

行政院國家科學委員會專題研究計畫 成果報告

探討 MBNL 蛋白質調控選擇性剪接的分子機轉 研究成果報告(精簡版)

計畫類別：個別型
計畫編號：NSC 96-2311-B-040-002-
執行期間：96年08月01日至97年07月31日
執行單位：中山醫學大學生物醫學科學學系

計畫主持人：劉玉凡
共同主持人：蕭光明
計畫參與人員：碩士級-專任助理人員：張庭諺
碩士班研究生-兼任助理人員：蔡德峰

處理方式：本計畫涉及專利或其他智慧財產權，2年後可公開查詢

中華民國 97年10月30日

行政院國家科學委員會補助專題研究計畫

成果報告 期中進度報告

探討 MBNL 蛋白質控制選擇性剪接的分子機制

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 96-2311-B-040-002-

執行期間：96年 08月 01日至 97年 07月 31日

計畫主持人：劉玉凡 助理教授

共同主持人：

計畫參與人員：研究助理 張庭諺，研究生 蔡德峰

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：中山醫學大學生物醫學系

中華民國 97 年 10 月 30 日

簡介

Muscleblind-like(MBNL)基因是屬於一群位於細胞核內 RNA-binding 的蛋白質家族，目前所知其主要的功能是扮演調控 pre-mRNA 選擇性剪輯的角色(1)，雖然對於該基因家族的分析，是擁有不同的蛋白質模組，但彼此間一定共同擁有一個高度保留性的 CCCH 鋅指 (Zn finger) 的蛋白質模組，這個蛋白質模組也是 MBNL 主要與特定 RNA 結構結合的部位(2)，透過蛋白質與 RNA 的交互作用，對於相關基因在發育的過程中，作為基因選擇性剪輯調控的機制。

最早的 muscleblind (mbl) 基因的報導是在線蟲中被發現的，該基因共有四種不同的 isoforms (mbl A-D) 被報導出來(3)，其中在線蟲中 mbl 的基因，從胚胎發育的中後期就開始有表現，同時在內臟 (visceral)、體肌肉(somatic muscles)、腹神經索 (ventral nerve cord) 以及幼蟲感光受體 (larval photoreceptors) 皆有表現 mbl 基因(4)。另一方面 mbl 基因的突變，會導致因感光受體分化的缺陷，以及 Z-band 形成導致間接肌肉性的附著不正常而有致死的現象，這些實驗結果都在說明著，於線蟲的發育過程中，mbl 基因扮演的角色，是在視覺受體及肌肉細胞分化終止的調控因子。

相較於線蟲僅有單一個 mbl 基因被發現，在較高等的哺乳動物如人類及小鼠中已有三個 muscleblind-like (MBNL1, MBNL2 and MBNL3) 基因被報導(5,6)，然而在人類的基因表現分析的結果，其中 MBNL1 基因在心臟以及骨骼細胞表現最多，而 MBNL2 基因則是在大部分的組織中均有表現(5)，而 MBNL3 的表現則是相對低量，並且在骨骼細胞無法偵測到該基因的表現，對於 mbl 基因在小鼠中表現的模式，也跟在人類的發現類似(6)。目前的資料顯示，MBNL1 基因主要透過調控選擇性剪輯，參與促進肌肉分化的角色，

而 MBNL3 則是相反有抑制肌肉分化的作用(7)，而另一方面，MBNL2 的功能報導，透過與 RNA-dependent integrin $\alpha 3$ 等細胞結構及骨幹蛋白質等 mRNA 的交互作用，將該基因被運送到需要的表現位置有關(8)。

除了調控相關組織的分化作用外，MBNL 基因也報導與肌強直症(myotonic dystrophy, DM1, MIN 160900) 的致病機轉有關。肌強直症主要發病的原因，是由位於 myotonic dystrophy protein kinase (DMPK) 基因 3'UTR 區域的重複序列 (CTG) n 異常擴增的結果 (DM1)，或是位於 ZNF9 gene 的 intronic 區域的重複序列 (CCTG) n 異常擴增的結果(DM2, MIN 602668) (9)。帶有異常擴增的重複序列的 mRNAs 會停留在細胞核中，並進一步形成 RNA 的 foci，而在 DM1 及 DM2 病人的細胞核中三個 MBNLs 的蛋白質基因，都會被吸引至 RNA foci 的位置，因此被隔離的 MBNLs 基因，會造成其下游所調控選擇性剪輯的基因異常，以及無法運送特定 mRNAs 至適當地表現位置，而最後導致多系統性 (multi-systemic) 肌強直症的臨床病徵(9,10)。同時 mbnl 剔除的小鼠也呈相類似的臨床病徵，以及共同錯誤剪輯的基因產物等分子機制的特徵(10)。

雖然過去幾年的研究工作，對於 muscleblind-like 基因的分子層次功能，以及相關疾病的致病機轉，已有很大的突破以及明瞭，但是卻少有該基因在脊椎動物發生過程中，所扮演功能的相關報導。斑馬魚是很好研究胚胎時期以及發生過程的脊椎模型生物，同時至今對於斑馬魚 mbnl 的相關基因也未有相關的研究工作，因此本計劃利用生物資訊學以及分子生物學的方法，對於斑馬魚的三個 mbnl 基因的性質、蛋白質結構、基因剪輯的形式和表現的模式做進一步的研究與分析(11)。

結果

■ Identification of zebrafish mbnl genes

首先利用人類的三個 MBNLs 同源基因，以 TBLASTN 的方式對於斑馬魚的基因體，進行相似性比對的分析，從搜尋比對的過程中，總共發現四個推定(putative) mbnl 特定序列的模式，分別位於斑馬魚基因體 scaffolds 4, 157, 191 and 1,388 的組合序列上。並且也從 NCBI UniGene 轉譯基因資料庫中，也比對到 3 個基因叢集 (clusters)，分別是 Dr.141606 (scaffold 4 位於 Chr.1)、Dr.89831 (scaffold 157 and 191 位於 Chr.2，可組成更大的 scaffold) 以及 Dr.85253 (scaffold 1388 位於 Chr.14)。根據生物資訊分析的結果，設計適當的引子(primers)，選殖並完成三個斑馬魚的 mbnl 基因的定序工作 [Fig.1]。接下來透過人類、小鼠、爪蟾、斑馬魚、果蠅及線蟲等六個模型生物的 14 個 MBNL 同源基因，以及基因親源樹 (phylogenetic analysis) 分析的結果顯示，上述斑馬魚的三個 mbnl 基因分別屬於 mbnl2, mbnl1 以及 mbnl3 [Fig.2]。同時在三個斑馬魚 mbnls 基因，透過不同剪輯的方式共有 13 個轉錄基因被發現 [Fig.3]。目前相關的序列資訊，也已完成交存至基因銀行(GenBank) 資料庫中。

對於 mbnl 基因在基因體中的數量的分析，在無脊椎生物中僅有 1 個 copy (1)，相較於脊椎生物包括斑馬魚在內，本計劃的研究結果也僅有 3 個 mbnl 基因。雖然最近在河豚的基因體中有雙倍 (duplication) 複製的現象，在河豚基因體中有 5 個 mbnl 基因被報導，其中 mbnl3 有 1 個 copy，而 mbnl1 及 mbnl2 則有 2 個 copy 的發現(12)，說明有在物種魚類的演化過程中，在斑馬魚以及河豚基因體中發生了不同的結果，極有可能在斑馬魚及河豚共同的祖先演化之後才發生基因雙倍的過程。

■ Genomic organization and structural features of zebrafish mbnl genes

對於斑馬魚上的三個 *zmbnl1*, *zmbnl2* 及 *zmbnl3* 分別於基因體橫跨的範圍，依次為 87.6, 75.3 及 65.0 kb，依次分別由 10, 8 及 9 exons 所組成的基因 [Fig.3]。根據分析的結果 *zmbnls* 和人類 MBNL 基因在基因體上的結構組成大致類似，共同的特徵有：每個 *mbnl* 基因的第一個 intron 都很大 (約佔 30-60 kb 左右)，第二個 intron 通常很小，其次第二、三個 exons 及倒數第二個 exon 的大小以及剪輯點的位置，在 *zmbnl2* 及 *zmbnl3* 兩個基因上完全相同。對於三個 *zmbnl* 基因而言，第一及二對的 CCCH 鋅指蛋白質模組分別位於第一、二個及第四個 exon 上 [Fig.3]。對於斑馬魚的 3 個 *zmbnl* 基因彼此間，不但擁有相類似的基因結構及剪輯形式外，同時更與現在已知的脊椎生物也相類似的結果(1)。再進一步透過斑馬魚與人類基因體的比對，這三個基因附近的同源基因以及其組成也相同。

透過跨物種的比對 *mbnl* 基因上的序列，所有的 *mbnl* 的蛋白質都有高度保留性的 CCCH 鋅指蛋白質模組 [Fig.1]。兩個成對出現的鋅指蛋白質模組，是以固定的距離出現 Cys 及 His 兩種氨基酸，其出現的模式分別為 $CX_7CX_6CX_3H$ 及 $CX_7CX_4CX_3H$ ，在脊椎生物中所有的 *mbnl* 基因都有相同鋅指蛋白質模組。除了鋅指蛋白質模組外，斑馬魚的 *mbnl* 基因還有兩個高度相似的蛋白質序列，如先前文章所報導的 LEV 及 NGR boxes，其中 LEV box 出現的位置是緊接在第一、三個鋅指蛋白質模組之前 (1)，同時 NGR boxes 則是出現在第二個鋅指蛋白質模組附近 [Fig.1]。在緊接於鋅指蛋白質模組之後的是 alanine-rich 的區域，但這段序列只出現在所有的 *mbnl1* 及 *mbnl2* 基因，但在 *mbnl3* 及無脊椎生物的同源基因卻沒有出現 [Fig.2]。特別值得一提的斑馬魚的 *zmbnl1* 基因，在 2 個成對的鋅指蛋白質模組間，有連續兩個 poly alanine 的區域，這個序列的特徵，目前僅有在

斑馬魚有被報導出來。雖然目前對於 LEV/NGR 及 alanine-rich 區域，高度保留性的序列功能還不太清楚，但是也有一些相關文章對於 alanine-rich 序列功能有些報導，其中 AUF1 也是一種與 RNA 結合的基因，該蛋白質上的 alanine-rich 區域的序列，與雙體 (dimerization) 的形成有關，並且可以加強與目標 RNA 接合的親和力 (13)。另一篇文章，黏菌(Dictyostelium) 的 myosin II 也有兩個專一性的 alanine-rich 的區域，會相互結合形成四倍體的 coiled-coil 結構，這些研究工作可以作為繼續探討 mbnl 基因(14)，透過這些高度保留性的蛋白質模組的交互作用，以及它們對於選擇性剪輯調控的影響。

在 mbnl 基因上，除了上述幾個蛋白質模組被發現外，本計劃亦首度發現一個位於蛋白質 C-端的新穎蛋白質模組，這是一個 Ser/Thr-rich 的區域，整個蛋白質模組長度約為 25 個氨基酸 [Fig.1]，這個區域的序列僅在脊椎生物中的 mbnl 基因具有高度的保留性，但在無脊椎生物中則沒有發現。進一步分析發現該區域的親、疏水性的氨基酸組成，發現間隔固定的氨基酸頻率出現 [Fig.4A]。這個序列的特徵，與前面文獻對於 ALPS 蛋白質模組的報導相類似，該蛋白質模組透過其親水性/疏水性的雙性 α -helix 特性結構，可以使得細胞膜彎曲，與細胞內物質運輸功能有關(15)。雖然在這個區域中發現，會有傾向於破壞蛋白質 α -helix 的結構 proline 胺基酸出現，但是透過生物資訊學工具 - helical wheel 的分析 [Fig.4B]，進一步證明這個區域的序列，具有類似於 ALPS 蛋白質模組的特徵。當然，這個區域的序列是否具有與細胞膜結合的能力?或是受到膜的誘導下會形成雙性 α -helix 結構(16)? 仍需要進一步的實驗證明。但整體而言，這些保留性的蛋白質模組，不但顯示了 mbnl 相關的基因是從同一個始

祖基因所演化出來，同時在物種的演化過程中，在既有的功能上，更衍生出更多與剪輯以及運輸的調控機制功能。

■ Alternative splicing of zebrafish mbnl primary transcripts generates extensive protein diversity

透過選擇性剪輯的機制 mbnl 基因本身可以產生蛋白質的多樣性，而這些延伸出來多樣性的蛋白質，本計劃利用分子生物學的方法選殖及定序，分別對於 zmbnl1, zmbnl2 及 zmbnl3 基因依次各有 4, 4 及 5 種不同的剪輯方式[Fig.3]。其中主要的 mbnl 亞型 (isoforms) 的蛋白質都幾乎都帶有兩個完整成對的鋅指蛋白質模組，除了 zmbn2A 及 zmbnl2D 兩個蛋白質亞型，因為第二、三、四個 exons 的選擇性剪輯的原因，造成其轉錄出來的蛋白質缺乏第 1 個成對的鋅指蛋白質模組。而且這個選擇性剪輯的結果，在人類的 MBNL 同源基因也是同樣的情形，也就是缺乏第 1 個成對的鋅指蛋白質模組，目前僅發現在 mbnl2 的基因上，而沒有發現在 mbnl1 及 mbnl3 基因上出現(17)。同時根據前面的文獻報導，利用 deletion 的分析方法及 yeast three-hybrid 的系統，顯示只要欠缺任何一個成對的鋅指蛋白質模組，該蛋白質就失去了與 CUG/CCUG RNA 重複序列結合的能力(17)。然而，也有文獻報導 TTP 基因僅帶有一個 CCCH 鋅指蛋白質模組，但也可以於特定的 RNA 序列結合。因此推論在斑馬魚中，缺乏第 1 個成對的鋅指蛋白質模組的亞型蛋白質，也可能扮演調控的角色，不是直接與具有 CUG/CCUG 重複序列的 RNA 結合，而是藉由其他的 RNA 序列(18)。

斑馬魚的 mbnl 亞型蛋白質，除了鋅指蛋白質模組的數量有不同的差異外，亞型間主要的差異發生在蛋白質 C-端的 exons，例如：zmbnl1 的第五、七個 exons 的選擇性剪輯，可以產生 4 種蛋白質亞型，這樣的剪輯模式在人類的 MBNL1 基因也有同樣的發現，更重要的是這些亞型的開放式架構 (ORF) 也都相同，這樣的現象也發生

在 zmbnl2 及 zmbnl3 兩個基因上，在斑馬魚幾乎所有 mbnl 基因的亞型都擁有 Ser/Thr-rich 的區域，但是除了 zmbnl3A 因為選擇性剪輯的因素，以及 zmbnl3D 因為第五個 exon 的延伸造成轉錄框架位移(frameshift) 外，沒有 Ser/Thr-rich 的區域外，顯示這個區域的對於 mbnl 基因的重要性(1)。

相較於脊椎生物 mbnl 基因的亞型多樣性特徵，無脊椎生物的同源基因 mbnl 基因就顯得簡單多了，在線蟲的基因體僅有 1 個 mbnl 基因帶有 2 個蛋白質亞型，在果蠅的基因體也僅有 1 個 mbl 基因，同時帶有 4 個蛋白質亞型(3)。因此隨著物種的演化，mbnl 基因的複雜度(complexity) 也隨著增加，mbnl 基因亞型的數量，也開始隨著細胞的分化，開始透過不同的剪輯模式來調整其功能，甚至這些亞型彼此間也可能透過某種拮抗(antagonistic) 或是回饋(feedback) 作用，來達到彼此調控的目的。

討論

整體而言，本計劃透過生物資訊學以及分子生物學的方法，藉由選殖及定序基因的方法，將斑馬魚的 3 個 mbnl 相關基因，以及其 13 種不同選擇性剪輯的蛋白質亞型給完成。雖然斑馬魚 mbnl 基因與現有脊椎生物的同源基因，在序列方面具有高度的保留性，但是本計畫不只完成基因選擇及定序的工作，同時也是第一個針對 mbnl 基因在胚胎發生初期的研究工作，以及不同的剪輯亞型在斑馬魚不同組織間的表現情形，雖然大部分的基因表現模式，與已經的哺乳類動物的表現模式相類似，但是並不完全是相同的，顯示斑馬魚的 mbnl 基因除了與現有已知的功能外，還可能胚胎時期扮演一些未知功能的角色，值得繼續研究與探討，並藉此研究工作的報導，希望藉由斑馬魚的 mbnl 基因的研究與分析，藉此對於脊椎生物早期發育的過程，能提供更多的資料。

計畫成果自評

完成預定本計畫第一年的進度，並且在計畫中所提的基因選殖，以及定序了斑馬魚 3 個 mbnl 基因，以及其 13 種不同選擇性剪輯的蛋白質亞型，並且進一步建立探討 MBNL 蛋白質調控選擇性剪接的不同胚胎及發育時期的表現情形。同時結合下一年度的初步結果，進一步發現 C-端的 Ser/Thr-rich 區域的蛋白質模組，透過生物資訊學的分析，初步決定該蛋白質模組，是透過其親水性/疏水性的雙性 α -helix 特性結構，是否這個區域的序列具有與細胞膜結合的能力？或是受到膜的誘導下會形成雙性 α -helix 結構？仍需要進一步延伸性的計畫，繼續支持以及相關實驗的證明。

初步將本計畫生物資訊部份的研究成果，與合作的實驗室聯名於今年初發表於相關的國際期刊上，*Comparative Biochemistry and Physiology, Part B* 151 (2008) 118-124。接下來對於探討 MBNL 蛋白質，調控選擇性剪接的分子機轉做進一步的整合，根據初步的實驗結果，發現 MBNL 基因與斑馬魚氯離子通道基因，N-端的 exons 選擇剪輯 (spliced) 的機制有關，是否這些不同剪輯的 zCIC-1 的蛋白質產物？其氯離子通道功能是否改變？以及與離子通道 (Ion channel) 與離子交換子 (Ion exchanger) 的特性作用是否有關？都需要做進一步的實驗來探討，從蛋白質功能方面決定離子通道上受 MBNL 選擇性剪輯調控的 Exons 後的改變後，下一階段與系上合作實驗合作，採取分子模擬的方式或結晶學的方法，進一步以結構為基礎的研究方法，來探討離子通道選擇性剪輯調控的機制。

本研究工作的主要價值，透過生物資訊工具的應用，將可輔助實驗上來找尋模型生物的同源基因，並且整合各個實驗室所長的分析能力，共同來解決生物上未知的問題，同時也可藉由蛋白質模組的分析，根據模型提出可能蛋白模組的功能，作為相關實驗結果的合

理探討與研究工作的延續，並且同時透過與實驗室間的合作，篩選相關 MBNL 所調控的下游基因，並找尋選擇性剪輯的作用與疾病的關係。

參考文獻

1. Pascual, M., Vicente, M., Monferrer, L. and Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation*, **74**, 65-80.
2. Worthington, M.T., Amann, B.T., Nathans, D. and Berg, J.M. (1996) Metal binding properties and secondary structure of the zinc-binding domain of Nup475. *Proc Natl Acad Sci U S A*, **93**, 13754-13759.
3. Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M. and Mlodzik, M. (1997) muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development*, **124**, 4321-4331.
4. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M. and Baylies, M.K. (1998) The muscleblind gene participates in the organization of Z-bands and epidermal attachments of Drosophila muscles and is regulated by Dmef2. *Dev Biol*, **195**, 131-143.
5. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. and Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum Mol Genet*, **11**, 805-814.
6. Kanadia, R.N., Urbinati, C.R., Crusselle, V.J., Luo, D., Lee, Y.J., Harrison, J.K., Oh, S.P. and Swanson, M.S. (2003) Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. *Gene Expr Patterns*, **3**, 459-462.
7. Squillace, R.M., Chenault, D.M. and Wang, E.H. (2002) Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev Biol*, **250**, 218-230.
8. Adereth, Y., Dammai, V., Kose, N., Li, R. and Hsu, T. (2005) RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1. *Nat Cell Biol*, **7**, 1240-1247.
9. Ranum, L.P. and Day, J.W. (2004) Myotonic dystrophy: RNA pathogenesis comes into focus. *Am J Hum Genet*, **74**, 793-804.
10. Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W. and Swanson, M.S. (2003) A muscleblind knockout model for myotonic dystrophy. *Science*, **302**, 1978-1980.
11. Liu, Y.F., Liu, H.Y., Tu, L.C., Lin, C.W., Hsiao, K.M. and Pan, H. (2008) Zebrafish muscleblind-like genes: identification, structural features and expression. *Comp Biochem Physiol B Biochem Mol Biol*, **151**, 118-124.
12. Fernandes, J.M., Kinghorn, J.R. and Johnston, I.A. (2007) Characterization of two paralogous muscleblind-like genes from the tiger pufferfish (*Takifugu rubripes*). *Comp Biochem Physiol B Biochem Mol Biol*, **146**, 180-186.
13. DeMaria, C.T., Sun, Y., Long, L., Wagner, B.J. and Brewer, G. (1997) Structural determinants in AUF1 required for high affinity binding to A + U-rich elements. *J Biol Chem*, **272**, 27635-27643.

14. Liang, W., Warrick, H.M. and Spudich, J.A. (1999) A structural model for phosphorylation control of Dictyostelium myosin II thick filament assembly. *J Cell Biol*, **147**, 1039-1048.
15. Drin, G., Casella, J.F., Gautier, R., Boehmer, T., Schwartz, T.U. and Antony, B. (2007) A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat Struct Mol Biol*, **14**, 138-146.
16. Bigay, J., Casella, J.F., Drin, G., Mesmin, B. and Antony, B. (2005) ArfGAP1 responds to membrane curvature through the folding of a lipid packing sensor motif. *EMBO J*, **24**, 2244-2253.
17. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N. and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum Mol Genet*, **13**, 495-507.
18. Michel, S.L., Guerrero, A.L. and Berg, J.M. (2003) Selective RNA binding by a single CCCH zinc-binding domain from Nup475 (Tristetraprolin). *Biochemistry*, **42**, 4626-4630.

FIGURES

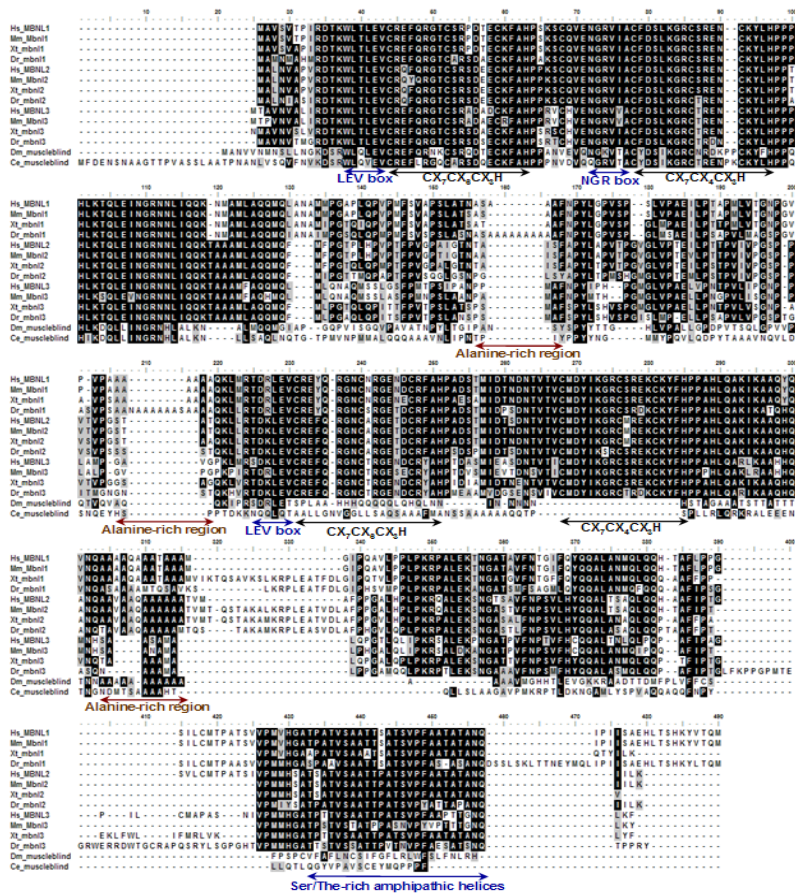


Fig.1 Sequence comparison among Muscleblind family members. Alternative spliced protein isoforms for each model animals orthologues. Sequences used for comparison are human (*Homo sapiens*) Hs_MBNL1(NP_066368.2), Hs_MBNL2 (NP_659002.1) and Hs_MBNL3(NP_060858.2); mouse (*Mus Musculus*) Mm_Mbnl1(NP_064391.2), Mm_Mbnl2 (NP_780550.1) and Mm_Mbnl3 (NP_598924.1); frog (*Xenopus tropicalis*) Xt_mbnl1(ENSXETP00000046798), Xt_mbnl2 (ENSXETP00000050703) and Xt_mbnl3(ENSXETP00000011905); fly (*Drosophila melanogaster*) Dm_muscleblind; worm (*Caenorhabditis elegans*) Ce_muscleblind (NP_510746.2). Alternative splice protein isoforms zmbnl1_C, zmbnl2_C and zmbnl3_B exemplify fish (*Danio retio*) Dr_mbnl1, 2, 3 proteins, respectively. Characteristic for the whole family is the presence of tandem CCCH Zinc fingers (CX7CX4CX6H and CX7CX6C3H, where C: Cys; H: His; X represents any amino acid), LEV and NGR boxes are indicated. Also conserved in all vertebrates MBNL proteins is a potential Ser/Thr amphipathic helices in C-terminal region. Some MBNL protein isoforms include low complexity region such as alanine-rich regions of unknown function. Alignments were generated with CLUSTALW with minor adjustment on the basis of structural information. The multiple alignments was drawn by BioEdit package gets from Tom Hall at North Carolina State University. Identical amino acid are boxed in blank and similar amino acids in grey.

[Insert Running title of <72 characters]

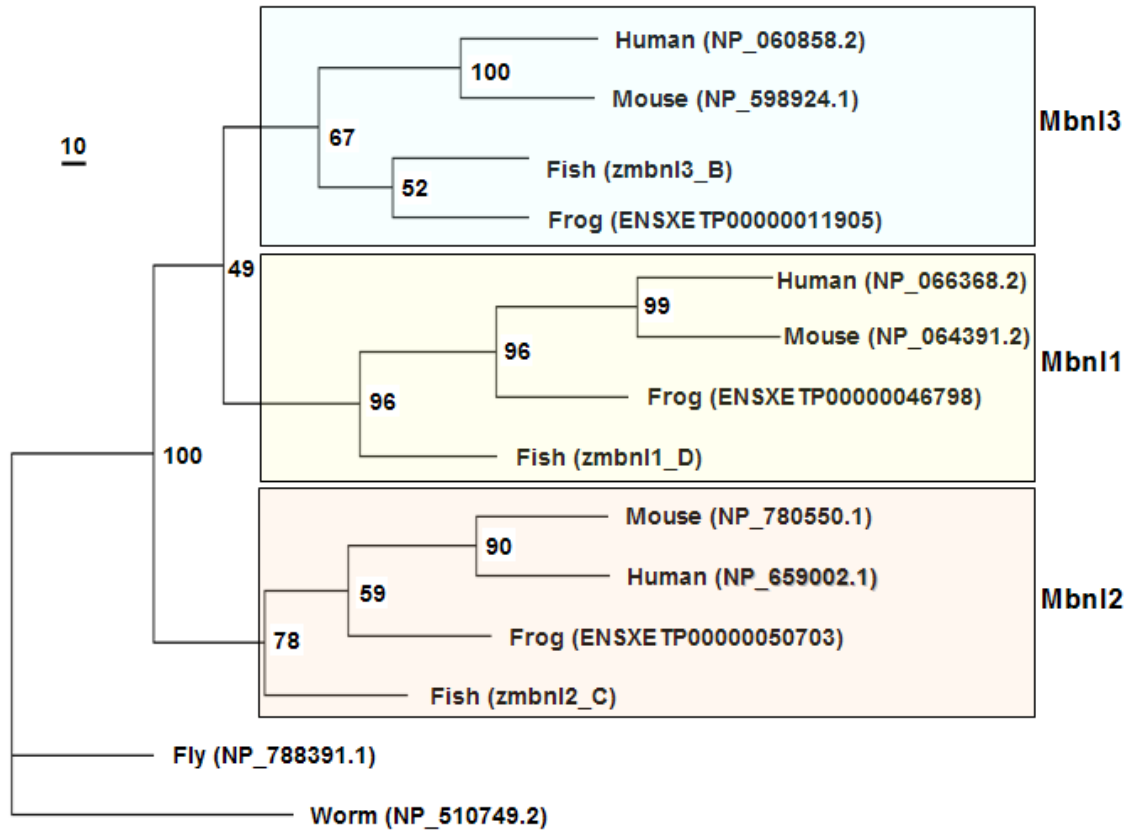


Fig.2 Phylogenetic analysis of MBNL proteins in model animals and fish. The similarity dendrogram of thirteen mbnl-related sequences were retrieved from protein database at the genomic database and aligned using a multiple sequence alignment method. Neighbor-joining algorithm was used and the numbers at the nodes were an indication of confidence level for the branches as determined by bootstrap analysis (100 bootstrap replicates). *C.elegans* and *Drosophila* muscleblinds were used as outgroup to root the tree. The length of the branch represents time to common ancestor. The scale bar on the top left corner of figure indicates 10% divergency along each branch. The genes of similar clusters were denoted by different color boxes.

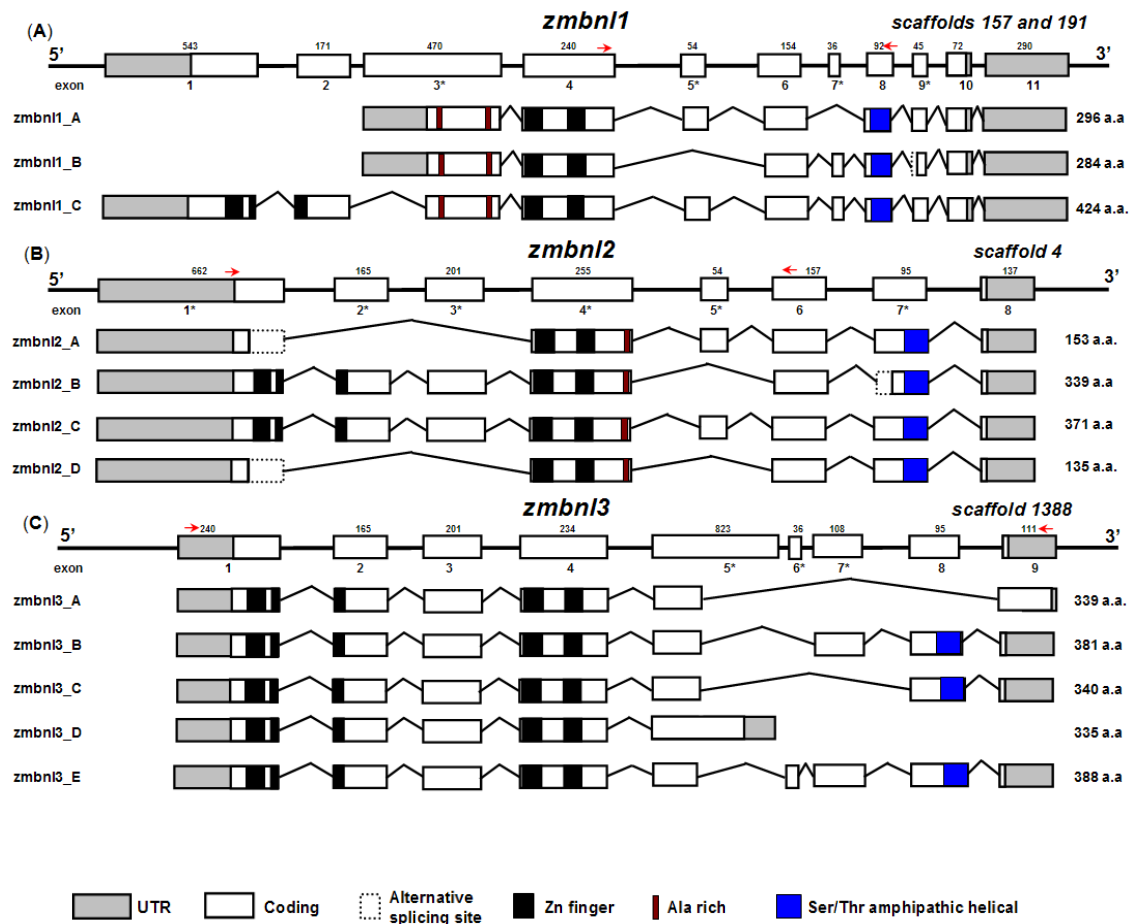


Fig.3 Schematic representation of the *D. retio mbnl*-related genes and their alternatively spliced transcripts. Alternative splicing and genome organization analysis for fish *zmbnl* primary transcripts that extensive protein diversities. Boxes linked by a horizontal line at the top of each panel represent exons and genomic DNA (*Zv.7* assembly), respectively, of *zmbnl1* (A), *zmbnl2* (B) and *zmbnl3* (C) genes. Exon-intron structure was represented from the alignment of cDNA and genomic sequences by sim4 program. The numbers on the bottom indicate exon numbers. Some untranslated exons are not presented in all *zmbnl* genes, as indicated. Only coding exons are represented for alternative splicing isoforms. Broken-lines linked specific exon combinations dedicated the combination of alternative splicing of fish *zmbnl* genes. Numbers indicate the length of each protein variants (a.a. amino acid). Exons are drawn approximately to scale, with numbers denoting the length (nucleotides) but the intervening genomic sequence is not. Untranslated regions (UTR), coding sequences, alternative splicing site, CCH Zinc finger domain (Zn finger), Alanine-rich regions (Ala rich) and Ser/Thr amphipathic helical are indicated as shown in the bottom. The primers used to amplify isoform-specific exons are denoted by arrows. Asterrisks denote the alternatively spliced exons.

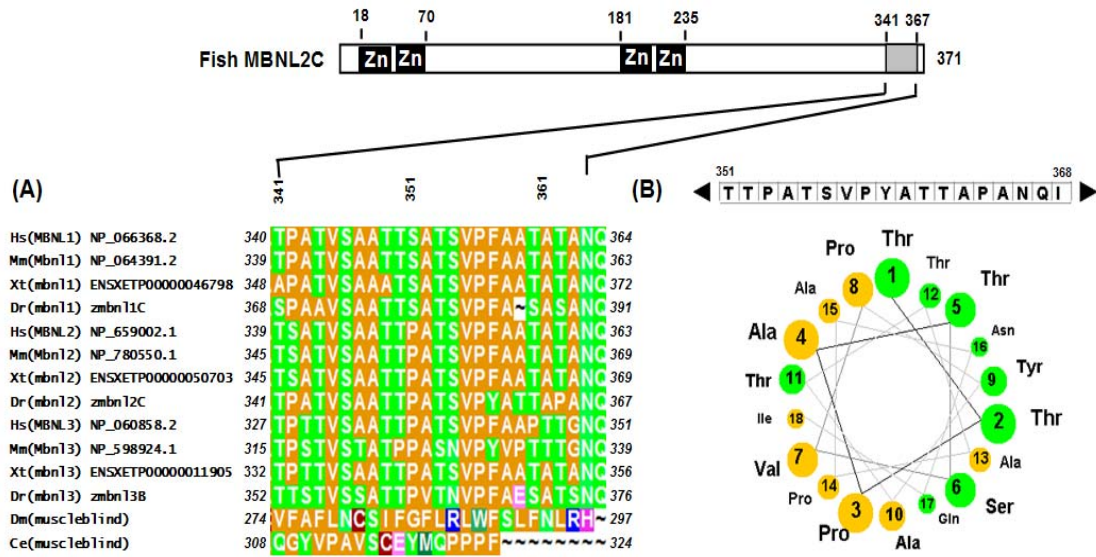


Fig.4 The C-terminal motif of mbnl-related genes adopts a Ser/The-rich amphipathic helical structure. **(A)** Multiple sequence alignment of the truncated C-terminal region of MBNL orthologues, including zebrafish MBNL2. The abbreviations used are as follows: HS *Homo sapiens*, Mm *Mus musculus*, Xt *Xenopus tropicalis*, Dr *Danio rerio* and Ce *Caenorhabditis elegans*. The NCBI Refseq and Ensembl accession number are indicated. Sequences were aligned using CLUSTALW and assembled using BioEdit program. Conserved residues are boxed with the following colour code: yellow, hydrophobic; green, serine and threonine; red, acidic. Dm and Ce muscleblind protein accession numbers are NP_788391.1 and NP_510746.2, respectively. **(B)** Helical-wheel representation of the 351-368 sequence of fish zmbnl2C with putative amphiphilic alpha-helix demonstrating partitioning of polar uncharged and hydrophobic amino acids (yellow and green circles, respectively). The helical wheel plot was generated with the aid of Java-applet at <http://cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html>.

[Insert Running title of <72 characters]