# The spectrin family of proteins: a unique coiled-coil fold for various molecular surface properties

Aurélie Nicolas, Olivier Delalande, Jean-François Hubert, Elisabeth Le Rumeur

HAL Id: hal-01020693

https://univ-rennes.hal.science/hal-01020693

Submitted on 8 Jul 2014

# The spectrin family of proteins: a unique coiled-coil fold for various molecular surface properties

Aurélie Nicolas[1, 2], Olivier Delalande[1, 2], Jean-François Hubert[1, 2] and Elisabeth Le Rumeur[1, 2, *]

[1]Université de Rennes 1, 35000 Rennes, France

[2]CNRS UMR6290, Institut de Génétique et Développement de Rennes, CS 34317, 35043 Rennes, France

*Corresponding author:

Elisabeth Le Rumeur

Institut de Génétique et Développement de Rennes (IGDR)

Faculté de Médecine, 35043 Rennes, Cedex, France

elisabeth.lerumeur@univ-rennes1.fr

**Abstract**

The spectrin superfamily is composed of proteins involved in cytolinker functions. Their main structural feature is a large central subdomain with numerous repeats folded in triple helical coiled-coils. Their similarity of sequence was considered to be low without detailed quantification of the intra- and intermolecular levels. Among the superfamily, we considered as essential to propose an overview of the surface properties of all the repeats of the five proteins of the spectrin family, namely α- and β-spectrins, α-actinin, dystrophin and utrophin. Therefore, the aim of this work was to obtain a quantitative comparison of all the repeats at both the primary sequence and the three-dimensional levels. For that purpose, we applied homology modelling methods to obtain structural models for successive and overlapping tandem repeats of the human erythrocyte α- and β-spectrins and utrophin, as previously undertaken for dystrophin, and we used the known structure of α-actinin. The matrix calculation of the pairwise similarities of all the repeat sequences and the electrostatic and hydrophobic surface properties throughout the protein family support the view that spectrins and α-actinin on one hand and utrophin and dystrophin on the other hand share some structural similarities, but a detailed molecular characterisation highlights substantial differences. The repeats within the family are far from identical, which is consistent with their multiple interactions with different cellular partners, including proteins and membrane lipids.

## Introduction

The proteins of the spectrin superfamily contain a calponin type actin binding domain at their N-termini followed by multiple spectrin repeats that could specify an actin crosslinking characteristic distance. The superfamily includes plakins, nesprins, plectins, α- and β-spectrin, α-actinin, dystrophin and utrophin (Jefferson et al., 2007). This superfamily of proteins is characterised by the presence of multiple copies of spectrin-like repeats arranged in tandem (Parry et al., 1992, Baines 2003, Jefferson et al., 2007). The spectrin-like repeats (or repeats) are composed of three alpha-helices (A, B and C) wrapped into a coiled-coil. In the present study, we focused among the superfamily on the five proteins members belonging to the so-called spectrin family, namely α- and β-spectrin, α-actinin, dystrophin and utrophin. The first two proteins can form dimers / tetramers while the last two remain as monomers (Broderick and Winder, 2005)

Although the sequence similarity of the repeats among the spectrin family has not been studied quantitatively in detail, it is considered to be rather low (Winder et al., 1995). The unique feature of this coiled-coil structure is that the residues of the primary sequence are distributed in heptads, i.e., on seven successive residues (a to g), where the two residues in the 'a' and 'd' positions are hydrophobic while the others are mainly hydrophilic. The coiled-coil folding of the repeats is principally maintained by the heptad hydrophobic residues, buried away from the molecular surface and being in close contact. Therefore, the accessible residues of the repeats are mostly hydrophilic and situated in positions other than 'a' and 'd' (Parry et al., 1992; Lupas, 1996). This structural characteristic was demonstrated by the first crystals of spectrin solved by X-ray diffraction in 1993 (Yan et al., 1993) and more recently for the dystrophin and utrophin single repeat 1 (Muthu et al., 2012).

Two repeats are linked to each other to compose a filament consisting of a long common helix formed by the third helix of the first repeat and the first helix of the second repeat. This linking has been solved by crystallography for several spectrin tandem or multiple repeats (Davis et al., 2009; Grum et al., 1999; Ipsaro and Mondragon, 2010; Ipsaro et al., 2009; Ipsaro et al., 2010; Kusunoki et al., 2004a; Kusunoki et al., 2004b; Stabach et al., 2009) and for the four repeats of α-actinin (Djinovic-Carugo et al., 1999; Ylanne et al., 2001). The remarkable feature of all these structures except the one of Grum (Grum et al., 1999) is that they appeared as dimers.

Although α-actinin has only 4 repeats, which have all been solved by X-ray crystallography (Djinovic-Carugo et al., 1999; Ylanne et al., 2001), the other proteins have a high number of repeats, and no crystal for the complete protein have been successfully obtained for any of them. However, these structures could be crucial for the understanding of the interactions or the mutation modifications of these proteins. In that context, we previously proposed homology models for all the dystrophin repeats (Legrand et al., 2011). Herein, we have applied this approach to an exhaustive study of the repeats of human utrophin and human erythrocyte α- and β-spectrin. We compared them to the previously

obtained models of dystrophin repeats and to the X-ray solved α-actinin repeat structure. We quantitatively assessed the similarity between these proteins at both the primary structure and the three-dimensional levels using electrostatic clustering methods and molecular hydrophobicity potential comparisons. We showed that spectrins and α-actinin on one hand and utrophin and dystrophin on the other hand shared some structural similarities, but a detailed molecular characterisation highlighted substantial differences. These differences are also related to a particular organisation of the exon and repeat coding in dystrophin and utrophin not observed in the other three proteins of the spectrin family.

## Materials and Methods

*Sequence alignment and isoelectric point calculations* – The sequences of human skeletal muscle utrophin and dystrophin and α- and β-spectrin from human erythrocytes were retrieved from the NCBI Protein Database, and sequences of α-actinin were retrieved from the deposited Protein Data Bank data (PDB: 1HCI) (see Fig. 1A for a schematic drawing of the proteins). To define the boundaries of the repeats, the alignment of dystrophin and utrophin repeats from Winder was used (Winder et al., 1995). The five proteins are schematically showed in Fig. 1A. Note that for α-spectrin, the repeat number according to Winder (Winder et al., 1995) was 20 complete repeats while in some papers, the SH domain accounts for a 10th "domain" of the protein which shifts the number to 21 "domains" (An et al., 2004; Li et al., 2010). We chose to number them from R1 to R20 according to Winder (Winder et al., 1995).

The sequences were aligned using the ClustalW2 program with the default parameters (Thompson et al., 1994; Larkin et al., 2007). The matrix similarity between the sequences was calculated using the Protdist program included in the PHYLIP package (Felsenstein, 1987). Isoelectric points were computed from the primary sequence using ProtParam on the ExPASy server (Gasteiger et al., 2005).

*Homology modelling* – The three-dimensional models were computed by the I–TASSER server (Zhang, 2008; Zhang, 2009). The sequences of two successive tandem repeats were submitted with an overlap of one repeat for the next submission, i.e., first repeat 1-2, and then repeat 2-3. The two tandem repeats of utrophin that include known hinges, R3-4 and R17-18, were omitted. The incomplete repeats -1 (α) and 17 (β) of the spectrins were not modelled. This strategy was used to obtain models for the potential helical linkers between adjacent repeat pairs. I-TASSER produced one to five models for each of the two-repeat sequences submitted, and only the model with the best C-score for each tandem repeat was retained. We did not use the X-ray solved structures of the erythrocyte spectrins to compare all the models obtained by the same method. However, the templates used by I-TASSER are all the X-ray crystallography structures of the spectrin repeats. In all, 74

models were obtained: 21 for dystrophin, 19 for utrophin, 19 for α-spectrin and 15 for β-spectrin. For α-actinin, we used the X-ray crystallographic four-repeat structure available in the PDB (Djinovic-Carugo et al., 1999) and subdivided it into three structures for each tandem repeat.

*Surface-property comparisons: hydrophobicity* - Surface hydrophobicity calculations were performed using the PLATINUM program, which is designed to match hydrophobic patches for the prediction of receptor-ligand complexes (Efremov et al., 2007; Pyrkov et al., 2009). This program allows the calculation and visualisation of molecular hydrophobic/hydrophilic surface properties using the concept of molecular hydrophobicity potential (MHP). The PLATINUM algorithm also provides the hydrophobic and hydrophilic composition (in %) of the molecular surfaces following the MHP definition. All the models were submitted to the PLATINUM web site, which provided potential maps visualised using the PyMOL program.

*Surface-property comparisons: electrostatics* - The web-based version of the PIPSA program (Protein Interaction Property Similarity Analysis) (Blomberg et al., 1999; Gabdoulline et al., 2007; Richter et al., 2008; Wade et al., 2001) was used to compare tandem repeats with respect to their electrostatic potentials. PIPSA quantifies the similarity in the surface properties of homologous proteins, which is particularly useful for comparing the surface properties of the tandem repeats of the spectrin family members. After the 3D alignment fit, the models were submitted to the PIPSA server (Richter et al., 2008), and the Adaptive Poisson-Boltzmann Solver (APBS) software was used to calculate the electrostatic potentials (Baker et al., 2001). To prevent the potential discrepancies observed in calculations performed at too high or low an ion concentration, we chose to present maps computed for an ionic strength of 50 mM. The similarity indexes (SI) for each pair of surface electrostatic potentials were calculated (Blomberg et al., 1999) and converted into a distance matrix ranging from 0 to 2 (Gabdoulline et al., 2007; Wade et al., 2001). The distance matrix was then subjected to a clustering procedure, and the corresponding heat map was generated. A visualisation of the electrostatic potentials projected on the molecular surfaces of the repeats was produced using PyMOL.

## Results

### Intramolecular sequence similarity

The sequences were aligned using ClustalW2 (Thompson et al., 1994; Larkin et al., 2007) (Supplementary Figure S1), and the similarity matrix between repeats was calculated using the Protdist program included in the PHYLIP package (Felsenstein, 1987). Only a poor sequence similarity was found between each single coiled-coil repeat of the five proteins. The pairwise sequence similarity scores of all repeats within each protein (Fig. 1B) indicated a median value below 0.2. The

highest internal sequence similarity scores were 0.38 and 0.30 for the α-spectrin and β-spectrin repeats, respectively. It is also interesting to highlight that the highest similarity recorded between repeats for the two spectrin molecules was observed for repeats that are distant in the filament: R5 and R13 (0.38), R1 and R14 (0.35) and R2 and R10 (0.35) for α-spectrin and R4 and R15 (0.30) for β-spectrin. The best similarity value for α-actinin repeats was very low at 0.19. Dystrophin and utrophin, despite a length comparable to the two spectrins (22-24 repeats vs 16-20 repeats for spectrins), showed a large range of sequence similarity scores among their repeats. The highest scores were no more than 0.25 for the dystrophin repeats (R8 and R17, R21 and R24) and 0.26 for the utrophin repeats (R12 and R2).
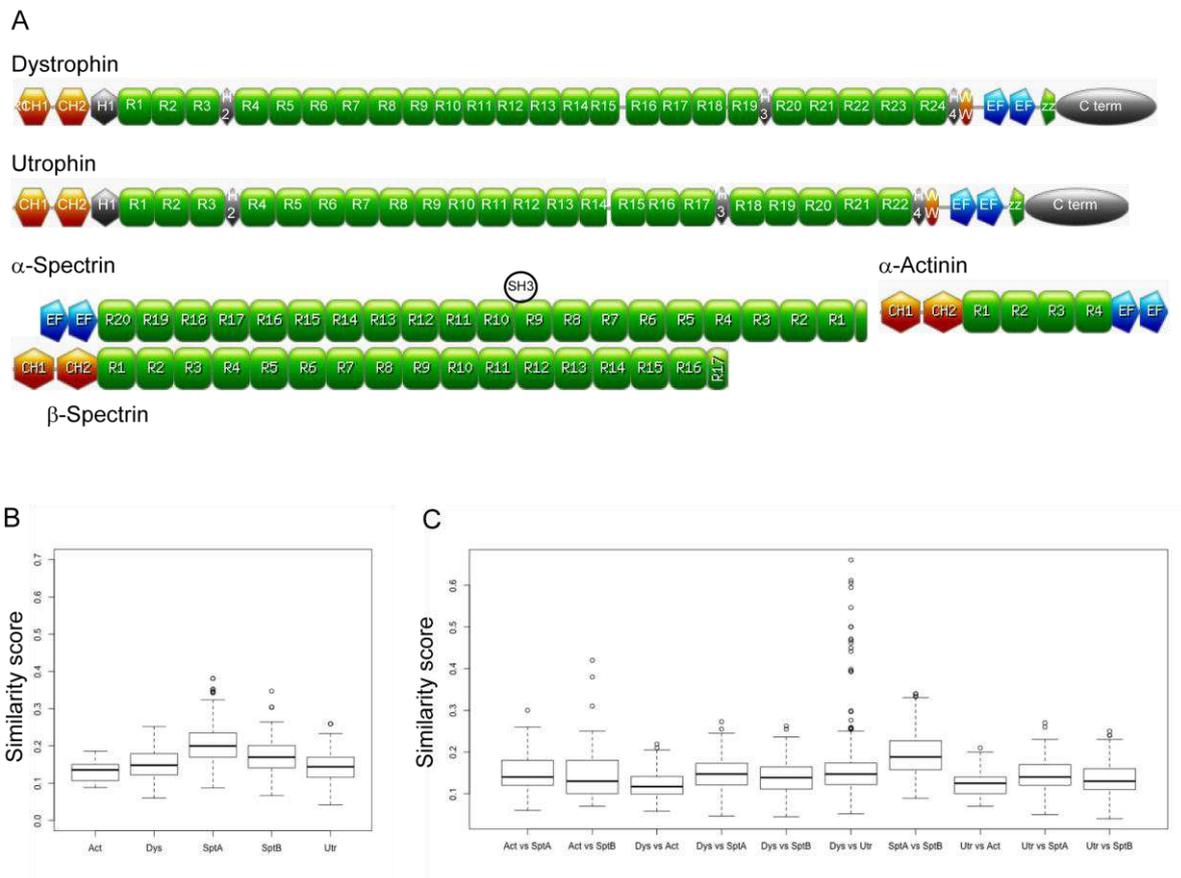


Figure 1. **Presentation of the five proteins and the similarity scores for their single repeats**. (A) Schematic drawing of the five proteins under study. Similarity scores are presented as box plots for (B) the pair-wise similarity of the repeats for each protein and (C) the pair-wise similarity of repeats of the five proteins in pairs. The box plots indicate the median values ± quartiles, and the vertical bar indicates the minimum and maximum values. When the values are larger than the 1.5 box size, individual values appear as points above or below the box plot. Act = α-actinin; Dys = dystrophin; SptA = α-spectrin; SptB = β-spectrin; Utr = utrophin.

*Intermolecular sequence similarity*

The highest median values of similarity between sequences of repeats in different proteins were observed for the two spectrins at 0.2 (Fig. 1C). For these two proteins, the highest sequence similarity was found for α-spectrin R9 and β-spectrin R13 (0.34); α-spectrin R16 and β-spectrin R4 (0.33), R12 (0.34) or R14 (0.33); α-spectrin R6 and β-spectrin R10 (0.33); and, finally, α-spectrin R12 and β-spectrin R16 (0.33). These scores for α-spectrin and β-spectrin defined protein domains with weak similarities between the two spectrins from α-spectrin R5 to R9 and from β-spectrin R9 to R13. We should note that the lowest sequence similarity score, with a mean value of 0.14, was found for α-spectrin R20 towards any of the β-spectrin repeats and β-spectrin R1 towards any of the α-spectrin repeats, underlying the specificity of these repeats involved in the α-spectrin/ β-spectrin hetero-dimer formation.

Dystrophin and utrophin did not show high sequence conservation with either the spectrins or α-actinin, with values well below 0.30 (Supplementary Figure S2). Clear high sequence similarities are found for the α-actinin repeats R1 and R2 with β-spectrin R1 and R2, with scores of 0.42 and 0.38, respectively. Surprisingly, the α-actinin repeat R2 is the closest to all α- and β-spectrin repeats. Finally, the most remarkable similarity was found between the dystrophin and utrophin coiled-coil repeats that can be seen on the boxplot with outliers distributed over the top bar (Fig. 1C). Dystrophin R1 and utrophin R1 present a sequence similarity up to 0.66, and the second highest value (0.61) was for dystrophin R8 / utrophin R8 and dystrophin R11 / utrophin R11.

Globally, the highest scores in sequence similarity between the dystrophin and utrophin proteins were found for the regions covering R1 to R3 (from 0.44 to 0.66), R8 to R13 (from 0.40 to 0.61) and R20 to R24 of dystrophin and R18 to R22 of utrophin (from 0.30 to 0.55), with 13 of them higher than 0.40. The regions covering R4 to R7 of dystrophin and utrophin are not very similar in sequence. A lack of similarity was also found for dystrophin R14 or R15 with utrophin R14 and for dystrophin R18 or R19 with utrophin R17. Finally, dystrophin repeats are very similar to utrophin repeats, more similar than to one another, which is also true for utrophin itself.

*Exon and repeats phasing in the five proteins*

In a previous work, we showed a partial phasing of the exon coding for the repeats of dystrophin and the repeat sequence (Nicolas et al., 2012). We therefore wondered if the same case holds for the proteins of the spectrin family. On the alignment of the repeats of these proteins, we differentiated the sequences coded by successive exons (Fig. 2). The sequences coded by even exons are in grey, while the sequences coded by odd exons coded sequences are uncoloured. It is clear that the helices B of the utrophin repeats are coded by two successive exons with the borders exactly aligned with the "b" and "c" residues of the 3rd heptad of the helices for all repeats except repeat 14. This pattern constitutes a partial phasing of the protein domains and the exons and is similar to what

we observed for dystrophin approximately in the middle of helices B. By contrast, there was no such phasing in the coding of the α- and β-spectrins or the α-actinin repeats (Fig. 2).
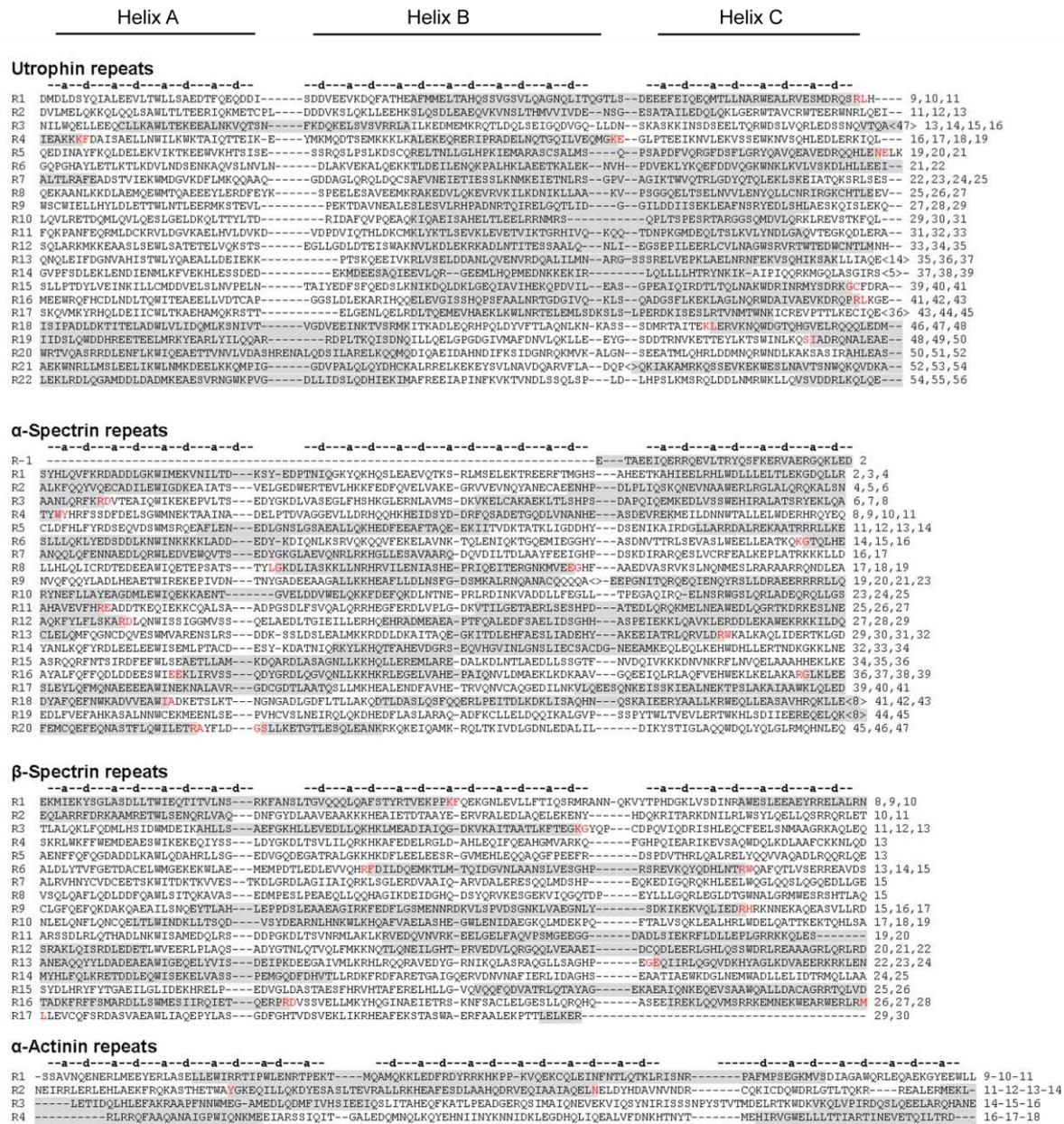


Figure 2: **Sequence alignment of the repeats of utrophin, α- and β-spectrin and α-actinin and the exons coding for these sequences.** The repeats were aligned by ClustalW using the default parameters according to Winder et al. (Winder et al., 1995). The repeat numbers are mentioned on the left side of the repeat sequence. The numbers of the exons coding for the repeats are mentioned at the right side of the repeat sequence. The repeat sequences coded by successive exons are underlined alternatively in grey (even) or left blank (odd). When the successive exons are not in-frame, the two residues coded by the two successive exons are coloured red.

*Isoelectric point of the single repeats of the spectrin family*

The isoelectric points calculated from the primary sequence ranged from 3.93 (UtrR1) to 9.01 (DysR3). The range of values is quite large for dystrophin, utrophin and β-spectrin, but there is a narrow range around the mean value for α-spectrin and α-actinin (Fig. 3). These two last proteins appear to be more acidic as a whole than the other three. However, the dystrophin values are significantly different from the utrophin and α-spectrin values, while the α-actinin and α- and β-spectrin values were not so different from each other. In addition, and as previously shown, there is an alternation of acidic and basic repeats on the region covering repeats 11 to 17 of dystrophin, which was not observed for the other proteins.
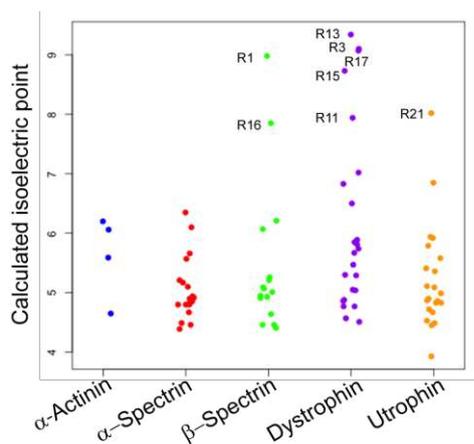


Figure 3. **Calculated isoelectric points of the single repeats for the five proteins.** The isoelectric points were computed by ProtParam of the ExPASy server. The individual values of the repeats appear as dots. Mean values (SD) are given above the plot. The values from dystrophin were significantly different from utrophin and from α-spectrin, as shown above the plot. Some values appearing outside the mean, such as for β-spectrin, dystrophin and utrophin, are tagged.

*Homology modelling*

Not surprisingly, the I-TASSER threading procedure identified spectrin repeats as the best structural templates for all the tandem repeats, specifically 1CUN (chicken-brain α-spectrin repeat R16-17) (Grum et al., 1999), 1U4Q (chicken-brain α-spectrin repeat R15-17) (Kusunoki et al., 2004b), 1S35 (erythroid β-spectrin R8-9) (Kusunoki et al., 2004a), 3EDV (β2-spectrin repeat R14-16) (Davis et al., 2009) and 1F57 (erythrocyte ß-spectrin R14-15) (Ipsaro et al., 2009). As shown above, the identity score was very low, ranging between 0.08 and 0.18. However, the sequence coverage was very good, with values ranging between 86 and 99%, and the C-scores ranged between -0.72 and 0.80, which indicated that all the models proposed adopt correct folds (Supplementary Table S1). The models were further assessed using PROCHECK to calculate Ramachandran plots. The percentage of residues in core ranged from 81.8% (dystrophin R16-16) to 98.5% (β-spectrin R15-16) and percentage of disallowed residues ranged from 0 to 2.6% (dystrophin R18-19) (Supplementary Table S1). These

results indicated that all models were of high quality. Furthermore, the I-TASSER model of R8-9 and R14-15 of erythrocyte β-spectrin were compared to the corresponding PDB structures 1S35 and 3F57 (Supplementary Fig. S3A-D). The Cα-atoms Root Mean Square Deviation (RMSD) between the structures and the corresponding models were respectively 1.60 Å and 1.36 Å, as these moderate deviations may originate from the last relaxation step of the I-TASSER modelling procedure. The electrostatic and hydrophobic surface properties are also compared and indicated a high likeness between structures and models.

The structural models for the coiled-coil repeats of both α- and β-spectrins all presented a regular α-helix fold at the linker region between two consecutive repeats. For instance, the models with the highest similarity tandem repeats, α-spectrin R6-7 and β-spectrin R10-11, presented a geometric deviation of 3.7 Å (Cα RMSD) for their protein backbone (Fig. 4).
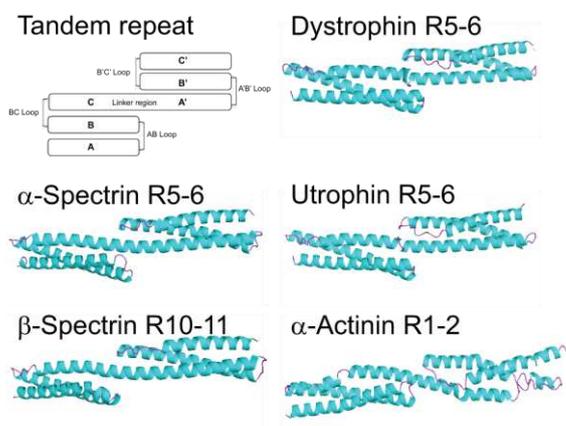


Figure 4. **Representative models of the five proteins**. A schematic drawing of a tandem repeat is shown on the left upper side. Helices A, B and C constitute the first repeat and A', B' and C' the second repeat. The helices are connected by loops named AB, BC, A'B' and B'C'. The linker region is common to the two consecutive repeats and may be either helical or not. The two models of α- and β-spectrin showed helical linkers, while the two models of dystrophin and utrophin showed non-helical linkers. Actinin was taken from the structure obtained by X-ray diffraction and showed helical linkers. In all models, the A helix of the first repeat is placed in front and at the bottom left of the image.

However, even though the models were obtained using spectrin X-ray crystallisation structures with helical linkers, non-helical linkers were observed for utrophin R5-6, R10-11 and R14-15, as previously observed for the dystrophin homologous tandem repeats (R5-6, R10-11 and R15-16). Finally, the homology models were well structured in the coiled-coil fold by comparison to the distorted α-actinin experimental structures. Compared to spectrin, dystrophin and utrophin models and structures, α-actinin presented more elongated structures. This is due to the fact that the sequences of the repeats are longer by about 6 to 10 residues than the repeats of the other proteins. Therefore, R1

and R2 of α-actinin have A-helices prolonged by 8 to 9 residues while R3 and R4 have their C-helices prolonged by 6 to 9 residues compared to spectrins, dystrophin or utrophin. This leads to B-helices less kinked than in the other structures and gives an elongated and compact structure for α-actinin tandem repeats.

*Surface properties - Electrostatic potentials*

The electrostatic surfaces of the tandem repeats were clearly dissimilar, and they showed a large variety in the distribution of positive and negative potential patches (Supplementary Figure S4). Such patches often appeared to extend over more than one repeat surface, while sometimes the two repeats in a tandem exhibited opposite electrostatic properties. To further quantify the surface electrostatic potential similarity of the tandem repeats, we analysed our models using the PIPSA program (Richter et al., 2008). The previously obtained dystrophin models were included in this analysis (Legrand et al., 2011). The resulting PIPSA heat map was divided into four major groups (Fig. 5).
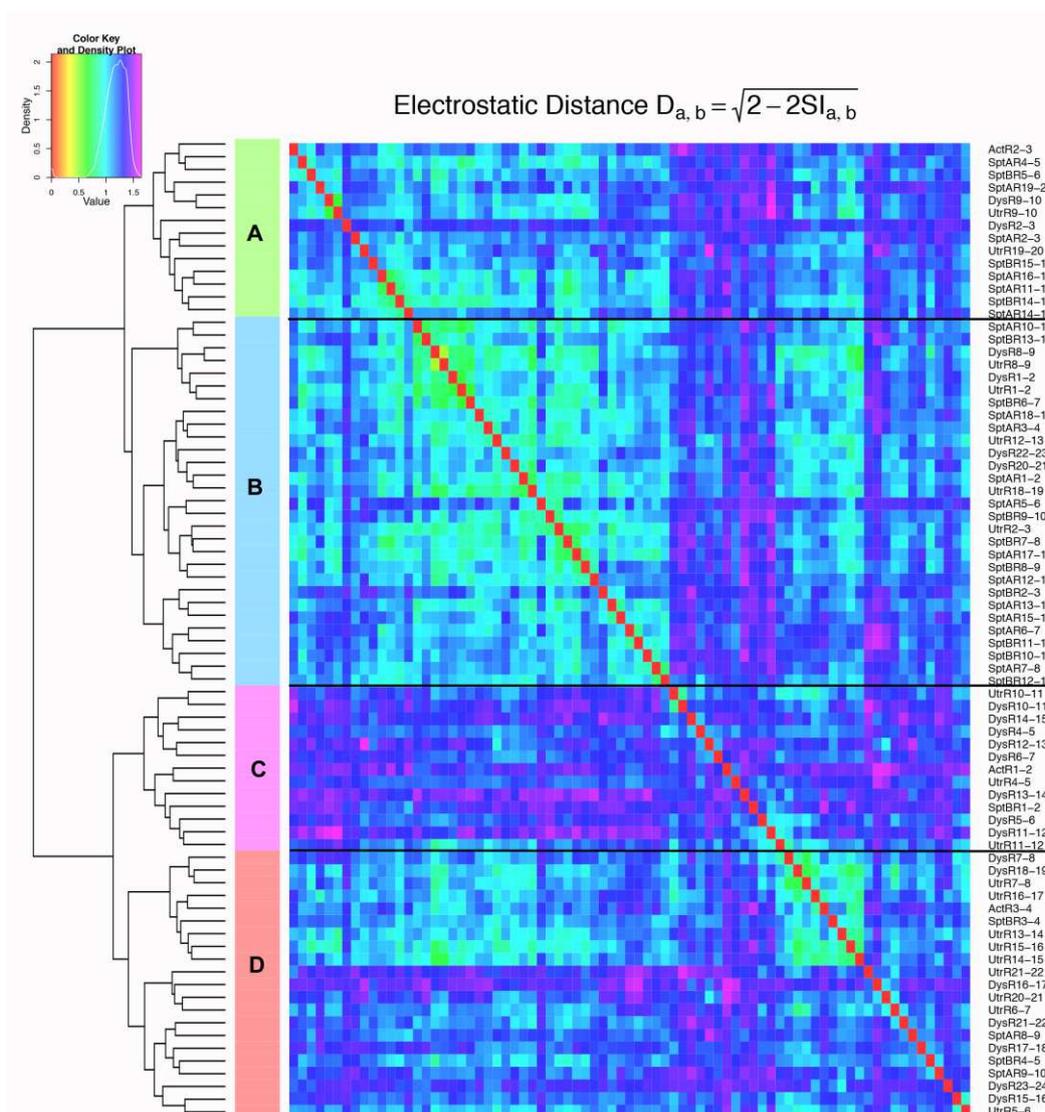
Figure 5. **Matrix visualisation of an all-pairwise comparison of the electrostatic potentials of the tandem repeat surfaces.** The matrix (heat map) is shown with colour coding corresponding to the distances calculated from the Hodgkin similarity indices for the electrostatic potentials (upper left side). Distance between molecules a and b is defined by the equation given on the top, where SI is the Hodgkin similarity index between molecules a and b. Tandem repeats with highly similar electrostatic potentials are clustered together, and four groups have been clustered (groups A, B, C, D).

Group A was characterised by strongly negative surfaces with positive patches distributed homogenously at the molecular surface (Fig. 6 and Supplementary Figure S4). The tandem repeats of group B had also negative surfaces but with positive patches distributed on one side of the molecular surface. Group C was characterised by an alternation of large positive and negative patches. The electrostatic surfaces in the group D were comparable to the ones in group C but with smaller positive patches.

Nicolas et al.                                                                                           Page 12
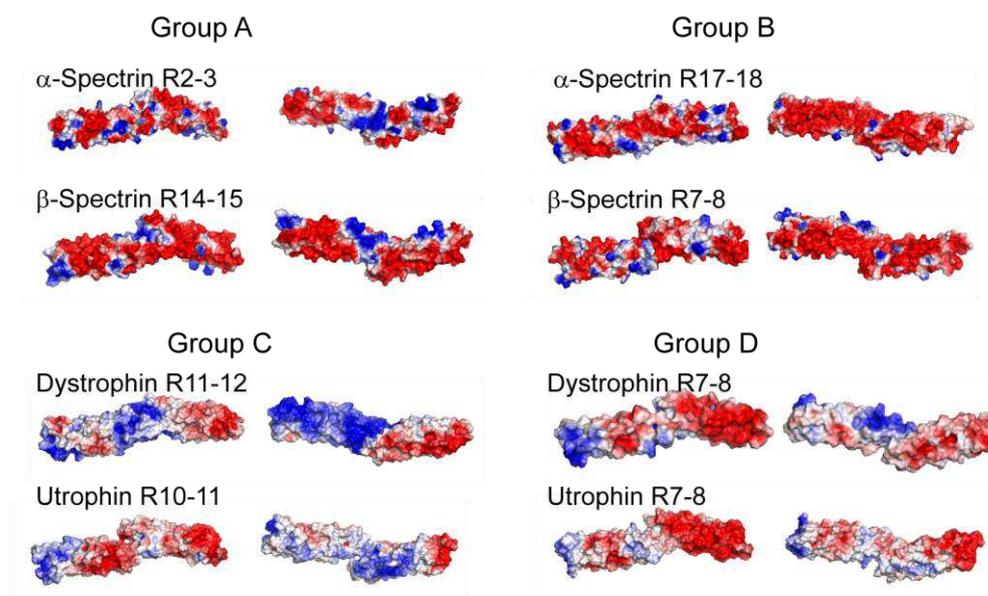
Figure 6. **Representation of the electrostatic potential projected on the solvent-accessible surface of tandem repeats representative of the four groups obtained by PIPSA analysis**. Each model was coloured using the APBS electrostatic potential calculated for an ionic strength of 50 mM, and the surface colours were clamped at -3 (red) and +3 (blue) $kTe^{-1}$. The molecules are presented in the same orientation as in Figure 4A.

The distribution of the tandem repeats of the five proteins in each group was clearly non-random (Table 1).

**Table 1.** Distribution of the five proteins in the four groups obtained by PIPSA analysis of the electrostatic surfaces of the tandem repeats.

| | Number of tandem repeats in each group (% of the total) | | | |
|---|---|---|---|---|
| | Group A | Group B | Group C | Group D |
| Dystrophin (n=21) | 2* (14%) | 5 (17%) | 8 (62%) | 7 (33%) |
| Utrophin (n=19) | 2 (14%) | 5 (17%) | 3 (23%) | 9 (43%) |
| α-Spectrin (n=19) | 6 (43%) | 11 (37%) | 0 | 2 (10%) |
| β-Spectrin (n=15) | 3 (21%) | 9 (30%) | 1 (8%) | 2 (10%) |
| α-Actinin (n=3) | 1 (7%) | 0 | 1 (8%) | 1 (5%) |
| Total (n=78) | 14 (100%) | 30 (100%) | 13 (100%) | 21 (100%) |

*Values indicate the number of tandem repeats of each protein in the different groups. The percentage of protein tandem repeats in each group is indicated in parentheses.
The α- and β-spectrin tandem repeats were predominantly distributed in groups A and B.

The three α-actinin tandem repeats were observed in three different clusters. The dystrophin tandem repeats were distributed rather homogeneously in groups B, C and D, and the utrophin tandem repeats were predominantly distributed in groups B and D. Group A predominantly consisted of α-spectrin, which made up 43% of its tandem repeats. Group B consisted primarily of α- and β-spectrins, respectively constituting 37% and 30% of its tandem repeats. Group C consisted predominantly of dystrophin models (62%), and the D group was composed of 43% utrophin and 33% dystrophin.

*Surface properties – Molecular hydrophobicity potentials*

As is often the case with cytoplasmic proteins, the overall tandem repeat surfaces were mostly hydrophilic with nonpolar surfaces covering 25 to 39% of the surface (Fig. 7). However, there was a large heterogeneity in and between the tandem repeats of the proteins. As a whole, α-actinin has the less hydrophobic repeat domain, with a mean hydrophobic surface representing 28% of its surface, while utrophin has the highest hydrophobic repeat domain with a mean hydrophobic surface of 36% of the whole molecular surface (p = 0.03 from the dystrophin repeat domain to p = $2.10^{-10}$ from α-spectrin). In addition, the values for the utrophin tandem repeats were in a small range of values from 32 to 38%, while the values for the other tandem repeats were in a larger range (their standard deviations ranged from 2.9 to 3.5 vs 1.6 for utrophin). Dystrophin repeat domain appeared the closest to utrophin but with three extra values of less than 30% hydrophobicity. The two spectrin repeat domains appeared identically hydrophobic with values in a range of 17 over 19 and 14 over 16 tandem repeats with 24 to 31% hydrophobic surface. This globally hydrophobic character arises from the presence of large patches at the molecular surface (Supplementary Figure S5). The highest value was observed for the tandem repeat R18-19 of dystrophin with a hydrophobic surface as large as 40% of the total surface.
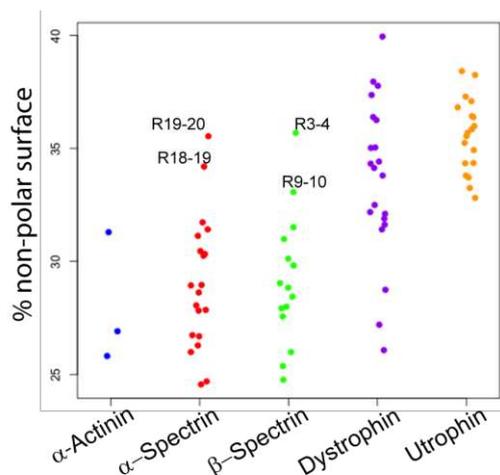


Figure 7. **Molecular hydrophobicity potentials of the tandem repeats of the five proteins.** The individual values of the repeats appear as dots. Mean values (SD) are given above the plot. *The values from dystrophin and utrophin are significantly different from each other and from the three other protein values.

**Discussion**

The spectrin family of proteins comprises the five well-known proteins analysed in the present study, although the spectrin superfamily includes other proteins such as plakins, nesprins or plectins that could be the subject of another study (Jefferson et al., 2007; Ortega et al., 2011). The main structural feature of all these proteins is the presence of repeats folded in coiled-coils due to a heptad pattern of hydrophobic residues (Parry et al., 1992, 2008). As in every family, the similarity of sequence is one of the major characteristics shared by their members. However, this similarity has been qualified as low without true quantification before our present study. Here, we reported that the α- and β-spectrin repeats showed the highest intramolecular sequence similarity values, which are likely related to a better conservation of the heptad pattern by comparison to the dystrophin and utrophin repeats. As mentioned previously (Winder et al., 1995), the length of the helices is more consistent in the spectrin repeats compared to the utrophin and dystrophin repeats, while the three helices of α-actinin repeats have very similar lengths compared to the other proteins. This intramolecular similarity of the spectrin repeats could also be related to the more precise placement of helix breaker residues in spectrins. Indeed, proline or glycine residues are clearly delimiting the stops in helix A and B and thus the start of the AB and BC loops. By contrast, dystrophin, utrophin and α-actinin exhibited low intramolecular sequence similarity. Dystrophin and utrophin showed relatively heterogeneous lengths of the three helices with highly variable delimitation of B and C helices, which likely accompanies the relatively low sequence similarity of the repeats.

As expected, the highest intermolecular sequence similarity was recorded between utrophin and dystrophin repeats, but dystrophin repeats R14, R15, R18 and R19 are the most divergent from the utrophin repeats. Despite a low similarity between α-actinin repeats and the repeats from other proteins, a high similarity was noticed between α-actinin R2 and all the spectrin repeats. Moreover, a high similarity score was found between α-actinin R1-2 and β-spectrin R1-2.

Correlated to these primary sequence variations in length and sequence similarity, we observed that all models of α- and β-spectrin tandem repeats showed helical linkers between successive repeats, in contrast to utrophin as shown here and to dystrophin as shown previously (Legrand et al., 2011). The helical linkers of spectrins should most likely be involved in the linearity of the spectrin filament, and this putative straight structure may be enhanced by the heterodimerisation of α- and β-spectrin. The efficiency of the heterodimer formation is likely increased by the singularity of sequences of both associated ending repeats, α-spectrin R20 and β-spectrin R1.

The non-helical linkers of dystrophin lead to kinks between successive repeats and define individualised regions all along the central domain (Legrand et al., 2011). This phenomenon is most likely the same for utrophin, which showed non-helical linkers in its dystrophin homologous repeats. This arrangement most likely enables a non-linear structure to be maintained through modifications

such as rotations or kinks between successive repeats. Finally, it should more effectively differentiate utrophin and dystrophin as flexible monomeric filaments, as opposed to the straight heterotetrameric filaments of spectrins. Similarly, α-actinin forms dimers that maintain a rather straight topology, although some flexibility is maintained (Djinovic-Carugo et al., 1999; Zaman and Kaazempur-Mofrad, 2004; Golji et al., 2009).

The evolution of spectrin family members has been previously studied from the protein sequence perspective (Pascual et al., 1997; Baines, 2003; Broderick and Winder, 2005). Spectrin family members were hypothesised to diverge from an ancestral α-actinin molecule. This α-actinin diverged into three branches, the first branch leading to the modern α-actinin, the second branch to dystrophin and utrophin and the third branch to the α- and β-spectrins. Indeed, the α- / β-spectrin ancestor split is visible in the α- and β-spectrins, with β-spectrin R1 and R2 diverging from the ancestral α-actinin repeats R1 and R2 and the α-spectrin R19 and R20 diverging from the ancestral α-actinin repeats R3 and R4 (Pascual et al., 1997). The acquisition of new repeats occurred by further duplication and elongation events.

In agreement with these results, a high similarity score was observed here between α-actinin R1-2 and β-spectrin R1-2. However, a lower similarity score was observed between α-actinin R3-4 and α-spectrin R19-20, while a high similarity between α-actinin R2 and all other spectrin repeats was observed. This discrepancy with previous reports regarding evolution allows us to hypothesise that the newly appeared repeats evolved after the duplication of the ancestral α-actinin R2.

Our results show that particular features were observed specifically in utrophin and dystrophin, such as partial phasing of the repeats and the respective coding exons, variable length in their helices and the presence of non-helical linkers between several identified repeats. These characteristics concur with the hypothesis of the early divergence of utrophin and dystrophin from the ancestral α-actinin and could also reveal a common ancestor diverging later into utrophin and dystrophin. The low similarity reported among all the repeats may be the consequence of the parallel evolution of these homologous proteins and the addition of new repeats with new functions. This concept could be supported by our analysis of the potential surface properties that also clearly help to distinguish these filamentous proteins.

The use of APBS and PIPSA allowed us to calculate and compare the electrostatic surface potentials of the tandem repeats from the five proteins. Considering all the models, despite substantial heterogeneity of the electrostatic surfaces, the analyses highlighted four distinct groups, A, B, C and D. Interestingly, the tandem repeats from the five proteins are not randomly distributed in these groups. The tandem repeats from the two spectrins appeared mostly in two groups, A and B, distinct from the two other groups, C and D, in which the tandem repeats from utrophin and dystrophin were clustered. This observation supports the idea that the repeats from the spectrins and the repeats from utrophin and dystrophin are far from similar and amplifies this idea to a structural perspective.

However, some interesting exceptions were observed. The tandem repeat R1-2 of α-actinin appeared in the same group C as the R1-2 of β-spectrin, with these four repeats being considered to derive from the same two repeats. This group C is more characterised by the presence of large basic patches at the molecular surface and is therefore quite different from the tandem repeats of groups A and B.

Remarkably, the four tandem repeats of dystrophin involving R11 to R15 appeared in the same group C as these two tandem repeats. Amman *et al.* (Amann et al., 1998) already proposed that the basic isoelectric points observed for this region of dystrophin defined a singular track, which was further characterised as a second actin binding domain (ABD2). However, the region is not entirely basic. It appears that there is an alternation of basic (R11, R13, R15 and R17) and acidic (R12, R14 and R16) repeats that could define a groove for the interaction between filamentous actin and dystrophin. This interacting sub-domain is clearly missing in the homologous region of utrophin, which does not interact with actin (Amann et al., 1999). This point was confirmed by the observation that the dystrophin R11 to R15 and utrophin R13 to R17 regions belong to different electrostatic clusters.

In accordance with the heptad pattern's projection of the polar residue side chains outside the coiled-coil, the tandem repeat surfaces are mostly hydrophilic. However, the tandem repeat surfaces also display hydrophobic patches of potential interest. What was highly striking was the observation that the coiled-coil domain of utrophin is a highly hydrophobic region compared to the corresponding domain of spectrins and α-actinin, with a mean difference of approximately 6%. However, dystrophin showed high heterogeneity among the repeats, from low hydrophobic tandem repeats such as R13-14 and R14-15 (26 – 27% of hydrophobic surface) to the highly hydrophobic tandem repeat R18-19 (40% of hydrophobic surface). The varied surface properties among the repeats are likely to be correlated to the varied interaction properties of the dystrophin repeats.

Indeed, as utrophin and dystrophin are monomeric proteins, the level of hydrophobic patches accessible to the solvent should be lower than the level that occurs in the monomer form of a dimeric protein, but this is not the case. This global feature in the physicochemical properties of the molecular surface is most likely a functional key for these peripheral proteins, which are able to dynamically bind to membrane phospholipids all along the molecule. It is effectively the case in dystrophin (Le Rumeur et al., 2010), by contrast to spectrins, in which only specific repeats interact with membrane lipids (An et al., 2004; Grzybek et al., 2006). These lipid-interacting repeats are the 8 to 9 of α-spectrin and the 2 to 4 and 12 to 14 of β-spectrin. The repeats 8-9 of α-spectrin are not in the most hydrophobic part of the protein while the repeats 2-4 and 12-14 of β-spectrin are among those that are the most hydrophobic of the molecule. Interestingly, the most hydrophobic tandems are those that interact with ankyrin (R14-15 with 31.5% hydrophobic surface) or are involved in the dimerization of the two spectrins (R1-2 of β-spectrin and R19-20 / 20-21 depending on the numbering of α-spectrin) (Li et al., 2010). The fact that utrophin has a repeat domain highly hydrophobic could be interpreted as utrophin being a molecule with numerous unknown binding partners.

From sequence similarity and molecular potential measurements, it appeared that the filamentous central domain of both utrophin and dystrophin (75% of overall protein) showed clear specificities in comparison with the same regions in the spectrins and α-actinin, such as regular exon phasing towards coiled-coil organisation, monomeric and non-straight filament structure and large hydrophobic and positive potentials. This observation is supported by the hypothesis concerning the evolution of utrophin, dystrophin and spectrins from a common α-actinin ancestor. The individualisation of each coiled-coil repeat, in terms of structure regularity or surface potential, would arise from the need to specifically interact with cellular partners, proteins or phospholipids. The diversity found in the coiled-coil repeats between spectrins and utrophin/dystrophin is also underlined by the great variety in the nature of the pathologies related to the dysfunction of one or the other protein (Gallagher, 2004; Bennett and Healy, 2008; Baines, 2008; Tuffery-Giraud et al., 2009; Nicolas et al., 2012). This study clearly indicates that the replacement strategies proposed for the therapy of genetic diseases, such as dystrophies, cannot be as simple as inducing an overexpression of a rescue homologous protein or by the conception of a chimeric protein including those pseudo-homologs.

In conclusion, our data support the view that spectrins and α-actinin on the one hand and utrophin and dystrophin on the other hand share some structural similarities. Nevertheless, a detailed molecular characterisation indicates that the repeats within the family are far from being identical, which is consistent with their multiple and specific interactions in the cell.

**Figure Legends**

# References

Amann, K.J., B.A. Renley, and J.M. Ervasti, 1998. A Cluster of Basic Repeats in the Dystrophin Rod Domain Binds F-actin through an Electrostatic Interaction. J Biol Chem 273: 28419-28423.

Amann, K.J., A.W. XGuo, and J.M. Ervasti, 1999. Utrophin Lacks the Rod Domain Actin Binding Activity of Dystrophin. J Biol Chem 274: 35375-35380.

An, X., X. Guo, h. Sum, J. Morrow, W. Gratzer, and N. Mohandas, 2004. Phosphatidylserine binding sites in erythroid spectrin: location and implications for membrane stability. Biochemistry 43: 310-315.

Baines, A.J., 2003. Comprehensive analysis of all triple helical repeats in beta-spectrins reveals patterns of selective evolutionary conservation. Cell Mol Biol Lett 8: 195-214.

Baines, A.J., 2008. Mechanisms of elliptocytosis: significant spectrin substitutions. Blood 111: 5417.

Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon, 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98: 10037-41.

Bennett, V., and J. Healy, 2008. Organizing the fluid membrane bilayer: diseases linked to spectrin and ankyrin. Trends Mol Med 14: 28-36.

Blomberg, N., R.R. Gabdoulline, M. Nilges, and R.C. Wade, 1999. Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. Proteins 37: 379-87.

Broderick, M.J., and S.J. Winder, 2005. Spectrin, alpha-actinin, and dystrophin. Adv Protein Chem 70: 203-46.

Davis, L., K. Abdi, M. Machius, C. Brautigam, D.R. Tomchick, V. Bennett, and P. Michaely, 2009. Localization and structure of the ankyrin-binding site on beta2-spectrin. J Biol Chem 284: 6982-7.

Djinovic-Carugo, K., P. Young, M. Gaudel, and M. Saraste, 1999. Structure of the alpha-actinin rod: molecular basis for cross linking of actin filaments. Cell 98: 537-546.

Efremov, R.G., A.O. Chugunov, T.V. Pyrkov, J.P. Priestle, A.S. Arseniev, and E. Jacoby, 2007. Molecular lipophilicity in protein modeling and drug design. Curr Med Chem 14: 393-415.

Felsenstein, J., 1987. Estimation of hominoid phylogeny from a DNA hybridization data set. J Mol Evol 26: 123-31.

Gabdoulline, R.R., M. Stein, and R.C. Wade, 2007. qPIPSA: relating enzymatic kinetic parameters and interaction fields. BMC Bioinformatics 8: 373.

Gallagher, P.G., 2004. Hereditary elliptocytosis: spectrin and protein 4.1R. Semin Hematol 41: 142-64.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, and A. Bairoch, 2005. Protein Identification and Analysis Tools on the ExPASy Server, p. 571-607, in: J. M. Walker, (Ed.), The Proteomics Protocols Handbook, Humana Press.

Golji, J., R. Collins, and M.R. Mofrad, 2009. Molecular mechanics of the alpha-actinin rod domain: bending, torsional, and extensional behavior. PLoS Comput Biol 5: e1000389.

Grum, V.L., D. Li, R.I. MacDonald, and A. Mondragon, 1999. Structures of two repeats of spectrin suggest models of flexibility. Cell 98: 523-535.

Grzybek, M., A. Chorzalska, E. Bok, A. Hryniewicz-Jankowska, A. Czogalla, W. Diakowski, and A.F. Sikorski, 2006. Spectrin-phospholipid interactions. Existence of multiple kinds of binding sites? Chem Phys Lipids 141: 133-141.

Ipsaro, J.J., and A. Mondragon, 2010. Structural basis for spectrin recognition by ankyrin. Blood 115: 4093-101.

Ipsaro, J.J., L. Huang, and A. Mondragon, 2009. Structures of the spectrin-ankyrin interaction binding domains. Blood 113: 5385-93.

Ipsaro, J.J., S.L. Harper, T.E. Messick, R. Marmorstein, A. Mondragon, and D.W. Speicher, 2010. Crystal structure and functional interpretation of the erythrocyte spectrin tetramerization domain complex. Blood 115: 4843-52.

Jefferson, J. J., C. Ciatto, L. Shapiro, and R.K. Liem, 2007. Structural analysis of the plakin domain of bullous pemphigoid antigen1 (BPAG1) suggests that plakins are members of the spectrin superfamily. J Mol Biol, 366: 244-57.

Kusunoki, H., R. MacDonald, and A. Mondragon, 2004a. Structural insights onto the stability and flexibility of unusual erythroid spectrin repeats. Structure 12: 645-656.

Kusunoki, H., G. Minasov, R. MacDonald, and A. Mondragon, 2004b. Independent movement, dimerization and stability of tandem repeats of chicken brain alpha-spectrin. J Mol Biol 344: 495-511.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins, 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

Le Rumeur, E., S.J. Winder, and J.F. Hubert, 2010. Dystrophin: More than just the sum of its parts. Biochim Biophys Acta 1804: 1713-1722.

Legrand, B., E. Giudice, A. Nicolas, O. Delalande, and E. LeRumeur, 2011. Computational Study of the Human Dystrophin Repeats: Interaction Properties and Molecular Dynamics. Plos One 6: e23819.

Li, D., S.L. Harper, H.Y. Maksimova, P.G. Gallagher, and D.W. Speicher, 2010. A comprehensive model of the spectrin divalent tetramer binding region deduced using homology modelling and chemical cross-linking of a mini-spectrin. J Biol Chem 285: 29535-45.

Lupas, A., 1996. Coiled coils: new structures and new functions. Trends Biochem Sci 21: 375-82.

Muthu, M., K.A. Richardson, and A.J. Sutherland-Smith, 2012. The crystal structures of dystrophin and utrophin spectrin repeats: implications for domain boundaries. PLoS One 7: e40066.

Nicolas, A., C. Lucchetti-Miganeh, R. Ben Yaou, J.C. Kaplan, J. Chelly, F. Leturcq, F. Barloy-Hubler, and E. Le Rumeur, 2012. Assessment of the structural and functional impact of in-frame mutations of the DMD gene, using the tools included in the eDystrophin online database. Orphanet J Rare Dis 7: 45.

Ortega, E. R.M. Buey, A. Sonnenberg, and J.M. de Pereda, 2011. The structure of the plakin domain of plectin reveals a non-canonical SH3 domain interacting with its fourth spectrin repeat. J Biol Chem 286: 12429-38.

Parry, D.A., T.W. Dixon, and C. Cohen, 1992. Analysis of the three-alpha-helix motif in the spectrin superfamily of proteins. Biophys J 61: 858-67.

Parry, D.A.D., R.D.B. Fraser, and John M. Squire, 2008. Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. J Struct Biol 163: 258-269.

Pascual, J., J. Castresana, and M. Saraste, 1997. Evolution of the spectrin repeat. Bioessays 19: 811-817.

Pyrkov, T.V., A.O. Chugunov, N.A. Krylov, D.E. Nolde, and R.G. Efremov, 2009. PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. Bioinformatics 25: 1201-2.

Richter, S., A. Wenzel, M. Stein, R.R. Gabdoulline, and R.C. Wade, 2008. webPIPSA: a web server for the comparison of protein interaction properties. Nucleic Acids Res 36: W276-80.

Stabach, P.R., I. Simonovic, M.A. Ranieri, M.S. Aboodi, T.A. Steitz, M. Simonovic, and J.S. Morrow, 2009. The structure of the ankyrin-binding site of beta-spectrin reveals how tandem spectrin-repeats generate unique ligand-binding properties. Blood 113: 5377-84.

Thompson, J.D., D.G. Higgins, and T.J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-4680.

Tuffery-Giraud, S., C. Beroud, F. Leturcq, R.B. Yaou, D. Hamroun, L. Michel-Calemard, M.P. Moizard, R. Bernard, M. Cossee, P. Boisseau, M. Blayau, I. Creveaux, A. Guiochon-Mantel, B. de Martinville, C. Philippe, N. Monnier, E. Bieth, P. Khau Van Kien, F.O. Desmet, V. Humbertclaude, J.C. Kaplan, J. Chelly, and M. Claustres, 2009. Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. Hum Mutat 30: 934-45.

Wade, R., R. Gabdoulline, and F. De Rienzo, 2001. Protein Interaction Property Similarity Analysis. Int J Quant Chem 83: 122-127.

Winder, S.J., T.J. Gibson, and J. Kendrick-Jones, 1995. Dystrophin and utrophin: the missing links! FEBS Letters 369: 27-33.

Yan, Y., E. Winograd, A. Viel, T. Cronin, S. Harrison, and D. Branton, 1993. Crystal Structure of the Repetitive Segments of Spectrin. Science 262: 2027-2030.

Ylanne, J., K. Scheffzek, P. Young, and M. Saraste, 2001. Crystal structure of the alpha-actinin rod reveals an extensive torsional twist. Structure 9, 597-604

Zaman, M.H., and M.R. Kaazempur-Mofrad, 2004. How flexible is alpha-actinin's rod domain? Mech Chem Biosyst 1: 291-302.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.

Zhang, Y., 2009. I-TASSER: fully automated protein structure prediction in CASP8. Proteins 77 Suppl 9: 100-13.