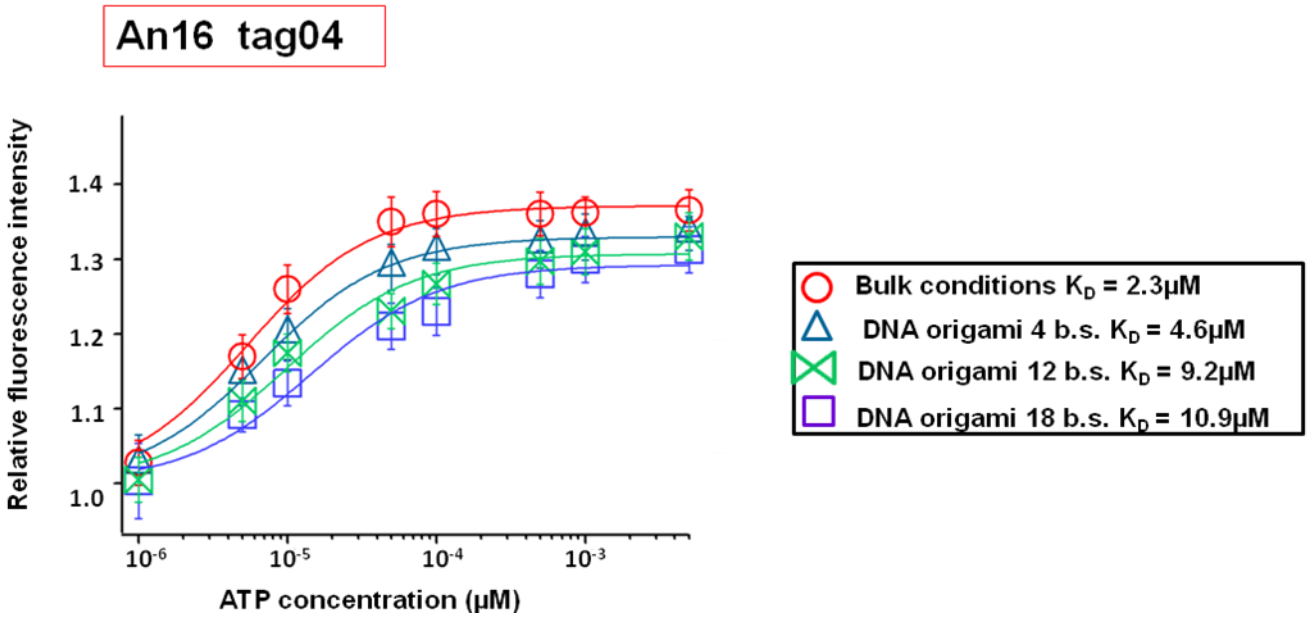


Figure 32: K_D evaluation on An16tag04 cRNP. Respectively, in bulk conditions (red circle); 4 binding sites (blue triangle); 12 binding sites (green ribbon); 18 binding sites (purple square).

cRNP/K_D (µM)	Bulk conditions	On DNA origami 4 binding sites	On DNA origami 12 binding sites	On DNA origami 18 binding sites
A26 tag01	142	56	48	47
A17 tag03	13.3	32.5	39	45
An15 tag02	0.8	1.5	5.1	6.2
An16 tag04	2.3	4.6	9.2	10.5

From this analysis it is evicted that the ATP binding affinity of covalent RNPs is retained after hybridization on DNA origami. Moreover a change in the K_D behavior is noticeable when the cRNP receptors are hybridized on DNA origami. A similar tendency could be underlined for An16tag04, An15tag02 and A17tag03. A26tag01, on the contrary, shows an opposite behavior. In order to acquire a

deeper understanding of the cluster effect on cRNPs receptors on DNA origami, a kinetic study was performed.

3.5.6. Kinetic measurements on DNA origami

Kinetic studies were carried out on A26tag01 and An16tag04 cRNPs evaluating the kinetics of the ligand binding event upon ATP titration in bulk conditions, on DNA origami 18 binding sites, 12 binding sites and 6 binding sites. The experiments were performed on a microplate reader; after 30 minutes incubation at room temperature, ATP was added and immediately the spectra recorded at 525nm (λ_{MAX}). The obtained saturation curves show that the ligand-binding event is a slow process that reaches saturation around 1 hour (from **Figure 33** to **Figure 42**).

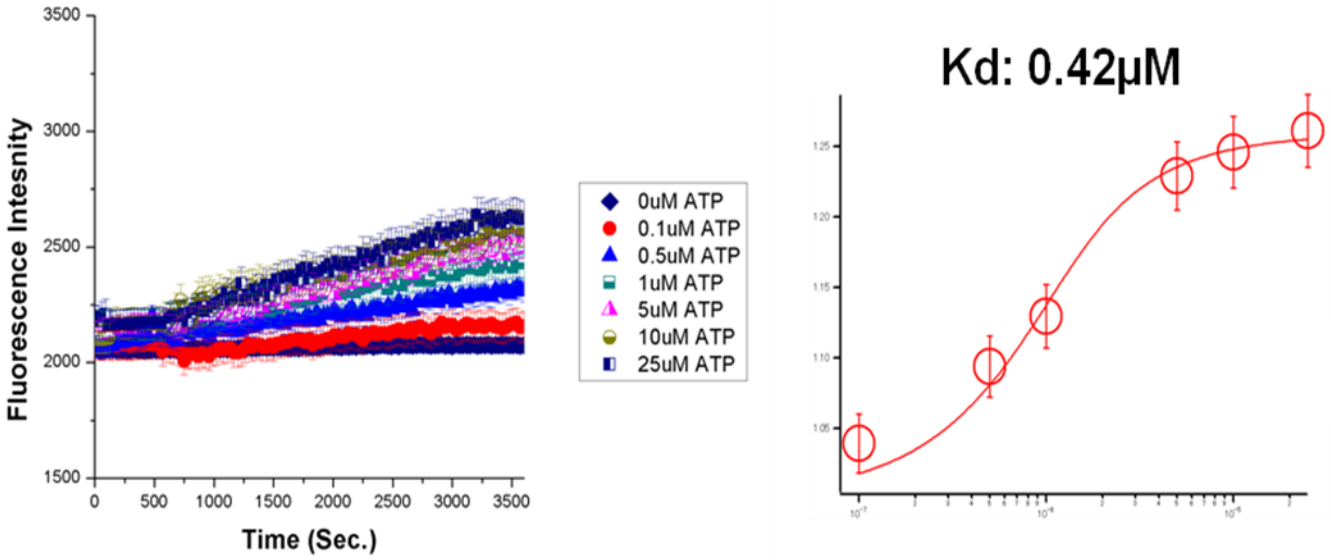


Figure 33: Fluorescence intensity change spectra upon ATP titration on An16tag04 in bulk conditions (ATP concentration range: 0 µM; 0.1 µM; 0.5 µM; 1 µM; 5 µM; 10 µM; 25 µM). Kd evaluated at the saturation point.

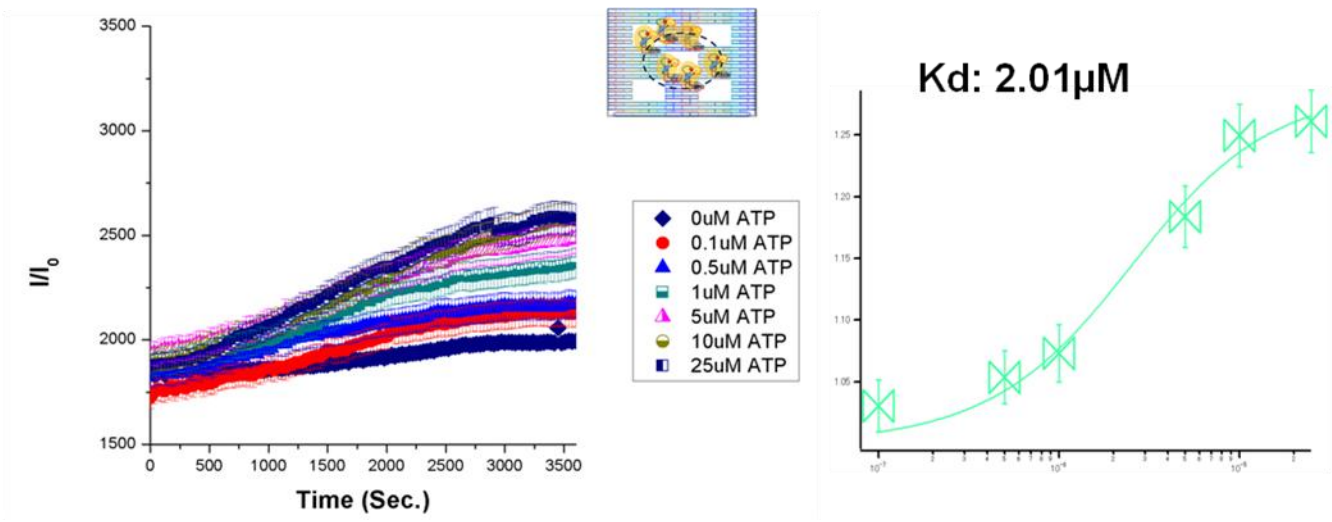


Figure 34: Fluorescence intensity change spectra upon ATP titration on An16tag04 on DNA origami 6 binding sites (ATP concentration range: 0 μM; 0.1 μM; 0.5 μM; 1 μM; 5 μM; 10 μM; 25 μM). Kd evaluated at the saturation point.

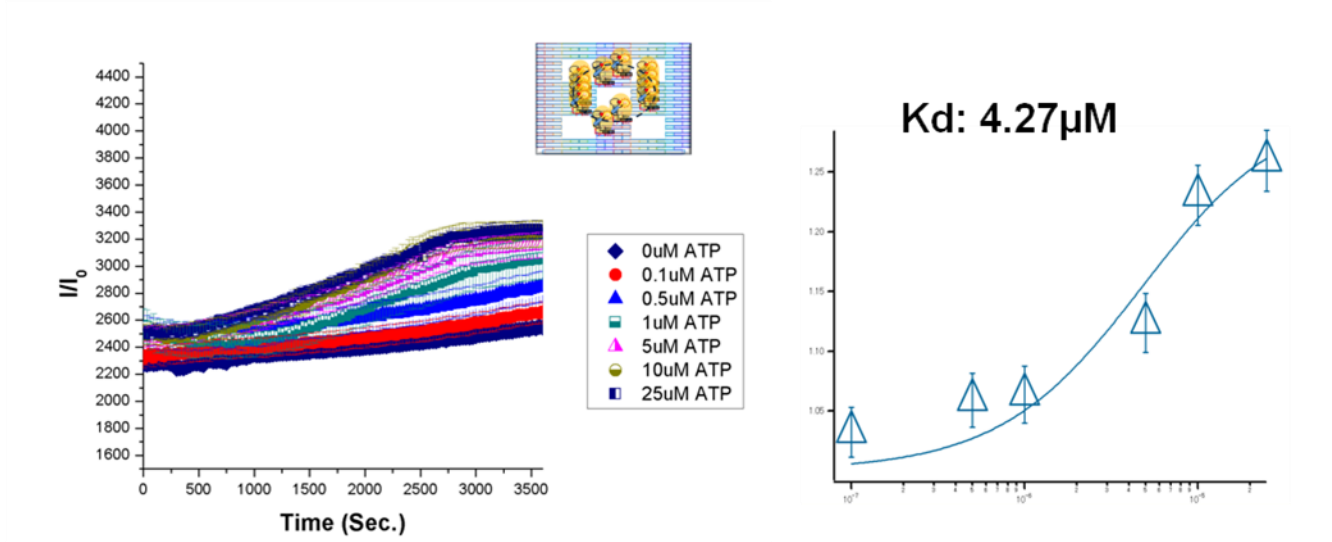


Figure 35: Fluorescence intensity change spectra upon ATP titration on An16tag04 on DNA origami 12 binding sites (ATP concentration range: 0 μM; 0.1 μM; 0.5 μM; 1 μM; 5 μM; 10 μM; 25 μM). Kd evaluated at the saturation point.

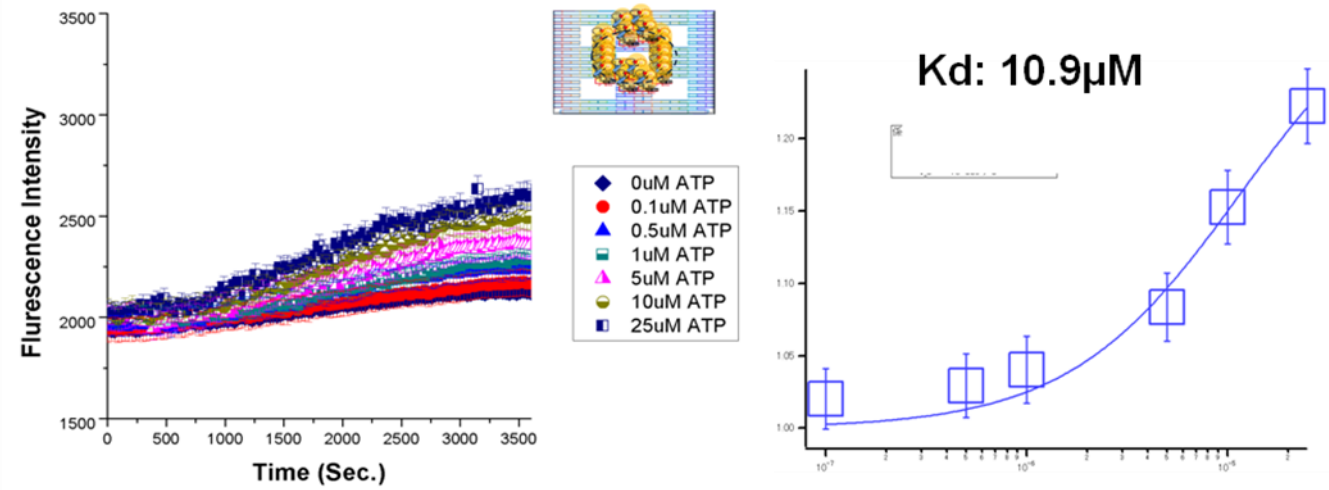


Figure 36: Fluorescence intensity change spectra upon ATP titration on An16tag04 on DNA origami 18 binding sites (ATP concentration range: 0 µM; 0.1 µM; 0.5 µM; 1 µM; 5 µM; 10 µM; 25 µM). Kd evaluated at the saturation point.

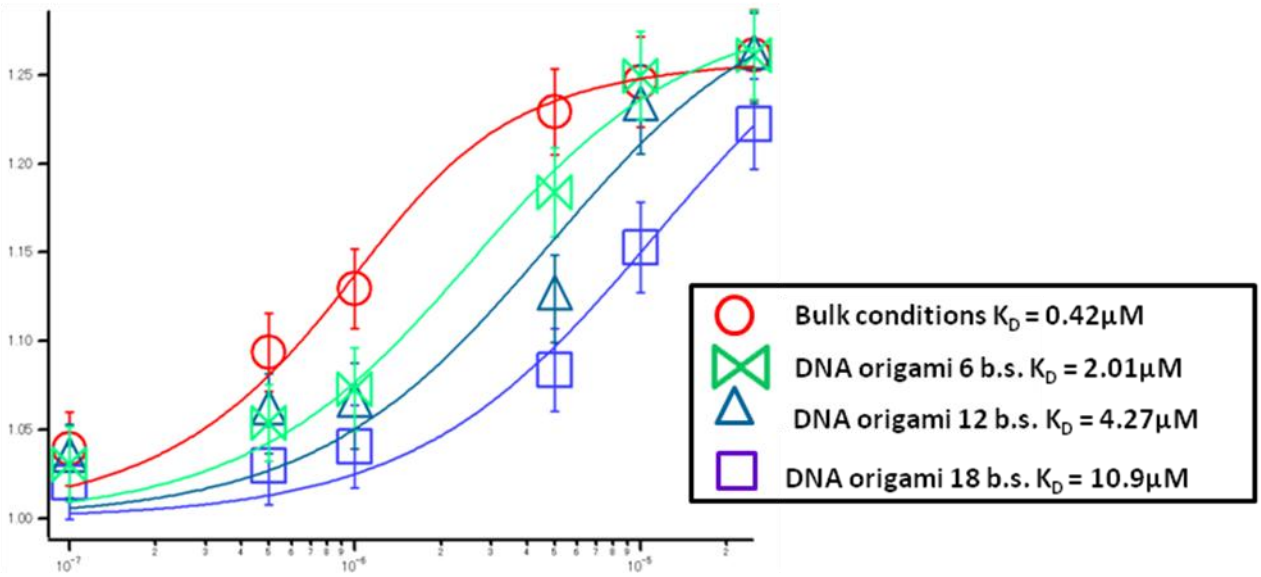


Figure 37: Total K_D evaluation on An16tag04. Respectively, red circles (bulk conditions); green ribbons (6 binding sites); blue triangles (12 binding sites); purple squares (18 binding sites).

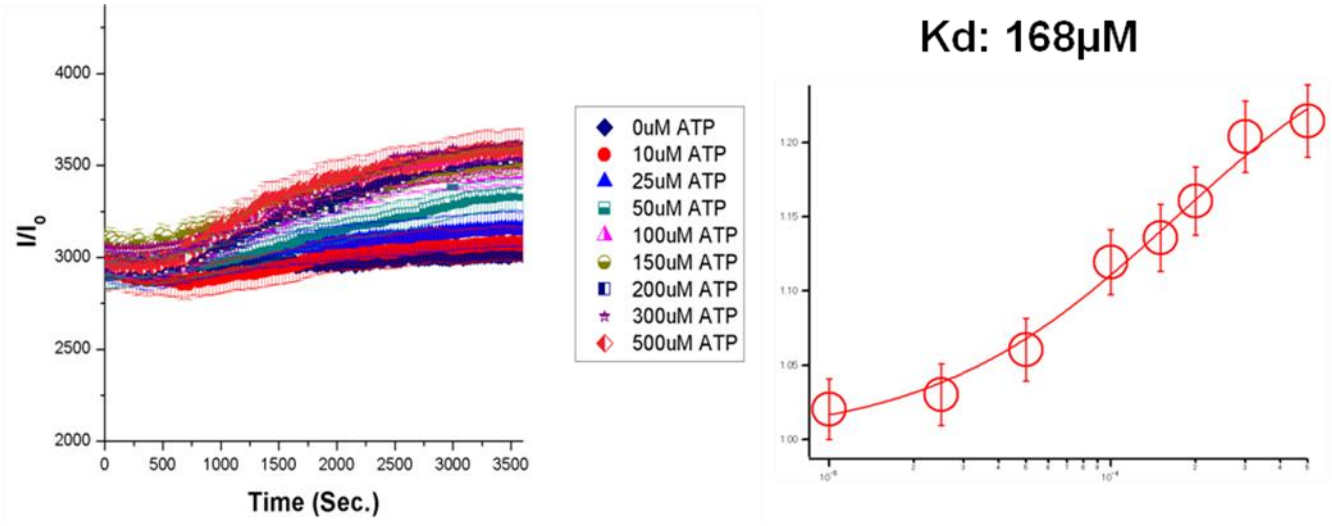


Figure 38: Fluorescence intensity change spectra upon ATP titration on A26tag01 bulk conditions (ATP concentration range: 0 µM; 10 µM; 25 µM; 50 µM; 100 µM; 150 µM; 200 µM; 300 µM; 500 µM). Kd evaluated at the saturation point.

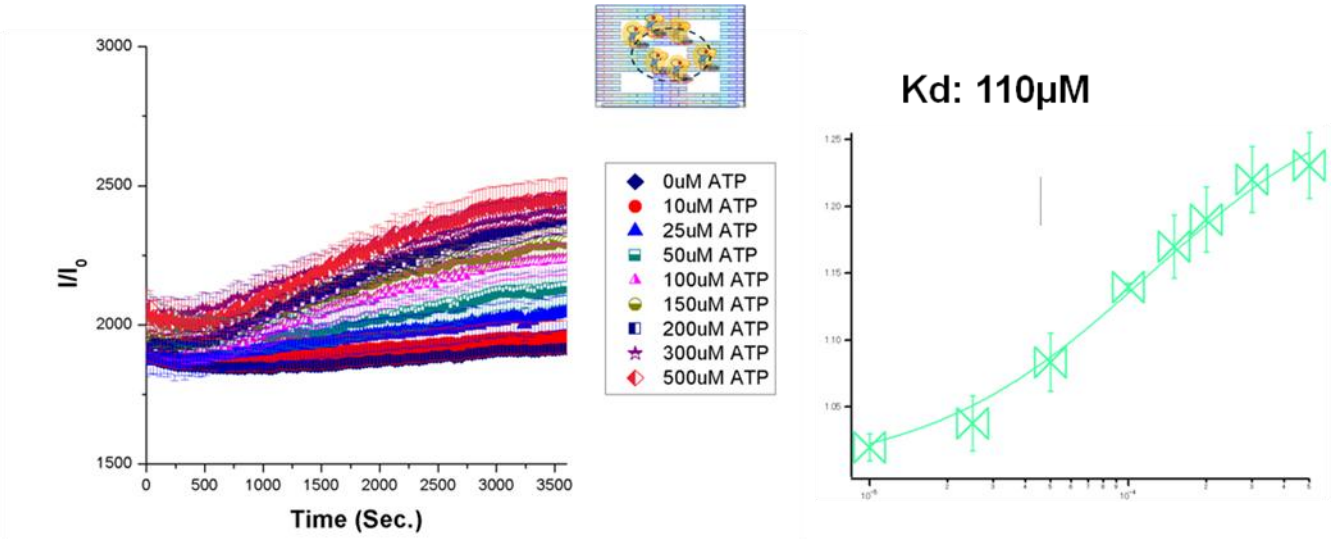


Figure 39: Fluorescence intensity change spectra upon ATP titration on A26tag01 on 6 binding sites (ATP concentration range: 0 µM; 10 µM; 25 µM; 50 µM; 100 µM; 150 µM; 200 µM; 300 µM; 500 µM). Kd evaluated at the saturation point.

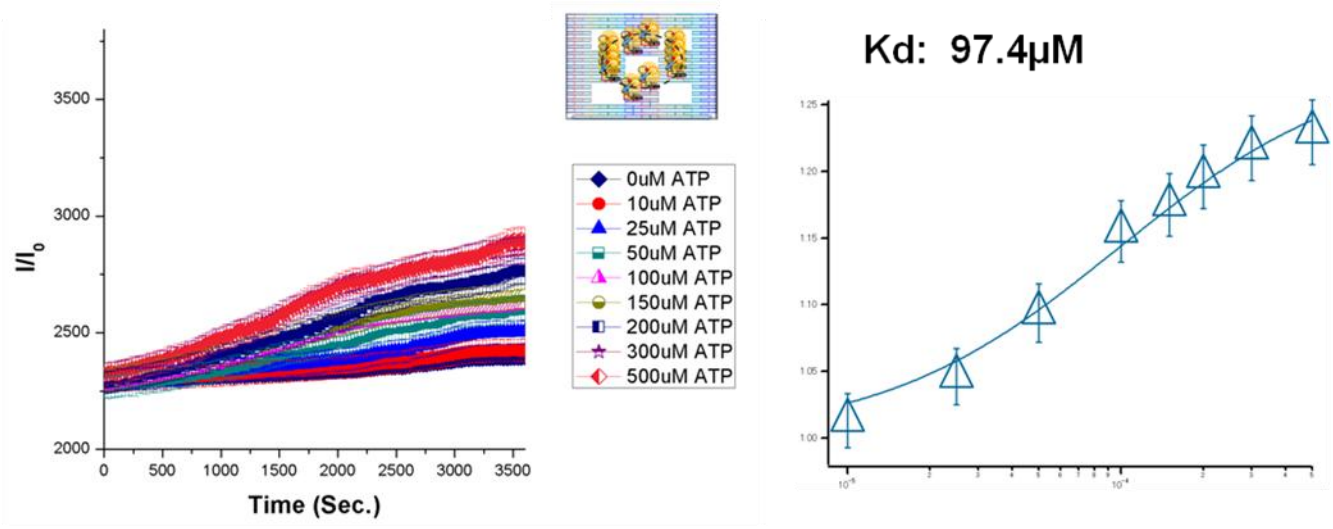


Figure 40: Fluorescence intensity change spectra upon ATP titration on A26tag01 on 12 binding sites (ATP concentration range: 0 μM; 10 μM; 25 μM; 50 μM; 100 μM; 150 μM; 200 μM; 300 μM; 500 μM). Kd evaluated at the saturation point.

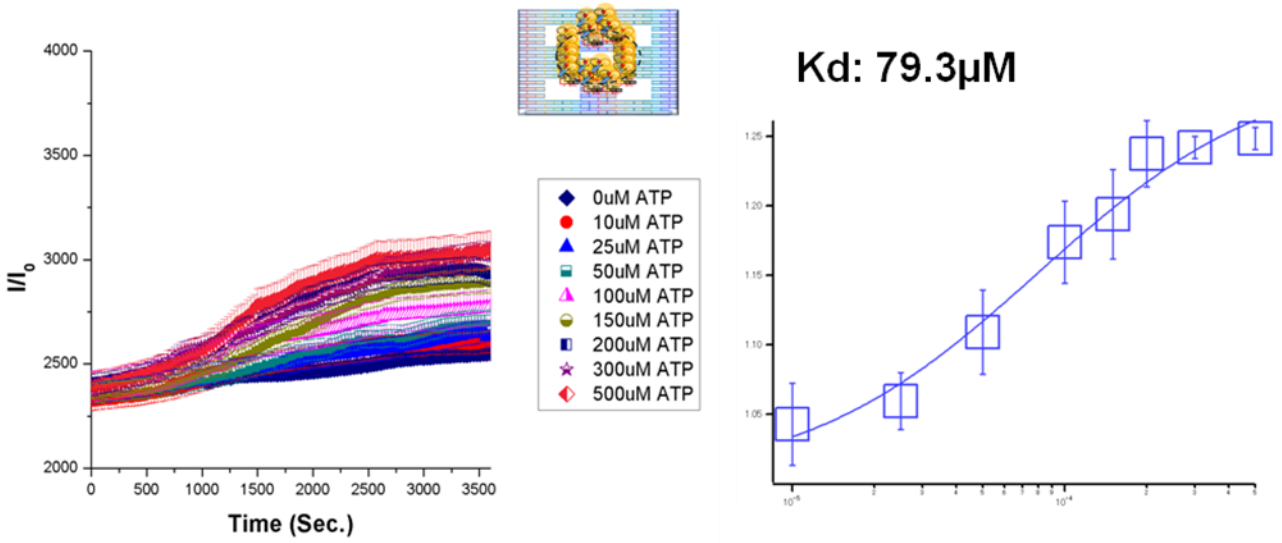


Figure 41: Fluorescence intensity change spectra upon ATP titration on A26tag01 on 18 binding sites (ATP concentration range: 0 μM; 10 μM; 25 μM; 50 μM; 100 μM; 150 μM; 200 μM; 300 μM; 500 μM). Kd evaluated at the saturation point.

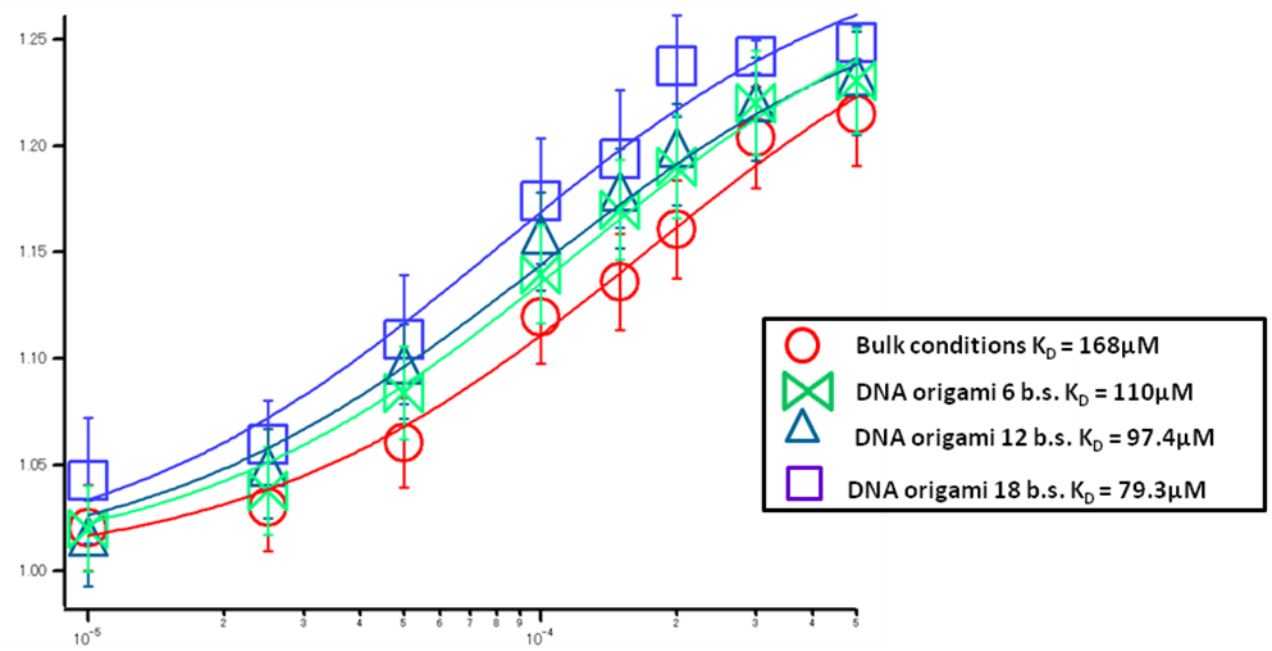


Figure 42: Total K_D evaluation on A26tag01. Respectively, red circles (bulk conditions); green ribbons (6 binding sites); blue triangles (12 binding sites); purple squares (18 binding sites).

The K_D values, obtained from the data at the saturation point (3600 seconds), are consistent with the results achieved from the ATP titrations previously performed after incubation at 4 degree Celsius.

3.6. Conclusions

As discussed in chapter 2, stepwise manner constructed covalently linked RNP receptors allowed ATP detection at low concentration. In this work we designed molecular switchboards with programmable placement of multiple ATP-binding biosensors at defined locations on DNA origami by using sequence specific tags as adaptors. The structure of molecular switchboard was confirmed by AFM analyses and the tag specificity was evaluated trough gel image studies. The fluorescent responses of the ATP-binding molecular switchboards revealed that the cRNP receptors maintain their activity once hybridized with DNA origami. Moreover, varying the receptors density a change in the sensitivity towards the target was observed. Receptors localization and clustering may have a dramatic influence on signaling. This phenomenon aroused a huge interesting in scientific environments but unfortunately, nowadays, the availability of *in vitro* models for studying receptors cluster effect are very limited. Molecular switchboard based on DNA origami may represent a convenient tool for analyzing receptors clusters outside the cell. In order to acquire a deeper understanding of the cluster effect on cRNP receptors on DNA origami, a kinetic study will be performed. Furthermore, a future development of this project considers the combination of receptors presenting different sensitivity towards the target on the same switchboard with the purpose to analyze how this might affect the target binding curve.

3.7.Materials and methods

3.7.1. Materials

DNA templates, respectively, A26tag01, An15tag02, A17tag03, An16tag04, and primers (for-A26ta01; for-An15tag02; for-A17tag03; for-An16tag04; Rev-S2) were purchased from Sigma Aldrich. The DNA amplification kit (5X Prime Star buffer (Mg^{2+} plus); 25mM dNTP mixture; Prime Star HS DNA Polymerase enzyme 250U 2.5U/µl) was bought from TaKaRa. T7-Scribe Standard RNA NT Kit which consists of 10X T7-Scribe transcription buffer; RNase-free water; RNase-free DNase 1 enzyme 1u/µL; T7-Scribe enzyme; 100mM Dithiothreitol and 25mM rNTP mixture, was purchased from CellScript. INC. Nacalai Tesque provided 1mol/L Tris-HCl buffer solution (pH 7.60); EtOH 99.5%; Isopropanol; 0.5M EDTA pH 8.00. TBE buffer components: Ethylenediaminetetraacetic Acid (Edetic Acid, EDTA); Boric Acid (Nuclease and Protease tested);Tris(hydroxymethyl)aminomethane (Neclease and Protease tested) and APS (Aminoperoxysufate) were obtained from Nacalai Tesque. Urea was bought from Wako. Sigma Aldrich supplied Oligo DNA strands for DNA origami; M13mp ss DNA was acquired from New England BioLabs.

3.7.2. DNA amplification

DNA template amplification solution consists of :

30 µl 5x Prime Star buffer (Mg^{2+}plus);

12 µl 25 µM dNTP mixture;

15 µl 20 µM for-primer;

15 µl 20 µM Rev-S2 primer;

73.5 µl dH_2O;

3 µl 0.2 µM template cDNA;

1.5 µl Prime Star HS DNA Polymerase enzyme 250U 2.5U/µl.

The solution was then tripartited into PCR microtubes (50µL each) In order to assess the optimum cycle number for DNA template amplification, a PCR cycle check was performed: protocol applied: 98ºC 10sec→55ºC 30sec→72 ºC 30sec; 24 cycles. In the example reported in **Figure 43**, 16 cycles were selected for obtaining the pure desired DNA sequences.

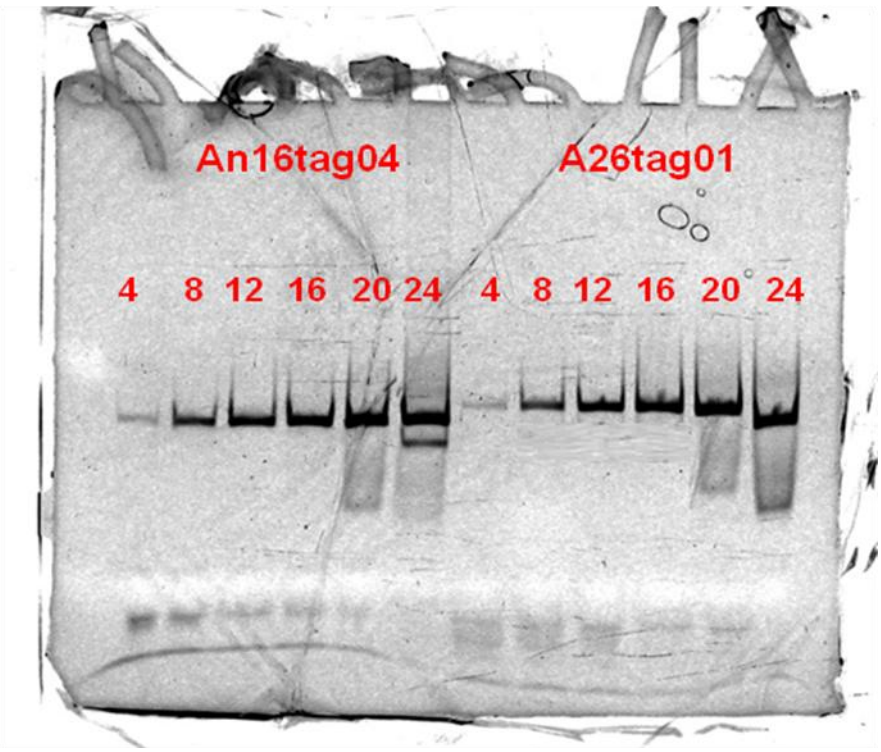


Figure 43: PCR cycle check on An16tag04 and A26tag01 DNA template. The numbers in red correspond to the cycle number.

Amplified DNA was precipitate by EtOH 2.5 eq. +10% AcONa pH 5.2 (samples centrifuged for 20 minutes at 150x100rpm); successively washed with 70% EtOH (centrifuged 15minutes at 150x100rpm) dried up under vacuum and redissolved in 20µl TE buffer. The sample purity was checked by nature 8% gel bisacrylamide gel as shown representatively in **Figure 44**.

Figure 44: DNA gel purity check

3.7.3. DNA transcription

RNA transcription mixture for 1µg DNA:

10 µl 10xT7 reaction buffer;

30 µl 25mM rNTP mixture;

10 µl 100mM dtt;

40 µl RNase water + Template DNA;

2.5 µl T7 enzime.

The samples were left at 37 ºC for 3 hours and then stored at 4 ºC , to stop the reaction 1µl DNase was added (20min 37 ºC). RNA was extracted using a phenol/chloroform solution and then precipitated using EtOH. Purification was accomplished using 8% urea (6M) gel. Pure RNA was extracted using 500µl TE buffer and centrifuged for 45min at 150x100rpm. It was, subsequently, precipitated using 2-isopropanol and gel-filtrated on a Micro Bio-Spin 6 Chromatography Column (4 min 10x100G). The purity was confirmed by gel check (**Figure 45**).

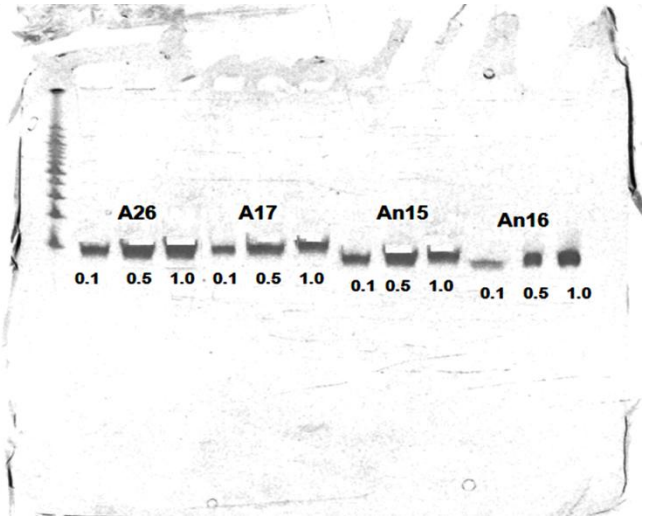


Figure 45: RNA purity check by 8% denatured bis-acrylamide gel.

3.7.4. DNA origami preparation

DNA origami preparation procedure was amply described at page 104-105; in this chapter I will focus on the gel size exclusion purification step optimization. With the aim of obtaining the highest DNA origami purity without detrimental effects on the yield, columns presenting filters of several materials and different pour size were compared by 1% agarose gel image as shown in **Figure 46** and **Figure 47**.

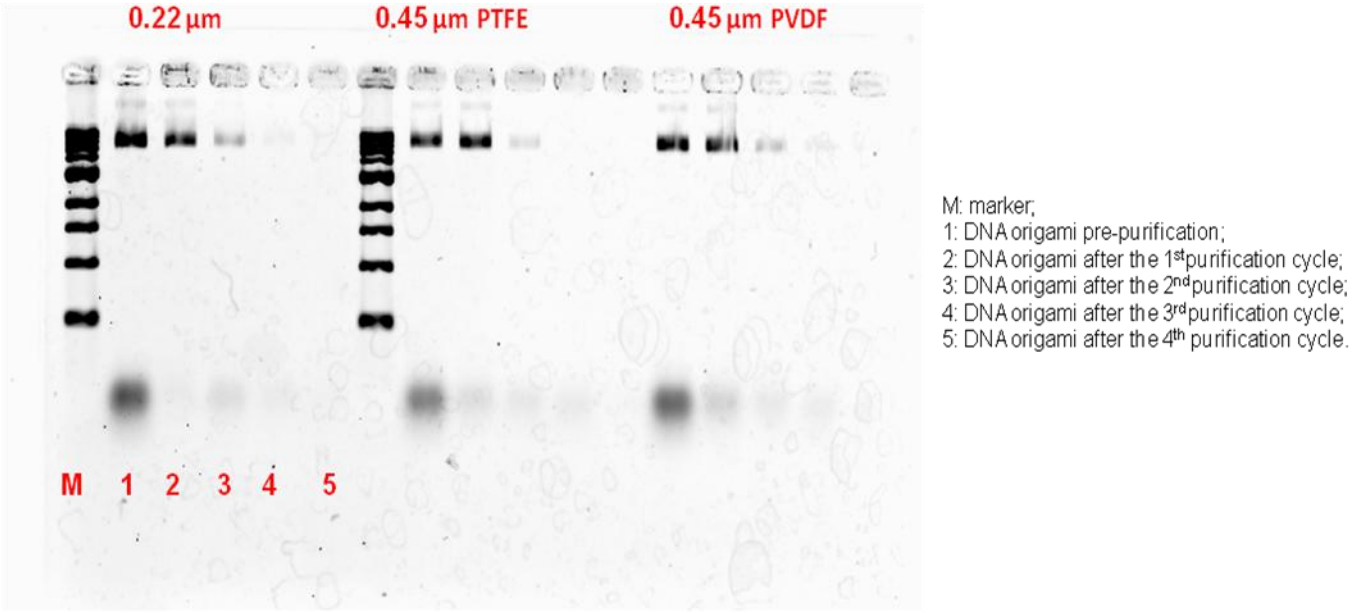


Figure 46: 1% agarose gel image (ETBr staining) of DNA origami pre and post gel purification step

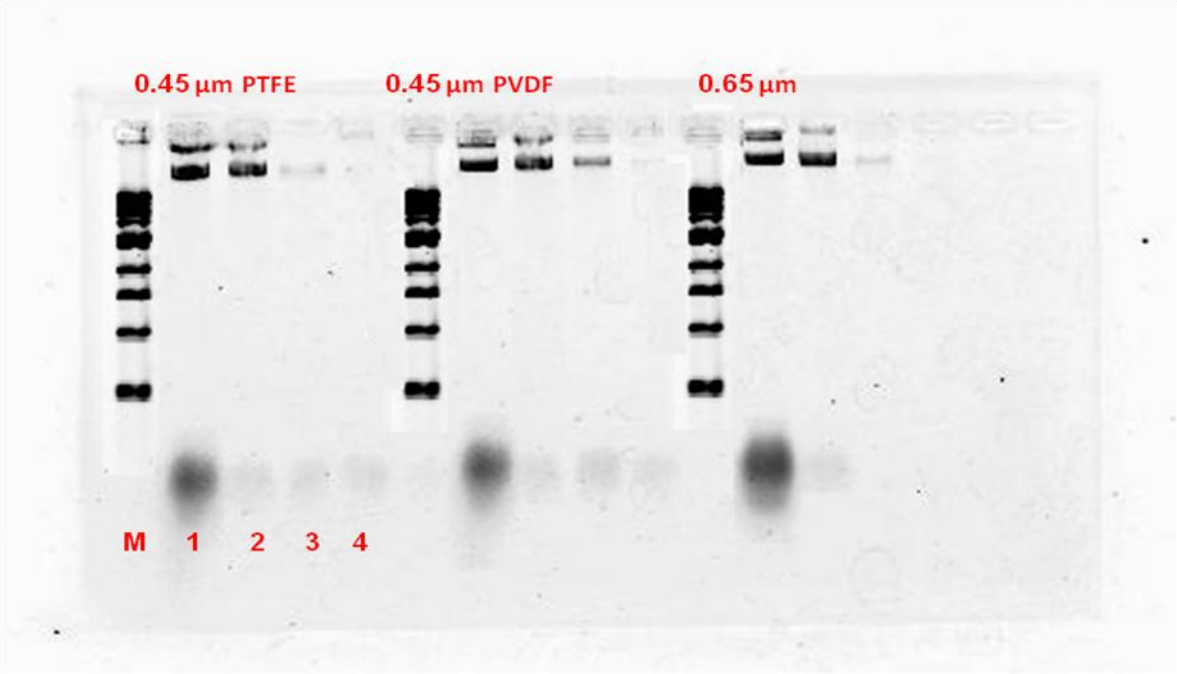


Figure 47: 1% agarose gel image (ETBr staining) of DNA origami pre and post gel purification step

In a second time also the gel media was analyzed. In **Figure 48** are reported three media having increasing porosity ranges.

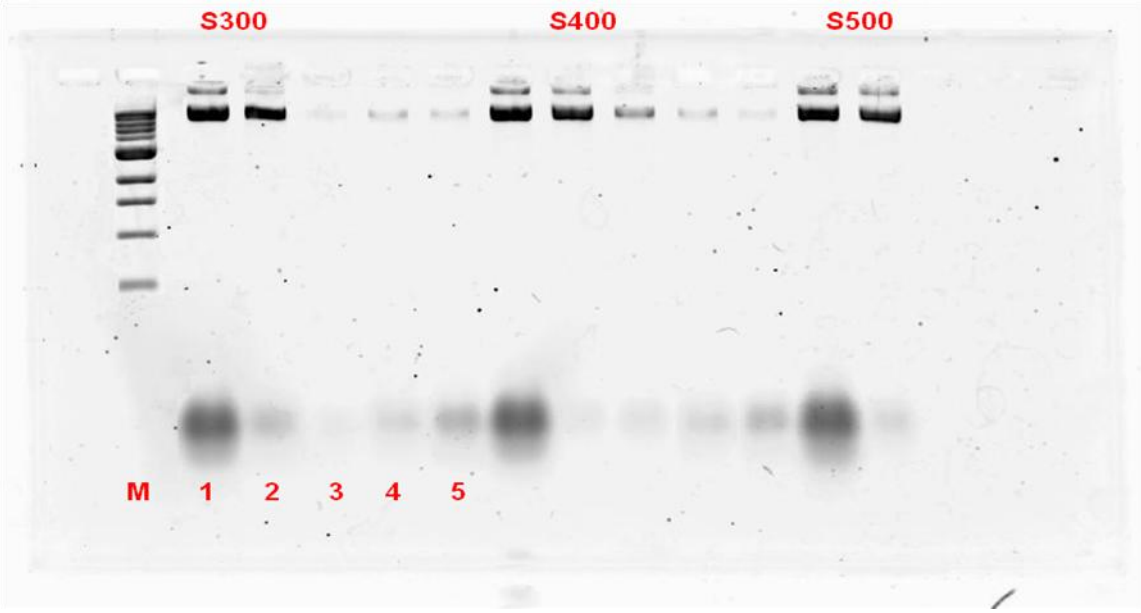


Figure 48: 1% agarose gel image (ETBr staining) of DNA origami pre and post gel purification step

It could be observed that the results are almost all comparable; in this optic we decided to use the combination that gave us the best yield/purity ratio: S400 gel and PVDF 0.65µm column.

3.7.5. **ATP titration on cRNPs : K_D measurement**

Covalently linked RNPs were analyzed both in bulk and on DNA origami surface in DNA origami buffer conditions (40mM Tris-HCl, 20mM acetic acid, 12.5mM $MgCl_2$ +0.005%Tween 20) at pH 7.60. The successful hybridization event and the approximate yield were evaluated from volume analyses on gel image and AFM analyses. Recovered cRNPs were then incubated at 4ºC with increasing concentration of ATP neutralized with NaOH on a microplate platform (30µL each sample). The fluorescence change measurements were carried out using Tecan Infinite Pro 200 plate-reader exciting the samples at 470nm; The Emission spectrum was scanned in a range from 500 to750nm at 4°C degree (**Figure 49**).

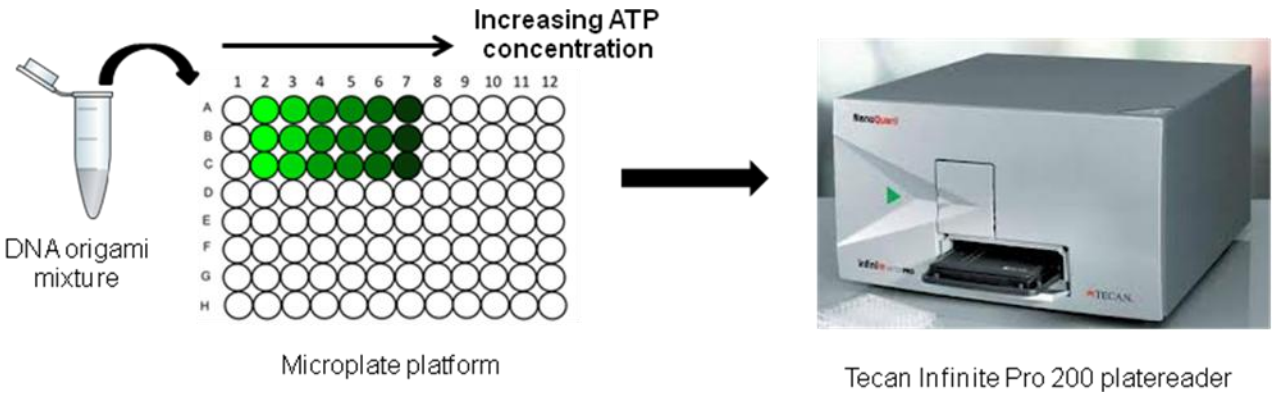


Figure 49: Schematic representation of the ATP titration experiments conducted on a microplate reader fluorimeter.

3.7.6. Dissociation constant calculation

The dissociation constant was achieved using Igor software, applying the following one-to-one binding fitting equation:

```
Function Kd (w,x) : FitFunc
  Wave w
  Variable x
```

$$f(x) = v\_0 + v\_2*(((v\_3+x+v\_1)-((v\_3+x+v\_1)^{\wedge 2}-4*v\_3*x)^{\wedge 0.5})/(2*v\_3))$$

where v_0 is the initial point at 0μM ATP addition (usually considered 1 in the I/I_0 plotting); v_2 corresponds to the expected K_D of the complex; v_3 is the intensity change during the titration and v_4 represents the substrate concentration. The results in bulk conditions are consistent with the results previously obtained.

3.7.7. Kinetics measurement

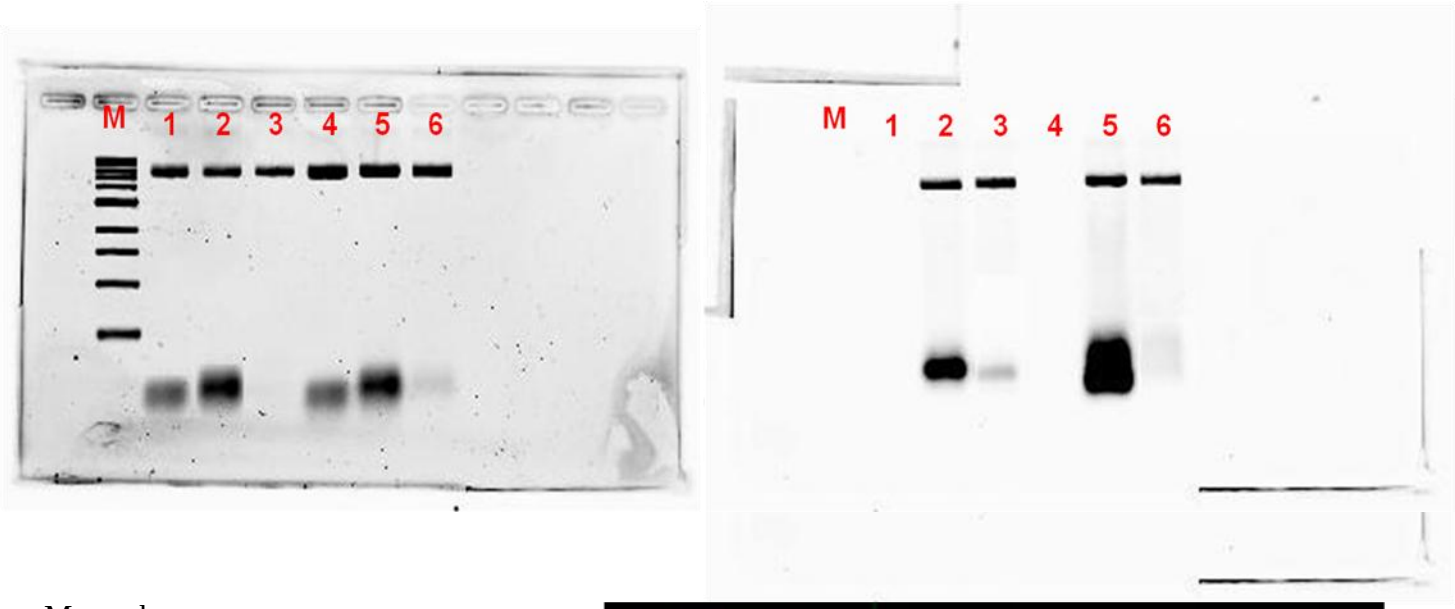
Experimental conditions:

- Instrument: Tecan Infinite pro 200;
- cRNP concentration: 60nM;
- DNA origami concentration: 10nM;
- Gain 80;
- Sample volume 100μl;
- Room temperature;
- Incubation time at room temperature: 30 minutes before adding ATP;
- Shaking: 10sec, amplitude 2mm (orbital);
- Fluorescent intensity measured every 30sec;
- Exc. Wv.: 470nm;
- Emission Wv: 525nm (λ_{MAX});

As for the K_D measurements, an ATP titration was performed on Tecan microplate reader. In order to compare bulk conditions with 12, 18 and 4 binding sites, sample concentration was fixed at 60nM cRNP (corresponding to the full binding of the 6 binding sites DNA origami). Sample concentration was evaluated from gel image volume analyses (**Figure 50**). After ATP addition fluorescence changes measurements were immediately carried out (every 30 second for 1hour) and successively plotted

against ATP concentration. The ATP range scanned is 0 µM; 0.1 µM; 0.5 µM; 1 µM; 5 µM; 10 µM; 25 µM (An16tag04) and 0 µM; 10 µM; 25 µM; 50 µM; 100 µM; 150 µM; 200 µM; 300 µM; 500 µM (A26tag01).

An16tag04 18 b.s. A26tag01 18 b.s.



M: marker;

1, 4: DNA origami without cRNP;

2,5: DNA origami with cRNP

Before purification;

3,6: DNA origami with cRNP after purification.

Yield: An16: 81%; A26: 87%

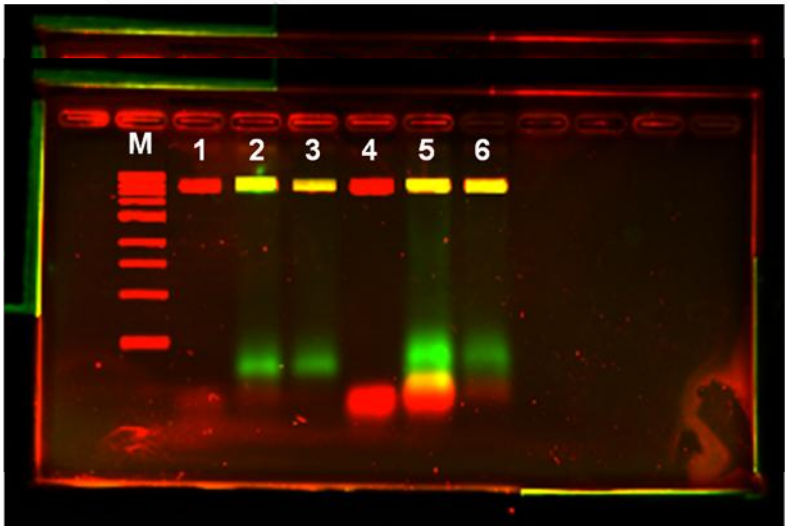


Figure 50: Gel image analyses of An16tag04 and A26tag01 on DNA origami 18 binding sites. From the top: EtBr staining (left side); fluorescence image (right side) and overlaid image (on the bottom).

3.7.8. References

1) Rothemund, P. W. K., *Nature*, 2006, 440, 297-302.
2) Seeman, N. C., *Mol. Biotechnol.*, **2003,** 37, 246-257.
3) Nakata, E., Liew, F.F., Uwatoko, C., Kiyonaka, S., Mori, Y., Katsuda, Y., Endo, M., Sugiyama, H., Morii, T., *Angew. Chem. Int. Ed*, 2012, 51, 2421-2424.
4) Hagihara, M., Fukuda, M., Hasegawa, T., and Morii, T., *J. Am. Chem. Soc.*, **2006**, *128*, 12932-12940.
5) Nakano, S., Fukuda, M., Tamura, T., Sakaguchi, R., Nakata, E., Morii, T., *J. Am. Chem. Soc*., **2013**, *135*, 3465-3473.
6) Koshland, D. E., Goldbeter, A., Stock, J. B., *Science*, **1982**, 217, 220-225.
7) Ferrel, J. E, *Trends in Biochemical Sciences*, **1996**, 460-466.
8) Bhattacharva, S., Bunick, C. G., Chazin, W. J., **2004,** *1742*, 69-79.
9) Vallée-Bélisle, A., Ricci, F., Plaxco, K. W., *J. Am. Chem. Soc.*, **2012**, 134, 2876-2879.
10) Binnig, G.; Quate, C. F.; Gerber, C., *Phys. Rev. Lett.*, **1986**, 56, 930.
11) Butt, H. J.; Downing, K. H.; Hansma, P. K. Biophys. J. 1990, 58, 1473.
12) Rajendran, A.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. *ACS Nano*, **2011**, 5, 665.
13) Zhong, Q.; Inniss, D.; Kjoller, K.; Elings, V. B., *Surf. Sci. Lett*., **1993**, 290, L688.
14) Marti, O.; Drake, B.; Hansma, P. K., *Appl. Phys. Lett*., **1987**, 51, 484.
15) Hansma, P. K.; Schitter, G.; Fantner, G. E.; Prater, C., *Science*, **2006**, 314, 601.
16) Viani, M. B.; Schaffer, T. E.; Paloczi, G. T.; Pietrasanta, L. I.; Smith, B. L.; Thompson, J. B.; Richter, M.; Rief, M.; Gaub, H. E.; Plaxco, K. W.; Cleland, A. N.; Hansma, H. G.; Hansma, P. K. *Rev. Sci. Instrum*., **1999**, 70, 4300.
17) Viani, M. B.; Pietrasanta, L. I.; Thompson, J. B.; Chand, A.; Gebeshuber, I. C.; Kindt, J. H.; Richter, M.; Hansma, H. G.; Hansma, P. K., *Nat. Struct. Biol.*, **2000**, 7, 644.
18) Picco, L. M.; Bozec, L.; Ulcinas, A.; Engledew, D. J.; Antognozzi, M.; Horton, M. A.; Miles, M. J., *Nanotechnology*, **2007**, 18, 44030.
19) Fantner, G. E.; Schitter, G.; Kindt, J. H.; Ivanov, T.; Ivanova, K.; Patel, R.; Holten-Andersen, N.; Adams, J.; Thurner, P. J.; Rangelow, I. W.; Hansma, P. K., *Ultramicroscopy*, **2006**, 106, 881.
20) Kindt, J. H.; Fantner, G. E.; Cutroni, J. A.; Hansma, P. K., *Ultramicroscopy*, **2004**, 100, 259.
21) Viani, M. B.; Schaffer, T. E.; Chand, A.; Rief, M.; Gaub, H. E.; Hansma, P. K., *J. Appl. Phys*., **1999**, 86, 2258.

22) Manalis, S. R.; Minne, S. C.; Quate, C. F., *Appl. Phys. Lett*., **1996**, 68, 871.

23) Lutwyche, M.; Andreoli, C.; Binnig, G.; Brugger, J.; Drechsler, U.; Häberle, W.; Rohrer, H.; Rothuizen, H.; Vettiger, P.; Yaralioglu, G.; Quate, C., *Sens. Actuators A: Phys*., **1999**, 73, 89.

24) Walters, D. A.; Cleveland, J. P.; Thomson, N. H.; Hansma, P. K.; Wendman, M. A.; Gurley, G.; Elings, V., *Rev. Sci. Instrum*., **1996**, 67, 3583.

25) Schitter, G.; Astrom, K. J.; DeMartini, B. E.; Thurner, P. J.; Turner, K. L.; Hansma, P. K., *Control Syst. Technol., IEEE Trans*., **2007**, 15, 906.

26) Ando, T.; Kodera, N.; Maruyama, D.; Takai, E.; Saito, K.; Toda, *A. Jpn. J. Appl. Phys*., **2002**, 41, 4851.

27) Ando, T.; Kodera, N.; Takai, E.; Maruyama, D.; Saito, K.; Toda, *A. Proc. Natl. Acad. Sci. U.S.A*., **2001**, 98, 12468.

28) Ando, T.; Kodera, N.; Uchihashi, T.; Miyagi, A.; Nakakita, R.; Yamashita, H.; Matada, K., e-*J. Surf. Sci. Nanotechnol*., **2005**, 3, 384.

29) Kodera, N.; Yamamoto, D.; Ishikawa, R.; Ando, T., *Nature*, **2010**, 468, 72.

30) Uchihashi, T.; Iino, R.; Ando, T.; Noji, H., *Science*, **2011**, 333, 755.