

## Supporting Information

# Cognate RNA-Binding Modes by the Alternative-Splicing Regulator MBNL1 Inferred from Molecular Dynamics

Àlex L. González, Daniel Fernández-Remacha, José Ignacio Borrell, Jordi Teixidó, Roger Estrada-Tejedor\*

Grup de Química Farmacèutica (GQF), IQS School of Engineering, Universitat Ramon Llull, Barcelona, Spain

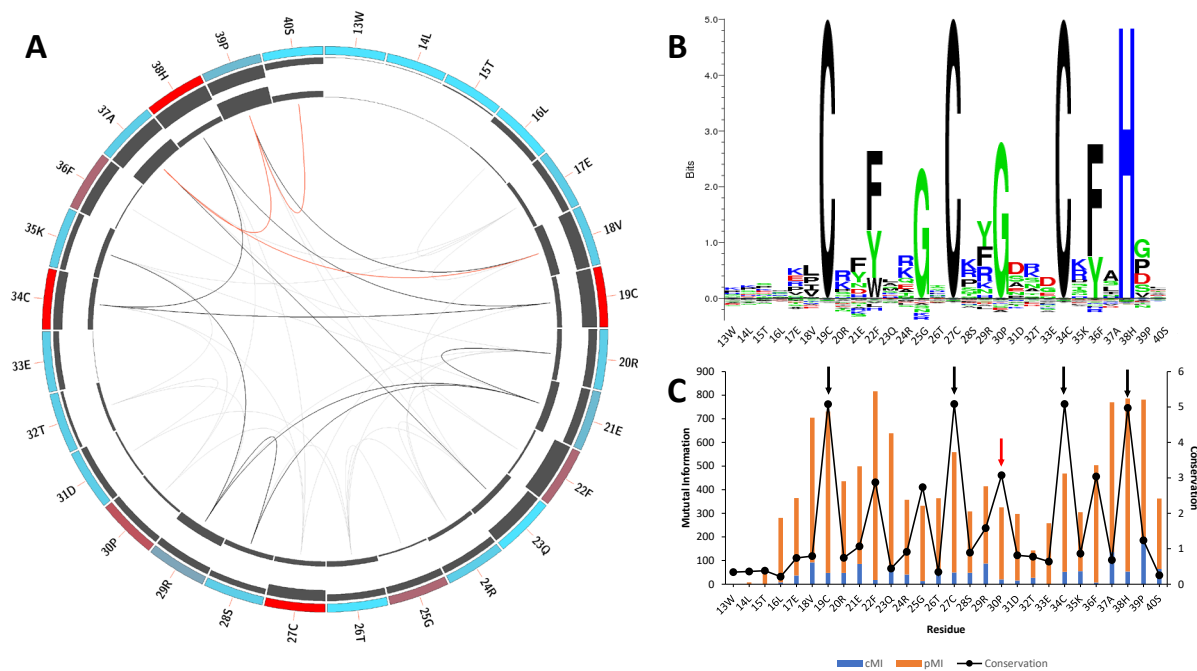
\* Correspondence: roger.estrada@iqs.url.edu

## Contents

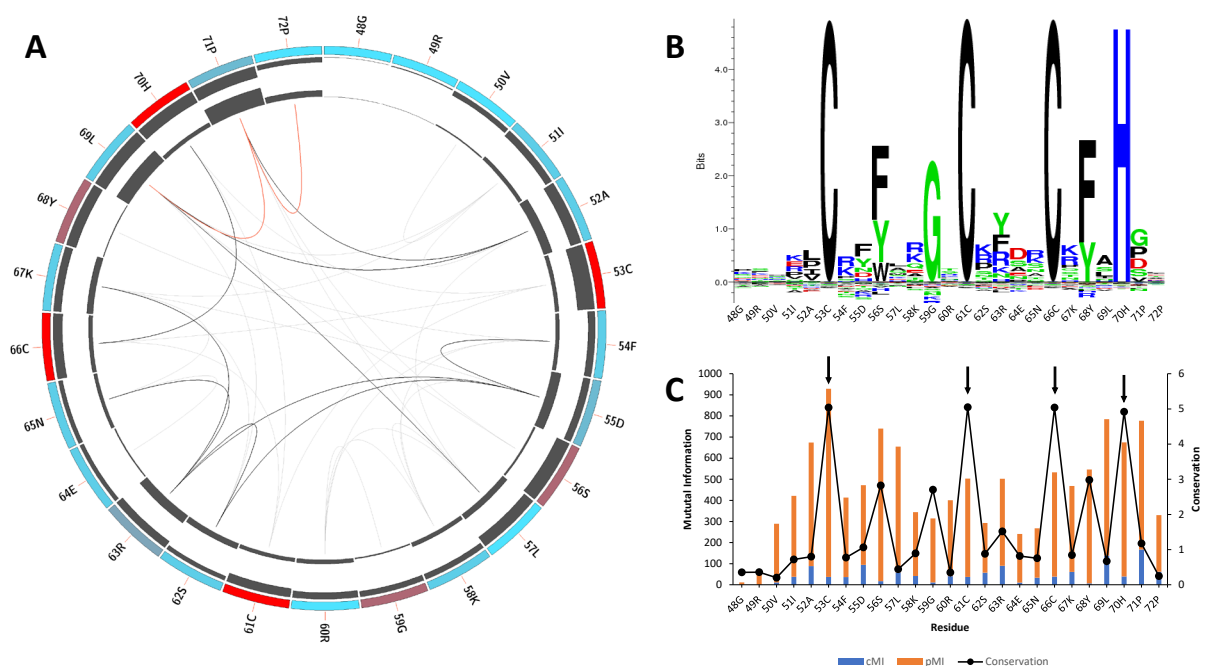
<b>Table S1.</b> Pearson correlation coefficient between RNA-bound and unbound local fluctuations for each ZnF domain. ....	2
<b>Figure S1.</b> Mutual information and conservation analysis of the CCCH domain in MBNL1 ZnF1 protein family. ....	3
<b>Figure S2.</b> Mutual information and conservation analysis of the CCCH domain in MBNL1 ZnF2 protein family. ....	4
<b>Figure S3.</b> Mutual information and conservation analysis of the CCCH domain in MBNL1 ZnF4 protein family. ....	5
<b>Figure S4.</b> Volume of the pocket computed during the sMD process for each ZnF .....	6

**Table S1.** Pearson correlation coefficient between RNA-bound and unbound local fluctuations for each ZnF domain. Tandem partner refers to each ZnF-partner into the same tandem (ZnF1 partners ZnF2, and ZnF3 partners ZnF4).

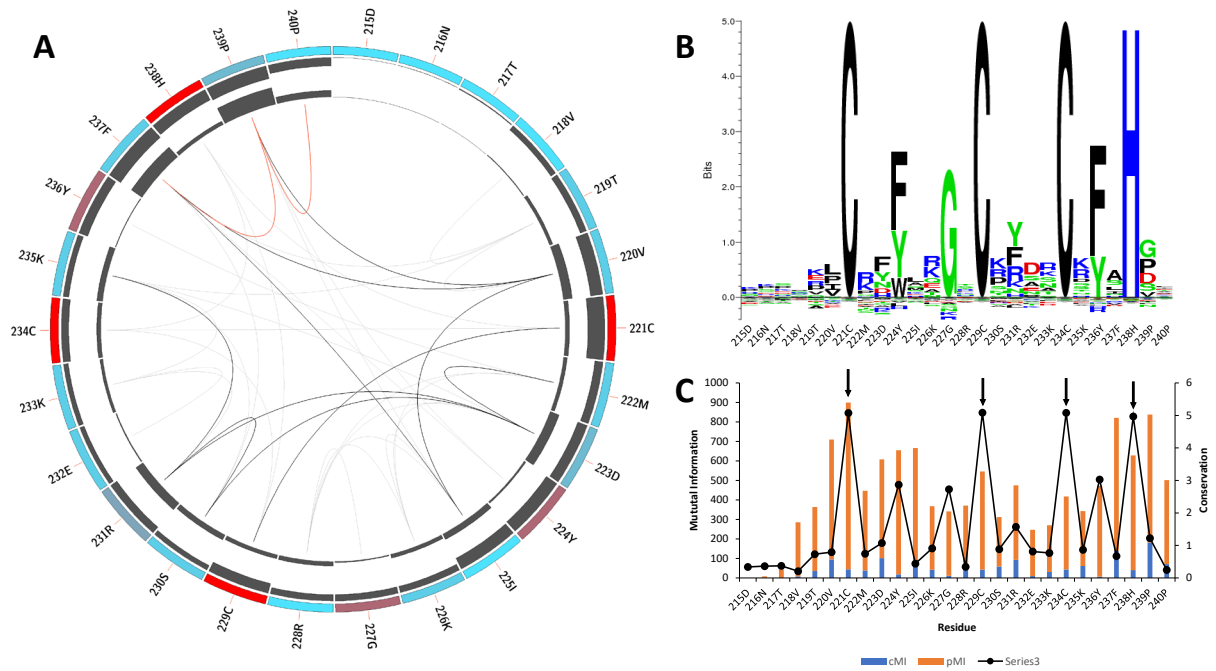
	<b>ZnF1</b>	<b>ZnF2</b>	<b>ZnF3</b>	<b>ZnF4</b>
<b>ZnF1 - RNA</b>	0.01	0.05	<b>0.31</b>	0.12
<b>ZnF2 - RNA</b>	0.06	0.03	0.21	-0.06
<b>ZnF3 - RNA</b>	0.12	0.04	<b>0.36</b>	-0.09
<b>ZnF4 - RNA</b>	0.13	-0.04	0.08	<b>0.76</b>
<b>Tandem partner - RNA</b>	0.05	0.11	0.06	<b>0.87</b>



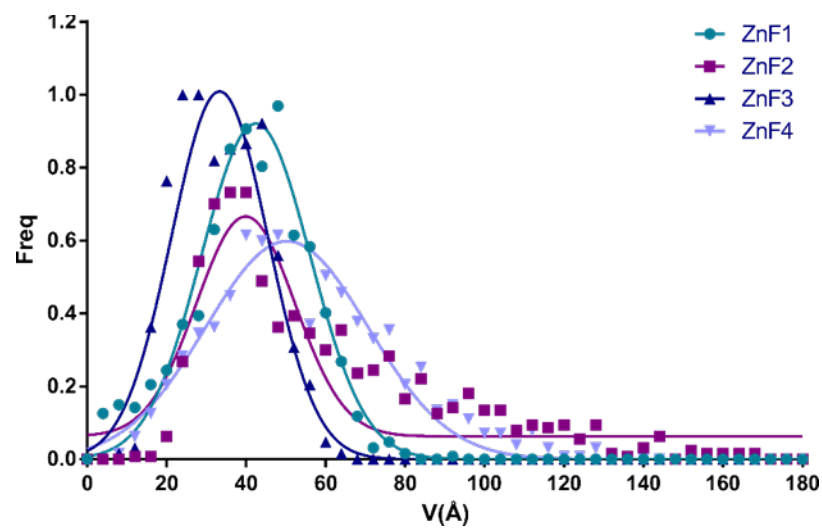
**Figure S1. Mutual information and conservation analysis of the CCCH domain in MBNL1 ZnF1 protein family.** (A) Circos representation of the CCCH domain family in ZnF1. Square boxes indicate the KL conservation score (from red-highest to cyan-lowest values). The cumulative mutual information (cMI) and proximity mutual information (pMI) scores are represented as histograms in the inner circle. Lines in the center connect pairs with a MI score higher than 6.5. Red lines represent the top 5%, black lines are between 95% and 70%, and gray lines indicate < 70%. (B) KL sequence logo of selected nodes. The second conserved Gly of the (F/Y)GG(F/Y) is noticeable for the alignment of ZnF1 Pro30 in the Multiple Sequence Alignment (MSA) from Pfam. Nevertheless, ZnF1 presents the (F/Y)G(F/Y) motif. (C) Histogram representing the conservation per residue, cMI and pMI within 5 Å threshold. Black arrows represent the CCCH motif and Pro30 is indicated in red. The sequence id numbers correspond to those in the circos representation.



**Figure S2. Mutual information and conservation analysis of the CCCH domain in MBNL1 ZnF2 protein family.** (A) Circos representation of the CCCH domain family in ZnF2. Square boxes indicate the KL conservation score (from red-highest to cyan-lowest values). The cumulative mutual information (cMI) and proximity mutual information (pMI) scores are represented as histograms in the inner circle. Lines in the center connect pairs with a MI score higher than 6.5. Red lines represent the top 5%, black lines are between 95% and 70%, and gray lines indicate < 70%. (B) KL sequence logo of selected nodes. (C) Histogram representing the conservation per residue, cMI and pMI within 5 Å threshold. Black arrows represent the CCCH motif. The sequence id numbers correspond to those in the circos representation.



**Figure S3. Mutual information and conservation analysis of the CCCH domain in MBNL1 ZnF4 protein family.** (A) Circos representation of the CCCH domain family in ZnF4. Square boxes indicate the KL conservation score (from red-highest to cyan-lowest values). The cumulative mutual information (cMI) and proximity mutual information (pMI) scores are represented as histograms in the inner circle. Lines in the center connect pairs with a MI score higher than 6.5. Red lines represent the top 5%, black lines are between 95% and 70%, and gray lines indicate < 70%. (B) KL sequence logo of selected nodes. (C) Histogram representing the conservation per residue, cMI and pMI within 5 Å threshold. Black arrows represent the CCCH motif. The sequence id numbers correspond to those in the circos representation.



**Figure S4.** Volume of the pocket computed during the sMD process for each ZnF. Data has been smoothened using Gaussian distributions.