



Doctoral Thesis

Probing dynamics in RNA-protein recognition - Structural and functional investigation of CUG-BP2 protein

Author(s):

Diarra Dit Konté, Nana

Publication Date:

2016

Permanent Link:

<https://doi.org/10.3929/ethz-a-010749930> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 23497

Probing dynamics in RNA-protein recognition
Structural and functional investigation of CUG-BP2 protein

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

NANA DIARRA DIT KONTÉ

MSc, University Pierre et Marie Curie, Paris

born on 06.07.1985

citizen of France and Mali

accepted on the recommendation of

Prof. Dr. Frédéric Allain
Prof. Dr. Andres Ramos
Prof. Dr. Gerhard Wider

2016

Abstract

RNA is the pivotal molecule in the central molecular biology dogma that acts as intermediary between the DNA that carries the hereditary information and the proteins. Therefore, messenger RNA processing is an important landmark in gene regulation. A plethora of RNA binding proteins control and modulate its different steps as component of essential ribonucleoparticles such as the spliceosome or as accessory factors. Thus, understanding the determinants of protein-RNA interactions is a prerequisite to gain insight in posttranscriptional gene regulation.

In the present work, we investigate the CUG binding protein 2 (CUG-BP2) that is involved in many aspects of RNA metabolism. It is an alternative splicing factor; it can regulate mRNA stability and translation but also modulates mRNA editing. CUG-BP2 interacts with multiple targets. It was initially identified as a CUG triplet binding protein but systematic evolution of ligands by exponential enrichment (SELEX) experiments showed that it prefers UG repeats and notably the UGUU motif. Moreover, its interaction with AU-rich sequences has been demonstrated to be essential in mRNA editing, stability and translation regulation.

CUG-BP2 belongs to the CUG-BP and ELAV like factor (CELF) family. The members have two RNA recognition motifs (RRM) separated from a third one by a divergent domain. The RRMs are extremely conserved among the family with over 90% identity between CUG-BP2 and CUG-BP1, its closest homolog. The solution structure of CUG-BP1 RRM3 in complex with (UG)₃ has been solved by NMR spectroscopy and the structures of CUG-BP1 RRM1 and RRM2 bound to RNAs containing the UGUU motif by X-ray crystallography. Although these structures provide very detailed insight in the binding mode to UG rich RNAs, it remains unclear how CUG-BP2 recognizes AU-rich RNAs.

The aim of this study is to characterize structurally and dynamically the interaction of the CUG-BP2 with short AU-rich RNAs by NMR spectroscopy and to understand how the protein discriminates between different targets. We also focus on gaining insight into the regulation and stabilization of mRNAs by CUG-BP2 *in vivo*.

In an attempt to facilitate the NMR investigation, we established simple procedures to obtain short labeled RNA that could prove useful for obtaining intermolecular restraints between the protein and the RNA target. We then used a modular approach to characterize the interaction of CUG-BP2 with AU-rich sequences originating from the 3' UTR of the COX-2 mRNA, and studied the N-terminal

tandem RRM12 and the C-terminal RRM3 separately. We obtained a preliminary structure of the RRM12 bound to 5'-AUUUAAUU-3' and established that both RRMs could interact with the RNA simultaneously and had correlated motion but the RNA was bound in multiple register, which was detrimental for the complex structure determination by NMR.

However, we solved the structure of the CUG-BP2 RRM3 in complex with 5'-UUUAA-3'. By comparison with the structure of CUG-BP1 RRM3 in complex with 5'-UGUGUG-3', we could show that RRM3 used the same binding interface and protein moieties to interact with different targets but that the conformation of the aromatic residues at the binding interface control the affinity for the cognate RNA. Using scalar coupling experiments and molecular dynamics, we could establish that RRM3 used aromatic side chain rearrangement as an original mean to discriminate different RNA targets.

We also investigated the role of the different domain of CUG-BP2 in the regulation of COX-2 mRNA translation and performed in cell luciferase gene reported assays. We could establish that interaction with the RNA was essential for the function and that the divergent domain helped amplifying the translation inhibition. We also demonstrated that CUG-BP2 was a cargo of the Karyopherin β 2 and that the nuclear localization signal at the C-terminus was a PY-NLS.

Résumé

L'ARN agit comme intermédiaire entre l'ADN qui porte l'information génétique et les protéines ce qui en fait la molécule clé du dogme central de la biologie moléculaire. Par conséquent, la maturation de l'ARN messenger est un jalon important dans la régulation de l'expression des gènes dont les différentes étapes peuvent être contrôlées et modulées par une pléthore de protéines de liaison à l'ARN. Ces dernières agissent en tant que composantes de particules ribonucléiques essentielles telles que le spliceosome ou comme facteurs accessoires. Ainsi, une meilleure appréhension des interactions protéines-ARN est indispensable pour comprendre la régulation post-transcriptionnelle des gènes.

Dans cette thèse, nous nous proposons d'étudier la protéine CUG-BP2 qui est impliquée dans de nombreux aspects du métabolisme de l'ARN tel que l'épissage alternatif, la stabilisation et la régulation de la traduction ou encore l'édition. CUG-BP2 interagit avec plusieurs cibles. Elle a été initialement identifiée comme une protéine de liaison aux triplets CUG mais il a été démontré par évolution systématique de ligands par enrichissement exponentiel que la protéine a une préférence pour les répétitions UG et notamment le motif UGUU. De plus, il a été établi que par son interaction avec des séquences riches en UA, CUG-BP2 régule l'édition, la stabilité et la traduction d'ARNm.

CUG-BP2 appartient à la famille CUG-BP and ELAV like factor (CELF) dont les membres ont trois motifs de reconnaissance de l'ARN (RRM), le deuxième et le troisième étant séparés par un domaine divergent. Les RRM sont extrêmement conservés parmi les membres de la famille avec plus de 90% d'identité de séquence entre CUG-BP2 et CUG-BP1, son plus proche homologue. La structure du RRM3 de CUG-BP1 en complexe avec (UG)₃ a été résolue par spectroscopie RMN et les structures des RRM1 et RRM2 de CUG-BP1 liés à des ARNs contenant le motif UGUU par cristallographie aux rayons X. Bien que ces structures donnent un aperçu très détaillé du mode de liaison aux ARN riches en UG, on ne sait pas comment CUG-BP2 reconnaît les ARN riches en UA.

Le but de cette étude est de caractériser structurellement et dynamiquement l'interaction de CUG-BP2 avec de courts ARNs riches en UA par spectroscopie RMN afin de comprendre comment la protéine distingue différentes séquences d'ARN. Nous nous intéressons également à la régulation et la stabilisation des ARNm par CUG-BP2 *in vivo*. Dans un premier temps, nous avons établi des procédés simples pour obtenir de courts ARNs marqués en vue d'obtenir des contraintes intermoléculaires entre la protéine et l'ARN cible et de faciliter les investigations par RMN. Nous avons ensuite caractérisé

séparément l'interaction des RRM12 et du RRM3 avec des séquences riches en AU provenant de la région 3' non-transcrite de l'ARNm de la cyclooxygénase 2 (COX-2).

Nous avons établi que les RRM1 et 2 peuvent interagir simultanément avec l'ARN et obtenu une structure préliminaire du RRM12 lié à 5'-AUUUAUU-3'. Cependant, les RRM1 et RRM2 ont de multiples surfaces d'interaction avec 5'-AUUUAUU-3', ce qui exclut toute détermination de la structure du complexe par RMN.

En revanche, nous avons résolu la structure du RRM3 de CUG-BP2 en complexe avec UUUAA 5'-3'. En la comparant avec la structure du complexe du RRM3 de CUG-BP1 avec 5'-UGUGUG-3', nous avons pu montrer que le RRM3 utilise la même surface d'interaction et les mêmes chaînes latérales pour interagir avec différents ARNs, mais la conformation des résidus aromatiques au niveau de la surface d'interaction varie en fonction de l'affinité pour l'ARN. En utilisant des expériences de couplage scalaire et de dynamique moléculaire, nous avons pu établir que le RRM3 utilise la réorganisation des chaînes latérales aromatiques pour distinguer les différents ARN cibles.

Nous avons également étudié le rôle des différents domaines de CUG-BP2 dans la régulation de la traduction de l'ARNm de COX-2. Nous avons pu établir que l'interaction avec l'ARN des RRM12 et RRM3 était essentielle et que le domaine divergent aidait à amplifier l'inhibition de la traduction. L'extrémité C-terminale quant à elle contient un PY-SLN ; un signal de localisation nucléaire qui est responsable du transport de CUG-BP2 par la karyophérine β 2.