



**HAL**  
open science

# Etude in silico de dystrophines tronquées dans les myopathies de Duchenne et de Becker

Aurélie Nicolas

► **To cite this version:**

Aurélie Nicolas. Etude in silico de dystrophines tronquées dans les myopathies de Duchenne et de Becker . Bio-Informatique, Biologie Systémique [q-bio.QM]. Université Rennes 1, 2012. Français. NNT: . tel-01343326

**HAL Id: tel-01343326**

**<https://theses.hal.science/tel-01343326>**

Submitted on 8 Jul 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Européenne de Bretagne*

pour le grade de  
**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**

*Mention : Biologie et Sciences de la Santé*

**Ecole doctorale Vie-Agro-Santé**

présentée par

**Aurélie NICOLAS**

Préparée à l'unité de recherche UMR CNRS 6290 IGDR  
Institut de Génétique et de Développement de Rennes  
UFR Sciences de la Vie et de l'Environnement (SVE)

---

**Etude *in silico* de  
dystrophines  
tronquées dans les  
myopathies de  
Duchenne et de  
Becker**

**Thèse soutenue à Rennes  
le 23 octobre 2012**

devant le jury composé de :

**Catherine ETCHEBEST**

Professeur – Université Paris 7 / *rapporteur*

**Pierre LEHN**

Professeur – Université de Bretagne Occidentale /  
*rapporteur*

**Sylvie ODENT**

Professeur (PU-PH) – Université de Rennes 1 /  
*examineur*

**Rumen ANDONOV**

Professeur – Université de Rennes 1 / *examineur*

**Frédérique BARLOY-HUBLER**

Chargé de recherche - CNRS / *co-directeur de  
thèse*

**Elisabeth LE RUMEUR**

Maitre de conférence (MCU-PH) - Université de  
Rennes 1 / *directeur de thèse*



# Remerciements

Ce travail a été réalisé dans l'équipe SIM en collaboration avec Frédérique Barloy-Hubler de l'équipe SP@RTE de l'Institut de Génétique et de Développement de Rennes (UMR CNRS 6290). Je remercie le CNRS de m'avoir accordé une bourse de thèse.

Pour commencer, je tiens à remercier Catherine Etchebest et Pierre Lehn pour avoir accepté d'être les rapporteurs de ce travail. Je remercie également Sylvie Odent et Rumen Andonov pour avoir accepté de faire partie de mon jury.

Je remercie tout particulièrement ma directrice de thèse Elisabeth Le Rumeur pour m'avoir encadrée durant ces trois années. Merci de m'avoir fait confiance et de m'avoir laissée une grande liberté pour travailler tout en étant toujours là quand j'en avais besoin. Ton enthousiasme communicatif m'a beaucoup encouragé. Bien que nous ayons des spécialités différentes, notre duo fonctionnait très bien, tu m'as énormément appris sur la dystrophine et les myopathies et, en échange, je t'ai initiée à la bioinformatique. Merci également de m'avoir épaulé pour préparer mon avenir qui s'annonce très bien, notamment grâce à toi. Pour résumer, si c'était à refaire, je resignerais avec toi.

Je remercie également ma co-directrice de thèse, Frédérique Barloy-Hubler qui était également mon encadrante de stage de M2. C'est notamment grâce à toi que j'ai obtenu cette bourse de thèse, sans toi ce travail n'aurait pas vu le jour. Bien que tu suivais cette thèse de loin, ton expertise en bioinformatique et ton regard extérieur m'ont été très utiles. Merci également pour tes bons petits plats et pour m'avoir fait découvrir la patate douce.

Je remercie chaleureusement mon encadrant non-officiel Olivier Delalande qui m'a détournée de la bioinformatique "traditionnelle" pour me plonger dans l'univers de la modélisation et de la dynamique moléculaire. Merci pour tout ce que tu m'as appris, ta disponibilité même tard le soir, ta gentillesse et ton humour parfois délirant.

Je tiens également à remercier Emmanuel Giudice de l'équipe TAF qui, avant d'être un collaborateur était mon enseignant. Tu m'as enseigné la théorie de la modélisation pendant mon master et maintenant tu m'aides à l'appliquer. Merci pour le temps passé, ta patience et les précieux conseils.

Je remercie nos collaborateurs de l'Institut Cochin, Rabah Ben Yaou, Jean-Claude Kaplan, Jamel Chelly et France Leturcq de nous avoir fait confiance pour mettre en place une collaboration recherche fondamentale - recherche clinique. Merci pour votre aide pour l'écriture de l'article.

Je tiens à remercier toute l'équipe SIM et les membres de l'équipe METRIQ avec qui nous partagions le couloir. Nous formions une grande équipe où la bonne humeur et l'entraide régnaient que ce soit aux pauses cafés, le midi ou à l'occasion de nos repas en dehors du travail. Je remercie l'ancien chef de SIM Arnaud Bondon et le nouveau chef Jean-François Hubert de m'avoir accueilli dans leur équipe. Je remercie mes voisines de bureau Céline Raguènes-Nicol et Fanny Noury de m'avoir fait une place et avec qui nous avons eu de grandes conversations sur tout et rien. Un grand merci à l'équipe de mots fléchés du midi : Angélique Chéron, Christophe Tascon, Olivier et bien sûr Elisabeth. Je remercie les thésards (anciens et actuels) avec qui nous avons partagé de bons moments Baptiste Legrand, Joe Sarkis, Fabien Balezeau qui m'a initiée à LaTeX et Matthieu Simon.

Un merci particulier à Sarah Ameziane qui est devenue une amie. Merci pour le soutien pendant notre période de rédaction, c'était plaisant de rédiger avec toi dans le bureau d'à côté. Je savais que je pouvais compter sur toi pour faire une petite pause quand je saturais. Merci aussi pour les cinés, les restaurants et toutes les autres choses qu'on a faites à l'extérieur.

Merci aux anciens membres de l'équipe B@SIC avec qui j'ai commencé ma thèse. Merci à Stéphane Avner pour avoir grandement approfondi mes connaissances en informatique. Merci à Céline Lucchetti-Miganeh pour tous tes conseils, ton aide pour les détails administratifs et surtout ta bonne humeur communicative. Merci aux deux David, David Thybert et David Goudenège qui finissaient leurs thèses alors que je commençais la mienne. Un merci particulier à David G. pour m'avoir initiée aux Macs, aux http request et autres trucs obscurs de geeks. Merci d'avoir été un bon ami en dehors du laboratoire et d'avoir été mon cobaye culinaire. Je n'ai pas refait de gâteau aux Oréos depuis que t'es parti.

Je remercie tout particulièrement mes amis qui m'ont soutenu sans toujours tout comprendre, vous avez toujours été là pour me changer les idées et me soutenir. Merci à ceux qui m'envoient leur bonnes ondes de plus ou moins loin particulièrement Delphine l'alsacienne d'adoption. Merci à ceux qui me soutiennent d'un peu plus près notamment Marina. Merci à Nathalie qui accourt dès que j'ai un problème aussi loufoque soit-il. Merci à tous les autres dont la liste serait trop longue à énumérer.

Je remercie également mes parents, mon frère et toute ma famille qui m'ont toujours soutenue et aidée dès que j'en avais besoin.

# Table des matières

<b>1</b>	<b>Introduction générale</b>	<b>9</b>
1.1	La dystrophine . . . . .	10
1.1.1	Le gène . . . . .	10
1.1.1.1	Caractéristiques du gène . . . . .	10
1.1.1.2	Les transcrits alternatifs . . . . .	11
1.1.2	La protéine . . . . .	12
1.1.2.1	Rappels sur la structure des protéines . . . . .	12
1.1.2.2	Structure de la dystrophine . . . . .	14
1.1.2.3	Domaines fonctionnels de la dystrophine . . . . .	18
1.1.2.4	Rôle physiologique et localisation de la dystrophine . . . . .	20
1.1.2.5	La famille des protéines homologues à la dystrophine . . . . .	20
1.2	Les dystrophinopathies . . . . .	23
1.2.1	Aspects historiques . . . . .	23
1.2.2	La myopathie de Duchenne . . . . .	24
1.2.2.1	Caractéristiques cliniques . . . . .	24
1.2.2.2	Caractéristiques histologiques . . . . .	25
1.2.3	La myopathie de Becker . . . . .	26
1.2.3.1	Caractéristiques cliniques . . . . .	26
1.2.3.2	Caractéristiques histologiques . . . . .	27
1.2.3.3	Les mutations du gène <i>DMD</i> . . . . .	28
1.2.4	Corrélation génotype - phénotype ? . . . . .	29
1.2.5	Les essais thérapeutiques dans la Myopathie de Duchenne . . . . .	31
1.2.5.1	La thérapie génique . . . . .	32
1.2.5.2	Le saut d'exons . . . . .	35
1.2.5.3	La thérapie cellulaire . . . . .	37
1.2.5.4	La translecture du codon STOP . . . . .	38
1.2.5.5	La surexpression de l'utrophine . . . . .	38
1.3	L'apport de la bioinformatique . . . . .	40
1.3.1	Dates importantes dans l'histoire de la bioinformatique . . . . .	40
1.3.2	Les bases de données . . . . .	41
1.3.2.1	Une nécessité pour la recherche actuelle . . . . .	41
1.3.2.2	Les bases de données générales et les base de données spécifiques . . . . .	42
1.3.2.3	L'intérêt d'une base de données spécifique à la dystrophine . . . . .	43
1.3.3	La modélisation moléculaire . . . . .	43
1.3.3.1	L'intérêt de la modélisation par homologie . . . . .	43
1.3.3.2	Les champs de force . . . . .	46
1.3.3.3	La dynamique moléculaire . . . . .	46
1.3.3.4	Quelques succès en modélisation moléculaire . . . . .	49
1.3.3.5	L'intérêt de la modélisation pour l'étude de la dystrophine . . . . .	50
1.4	Objectifs du travail . . . . .	51

<b>2</b>	<b>La modélisation des répétitions de la dystrophine</b>	<b>53</b>
2.1	Matériel et Méthodes . . . . .	53
2.1.1	Prédiction de la structure 3D des répétitions . . . . .	53
2.1.2	La dynamique moléculaire . . . . .	56
2.1.2.1	Les modèles de dystrophines tronquées choisis . . . . .	57
2.1.2.2	La dynamique moléculaire tout-atome avec NAMD . . . . .	59
2.1.2.3	La dynamique moléculaire gros grain avec GROMACS . . . . .	61
2.1.3	L'analyse des modèles des répétitions . . . . .	62
2.1.3.1	Les tandems de répétitions . . . . .	62
2.1.3.2	Les répétitions tronquées . . . . .	62
2.2	L'étude des répétitions natives . . . . .	64
2.2.1	L'analyse des séquences primaires . . . . .	64
2.2.2	L'analyse des modèles structuraux . . . . .	64
2.2.3	Les propriétés de surface des modèles structuraux . . . . .	66
2.2.4	La relaxation par dynamique moléculaire des modèles . . . . .	68
2.3	L'étude des répétitions tronquées . . . . .	71
2.3.1	Les répétitions hybrides et les répétitions fracturées . . . . .	71
2.3.2	L'analyse des modèles relaxés par DM . . . . .	71
2.3.2.1	La stabilité des protéines . . . . .	71
2.3.2.2	Les propriétés de surface . . . . .	83
2.3.2.3	L'analyse des contacts moléculaires . . . . .	90
2.3.2.4	Le cas particulier de la charnière H3 . . . . .	95
2.3.3	La validation par les dynamiques gros-grains . . . . .	98
2.3.3.1	La protéine $\Delta 45-49$ . . . . .	98
2.3.3.2	La protéine $\Delta 45-51$ . . . . .	100
<b>3</b>	<b>La base de données eDystrophin</b>	<b>105</b>
3.1	Matériel et méthodes . . . . .	105
3.1.1	Matériel . . . . .	105
3.1.2	Construction de la base de données . . . . .	106
3.1.2.1	Les données génétiques . . . . .	106
3.1.2.2	Les données sur la dystrophine native . . . . .	106
3.1.2.3	Les données cliniques . . . . .	107
3.1.2.4	Les données sur les dystrophines mutées . . . . .	108
3.1.2.5	Mises à jour de la base de données . . . . .	110
3.2	Résultats . . . . .	111
3.2.1	La base de données relationnelle . . . . .	111
3.2.2	L'interface d'eDystrophin . . . . .	113
3.2.2.1	La page <i>Wild-type dystrophin</i> . . . . .	113
3.2.2.2	Les pages <i>Mutated dystrophin</i> . . . . .	115
3.2.3	Vue d'ensemble du contenu d'eDystrophin . . . . .	118
3.2.3.1	Les statistiques d'eDystrophin . . . . .	118
3.2.3.2	Etude de cas . . . . .	119
3.2.4	Mise en évidence de l'alignement des répétitions sur l'alignement des exons . . . . .	125
<b>4</b>	<b>Discussion générale</b>	<b>127</b>
	<b>Bibliographie</b>	<b>141</b>
	<b>Annexes</b>	<b>159</b>

<b>A Exemple de résultats I-TASSER</b>	<b>159</b>
<b>B Alignement des séquences de R16-21 et des quatre protéines tronquées</b>	<b>161</b>
<b>C Article paru dans Plos One</b>	<b>165</b>
<b>D Article paru dans Orphanet Journal of Rare Diseases</b>	<b>177</b>

# Liste des abréviations

$\Delta$ 45-47 Protéine issue de la délétion des exons 45 à 47

ABD Domaine de liaison à l'actine

ADNc ADN complémentaire

BMD Myopathie de Becker

DCM Cardiomyopathie

DGC Dystrophine-glycoprotéine complexe

DM Dynamique moléculaire

DMD Myopathie de Duchenne

H1 à H4 Charnière 1 à 4 (Hinge 1 à 4)

IMD Myopathie intermédiaire

LGBM Laboratoire de Biochimie et de Génétique de l'Institut Cochin

nNOS Oxyde nitrique synthase neuronale (neuronal Nitric Oxide Synthase)

R1 à R24 Répétition 1 à 24

R17//18 Répétition reconstituée à partir de la partie N-terminale de R17 et de la partie C-terminale de R18

RMSD Ecart quadratique moyen (Root Mean Square Deviation)

# Chapitre 1

## Introduction générale

Ce travail de thèse s'articule autour de trois problématiques majeures :

- Comment la structure de la dystrophine permet-elle de comprendre sa fonction ?
- Que peut-on déduire des mutations identifiées chez les patients ?
- Comment l'étude de la structure de la dystrophine et de ses mutations permet-elle de corréler le génotype et le phénotype ?

Afin de répondre à ces questions, différentes approches ont été mises en place en parallèle. Cependant, pour faciliter la clarté de ce mémoire, j'ai choisi de les scinder en deux parties.

La première partie traite des approches de modélisation et de dynamique moléculaires effectuées sur les répétitions des dystrophines natives et mutées chez des patients. L'analyse des répétitions natives, initiée avant mon arrivée dans l'équipe mais à laquelle j'ai participé, sera présentée de façon succincte. Cependant, la méthodologie mise en place est la base de l'analyse des répétitions mutées que j'ai réalisée ensuite.

La deuxième partie présente la base de données eDystrophin créée afin d'analyser de manière automatisée les dystrophines mutées issues de mutations en phase présentes chez les patients atteints de dystrophinopathies. Cette base de données rassemble notamment les modèles structuraux de dystrophine décrits dans la partie précédente.

## 1.1 La dystrophine

La dystrophine, avec ses 3685 acides aminés, est l'une des plus grandes protéines humaines. Elle est encodée par le gène *DMD* (2,4Mb) situé sur le chromosome X et composé de 79 exons. Ce gène est sujet à de nombreuses mutations conduisant à deux dystrophinopathies majeures. La plus répandue est la myopathie de Duchenne avec une incidence de 1 pour 3500 naissances de garçons. Elle est responsable d'une dégénérescence musculaire progressive conduisant au décès des patients avant leur trentième année. La dystrophie musculaire de Becker est moins sévère, avec un phénotype clinique plus hétérogène. La myopathie de Duchenne fait l'objet de nombreuses études cliniques et génétiques afin d'aider à mettre au point des thérapies actuellement en phase d'essai clinique. Cependant, la dystrophine est très peu étudiée et son rôle au sein de la cellule musculaire n'est pas encore complètement précisé.

### 1.1.1 Le gène

#### 1.1.1.1 Caractéristiques du gène

Le gène *DMD* a été séquencé pour la première fois en 1986 [Monaco 1986, Koenig 1987]. Il est situé sur le bras court du chromosome X (locus Xp21.2) (figure 1.1) et représente 1,5% du chromosome. Ses 2,4 millions de paires de bases font de lui le plus grand gène humain [Monaco 1992, Coffey 1992]. Il est composé d'introns de tailles variables allant de 107b à plus de 200kb et de 79 exons. Les exons représentent seulement 0,6% de la séquence génomique et sont transcrits en un ARN de 14kb [Roberts 1993] (figure 1.2).

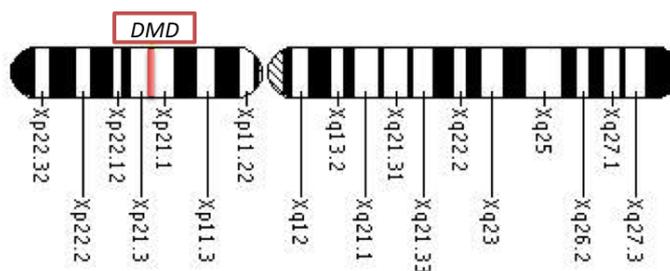


FIGURE 1.1 – Localisation du gène *DMD* sur le bras court du chromosome X.

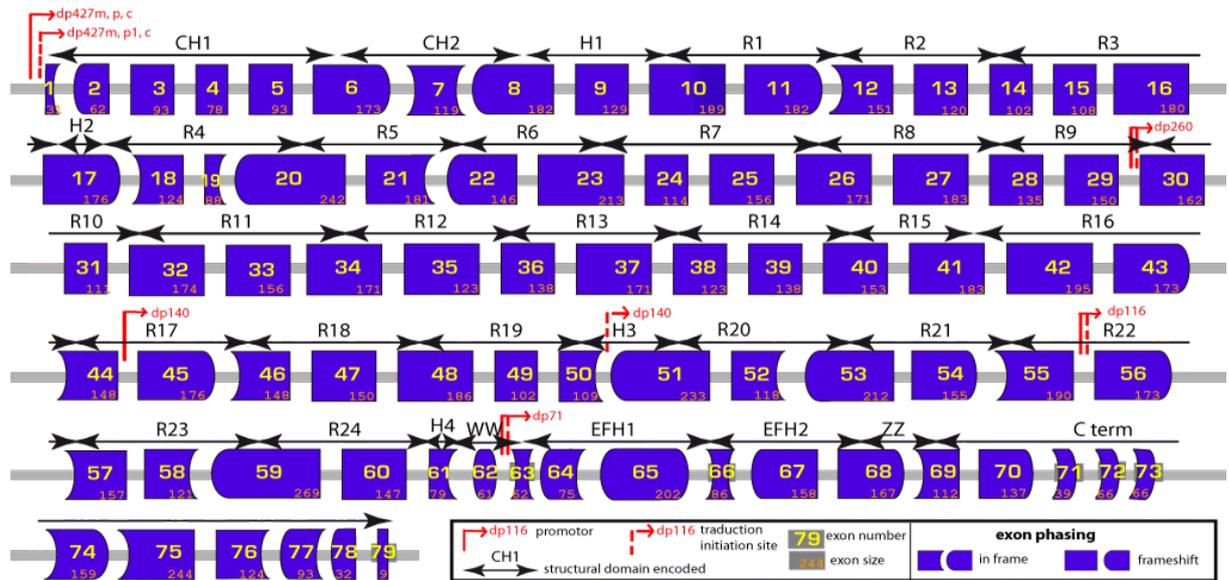


FIGURE 1.2 – Représentation des exons du gène DMD. Les promoteurs sont représentés par des flèches rouges pleines et les codons initiateurs par des flèches rouges pointillées. Les domaines structuraux successifs qu'ils encodent (cf. section 1.1.2.2) sont représentés par des flèches noires au dessus des exons.

### 1.1.1.2 Les transcrits alternatifs

La dystrophine possède 7 promoteurs spécifiques de tissus dont trois promoteurs codant pour des dystrophines pleine longueur. Ces trois promoteurs situés en amont du premier exon sont exprimés dans différents tissus : le promoteur M (Dp427m) dans les muscles striés et cardiaques [Koenig 1989], le promoteur C (Dp427c) principalement dans le cerveau [Nudel 1989, Chelly 1990], le promoteur P (Dp427p) principalement dans les cellules de Purkinje mais aussi dans le muscle squelettique [Chelly 1990]. Seul l'exon 1 diffère entre ces trois promoteurs. Un quatrième promoteur a été identifié par [Nishio 1994] : le promoteur L (Dp427l) mais des travaux plus récents indiquent que ce promoteur est artificiel [Whewey 2003], (Tableau 1.1).

Quatre promoteurs internes ont également été décrits et nommés en fonction de la taille de la protéine produite : Dp260 dans la rétine mais aussi dans le coeur et le cerveau [D'Souza 1995] ; Dp140 dans le cerveau, la rétine et les reins [Lidov 1995] ; Dp116 dans les cellules de Schwann [Byers 1993] et Dp71 dans toutes les cellules sauf les muscles squelettiques [Hugnot 1992].

Promoteur	Référence	Spécificité tissulaire	Transcrit alternatif
Dp427m	[Koenig 1989]	Muscle squelettique Coeur Cellules gliales	non
Dp427c	[Nudel 1989] [Chelly 1990]	Cerveau Rétine	non
Dp427p	[Chelly 1990]	Cellules de Purkinje Muscle squelettique	non
Dp260	[D'Souza 1995]	Rétine	non mais 2 exons 1 différents
Dp140	[Lidov 1995]	Cerveau Rein Rétine	Exon 71 & 78 -> dp140ab Exon 78 -> dp140b Exon 71 to 74 & 78-> dp140bc Exon 71 to 74 -> dp140c
Dp116	[Byers 1993]	Cellule de Schwann	non
Dp 71	[Hugnot 1992]	Ubiquitaire sauf muscles squelettiques	Exon 71 -> dp71a Exon 71 & 78 -> dp71ab Exon 78 -> dp71b Stop à l'exon 70 -> dp40

Tableau 1.1 – *Promoteurs et transcrits alternatifs du gène DMD.*

## 1.1.2 La protéine

### 1.1.2.1 Rappels sur la structure des protéines

#### Les propriétés des acides aminés

Les acides aminés naturels composant les protéines sont composés de deux groupements communs (carboxyle  $-\text{COOH}$  et amine  $-\text{NH}_2$ ) et d'une chaîne latérale qui diffère suivant les acides aminés. Le carbone  $\alpha$  est le carbone central qui relie ces trois groupements. La chaîne latérale confère des propriétés chimiques spécifiques aux acides aminés (hydrophobicité, charge). Ainsi, les acides aminés sont classés en fonction de leurs propriétés physico-chimiques (figure 1.3).

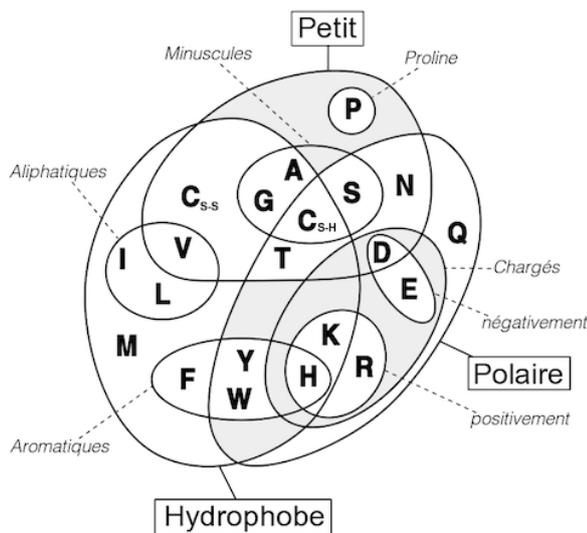
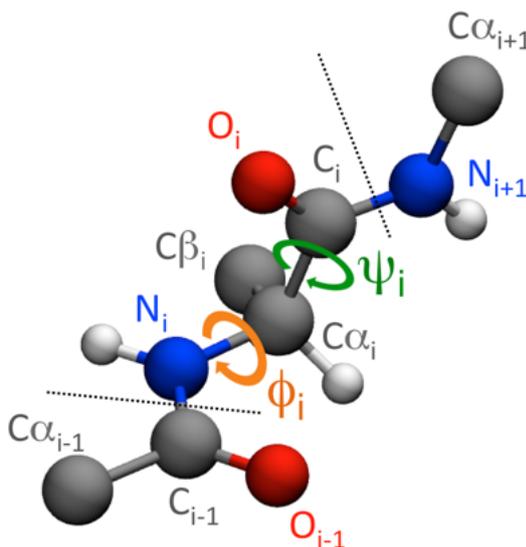


FIGURE 1.3 – Diagramme de Venn représentant les propriétés physico-chimiques des acides aminés.

### Les angles de torsion

Les acides aminés formant la chaîne polypeptidique sont liés par des liaisons peptidiques entre l'extrémité carboxyle d'un acide aminé et l'extrémité amine de son voisin. Ces liaisons peptidiques sont confinées dans un plan mais une certaine liberté de rotation existe autour des carbones  $\alpha$ . Les angles de torsion phi  $\phi$  et psi  $\psi$  sont définis respectivement entre  $N_i-C_i^\alpha$  et  $C_i^\alpha-C_i$  (figure 1.4).

FIGURE 1.4 – Les angles de torsion  $\phi$  et  $\psi$  qui définissent la conformation du squelette d'un acide aminé.

### Les niveaux de structuration des protéines

La structure des protéines présente quatre niveaux de complexité (figure 1.5). La structure primaire est définie par la séquence des acides aminés. La structure secondaire est

représentée par les repliements locaux de la séquence peptidique de la protéine définis par les angles de torsion. Les principaux repliements locaux forment des hélices  $\alpha$  et des brins  $\beta$  généralement reliés par des boucles. La structure tertiaire ou tridimensionnelle est représentée par le repliement de la chaîne polypeptidique dans l'espace sous l'influence des interactions existantes entre les éléments de structure secondaire. La structure tertiaire est la structure native des protéines, c'est-à-dire la forme sous laquelle la protéine est, en général, fonctionnelle. Enfin, la structure quaternaire décrit l'assemblage des protéines composées de plusieurs sous-domaines. Le repliement des protéines dépend de la composition en acides aminés de la séquence mais aussi de cinq grands types d'interactions : les liaisons hydrogènes, les interactions électrostatiques et hydrophobes, les forces de Van der Waals (dipôle temporaire de faible force) et les ponts disulfures entre les cystéines.

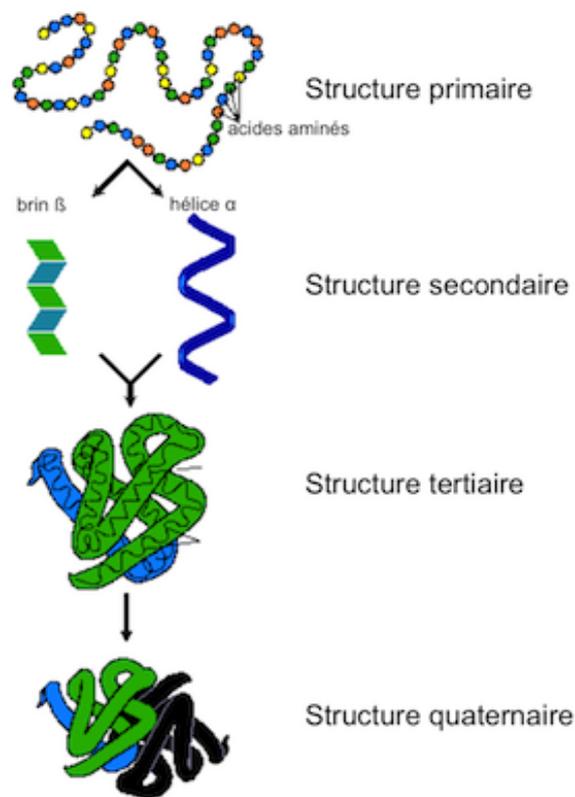


FIGURE 1.5 – Les quatre niveaux de complexité de la structure d'une protéine.

### 1.1.2.2 Structure de la dystrophine

La dystrophine appartient à la famille des spectrines caractérisées par un domaine N-terminal liant l'actine et par la présence de motifs répétés homologues à la spectrine appelés répétitions (voir section 1.1.2.5). Dans la littérature, la dystrophine est généralement décrite comme une protéine filamenteuse composée de quatre grands domaines : le domaine de liaison à l'actine (ABD), le domaine central avec ses 24 répétitions, la région riche en cystéine et le domaine C-terminal [Koenig 1988, Blake 2002]. Cependant, cette description de la dystrophine est peu détaillée et ne permet pas de distinguer les domaines structuraux

des domaines de liaison (figure 1.6).



FIGURE 1.6 – Schéma des domaines structuraux de la dystrophine.

### Le domaine ABD1

Le premier domaine de liaison à l'actine (ABD1) est composé de deux domaines homologues à la calponine : CH1 et CH2 réunis par une hélice  $\alpha$  (figure 1.7). Ce domaine a été cristallisé et sa structure a été résolue par diffraction aux rayons X [Norwood 2000, Banuelos 1998] et est répertoriée dans la Protein Data Bank ([www.pdb.org](http://www.pdb.org)) (PDB : 1DXX). Les domaines homologues à la calponine sont généralement présents par paire dans les protéines du cytosquelette et les protéines de transduction du signal [Castresana 1995, Gimona 1998].

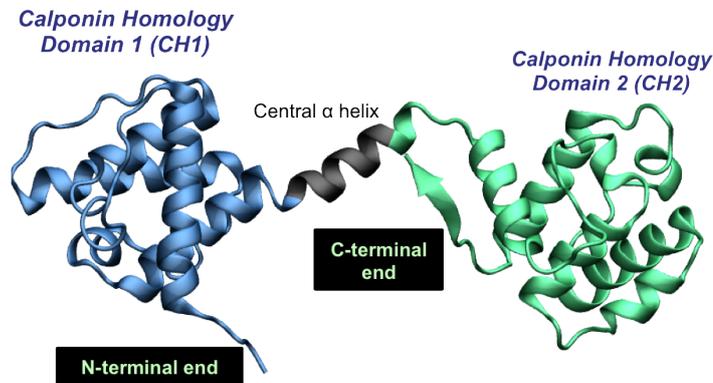


FIGURE 1.7 – Structure tridimensionnelle des domaines homologues à la calponine CH1 (bleu) et CH2 (vert) de la dystrophine reliés par une hélice (noir) [Norwood 2000] constituant le domaine de liaison à l'actine ABD1.

### Le domaine central

Le domaine central représente 76% de la séquence en acides aminés de la dystrophine et est composé de 24 répétitions homologues à la spectrine (R1 à R24) et de 4 charnières (hinges : H1 à H4) [Koenig 1988, Winder 1995a] (figure 1.8). Ces 4 charnières divisent le domaine central en trois sous-domaines (R1 à R3, R4 à R19, R20 à R24) [Koenig 1988].

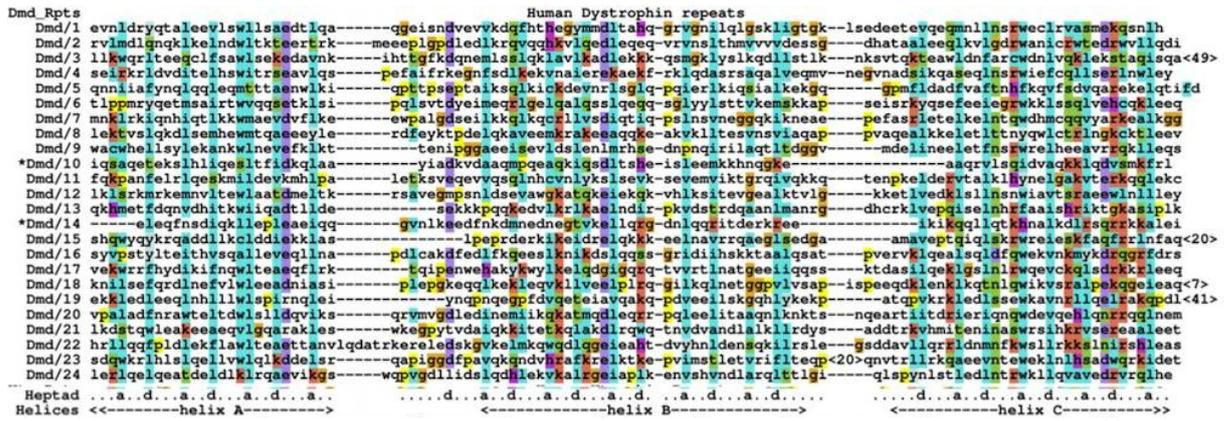


FIGURE 1.8 – *Alignement des séquences primaires des répétitions de la dystrophine proposé par Winder [Winder 1995a]. Chaque ligne représente une répétition. Le positionnement des résidus a et d des heptades est indiqué en-dessous ainsi que le positionnement relatif des hélices.*

Les séquences primaires des 24 répétitions sont très peu identiques entre elles mais sont toutes caractérisées par la succession de motifs en heptades. Une heptade est définie par sept acides aminés numérotés de a à g où les résidus en position a et d sont hydrophobes (figure 1.9). En se basant sur les structures 3D obtenues de répétitions de spectrine et sur les données de dichroïsme circulaire obtenues sur la dystrophine [Le Rumeur 2003, Legardinier 2009a, Mirza 2010], il est admis que ces répétitions d’heptades sont la base de la structure en faisceau de trois hélices (coiled-coil) des répétitions de la dystrophine.

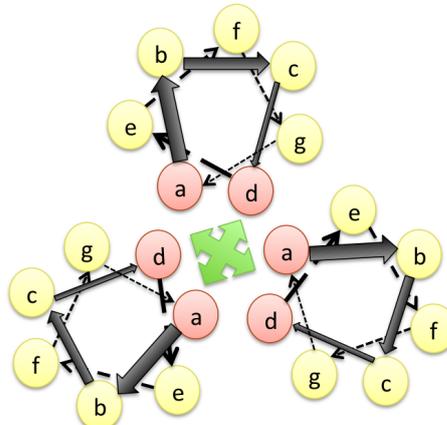


FIGURE 1.9 – *Schéma de l’organisation des heptades d’un faisceau de trois hélices. Les acides aminés a et d, en rouge, participent à des interactions hydrophobes (flèches vertes) qui maintiennent la structure en faisceau d’hélices.*

Ainsi, chaque répétition est composée de 3 hélices  $\alpha$  reliées par deux boucles, le tout organisé en faisceau de trois hélices. Ces répétitions s’enchaînent et sont liées par une hélice commune formée par la troisième hélice de la première répétition et la première hélice de la suivante (figure 1.10) et dont la partie centrale sera appelée "linker".

Jusqu’à présent, aucune équipe n’a pu obtenir une structure tridimensionnelle expérimentale de la dystrophine fragmentée ou entière, ni par cristallographie puis diffusion aux

rayons X, ni par RMN<sup>1</sup>. Il semblerait que la structure répétée du domaine central et sa flexibilité seraient un frein à l'obtention de structures expérimentales. Cependant, étudier la structure de la dystrophine permettrait de mieux comprendre sa fonction. Ainsi, nous avons décidé d'utiliser la modélisation par homologie pour étudier la structure du domaine central et cette approche est détaillée dans le chapitre 2. D'autres approches alternatives sont également utilisées au sein de l'équipe : la diffusion de rayons X aux petits angles (SAXS) et la cryo-microscopie.

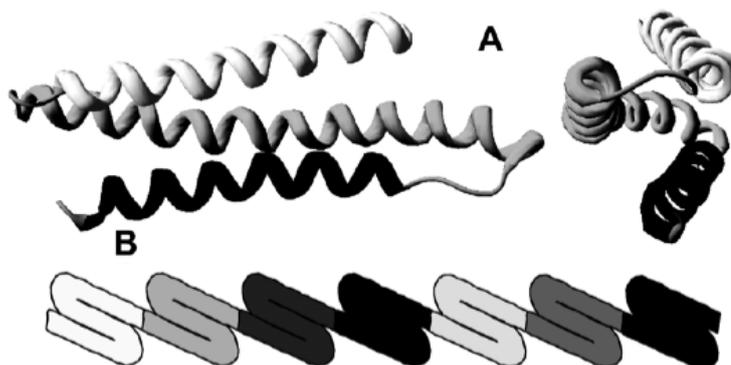


FIGURE 1.10 – Organisation structurale des répétitions homologues à la spectrine de la dystrophine [Saadat 2006]. A. Structure RMN d'une répétition de la spectrine (PDB : 1AJ3). B. Enchaînement de plusieurs répétitions en tandems qui constituent le coeur du domaine central.

### Le domaine riche en cystéines

Le troisième grand domaine structural de la dystrophine est le domaine riche en cystéines qui comprend quatre régions : un domaine WW, deux domaines EF et un domaine ZZ.

Le domaine WW est caractérisé par deux tryptophanes séparés par 21 acides aminés, ce domaine est présent dans de nombreuses protéines régulatrices et de signalisation [Bork 1994, Sudol 1996, Huang 2000].

Le domaine WW est suivi d'une hélice  $\alpha$  et de deux pseudo mains EF. Les mains EF sont des motifs hélice-boucle-hélice présents dans de nombreuses protéines liant le calcium [Ikura 1996]. Cependant le motif liant spécifiquement le calcium n'est pas présent dans la dystrophine et il semblerait donc que les mains EF de la dystrophine n'aient pas de propriété de liaison au calcium [Huang 2000]. L'ensemble de ces trois domaines a été cristallisé et une structure par rayons X a été obtenue (PDB : 1EG3) (figure 1.11).

Les mains EF sont suivies d'un domaine en doigt de zinc (ZZ). Ce type de domaine est présent dans de nombreuses protéines liant les cations divalents tels que  $Zn^{2+}$ , qu'elles soient nucléaires ou cytoplasmiques [Ponting 1996, Hnia 2007].

1. Les structures des répétitions 1 de la dystrophine et de l'utrophine viennent de paraître au moment de la rédaction de cette thèse [Muthu 2012].

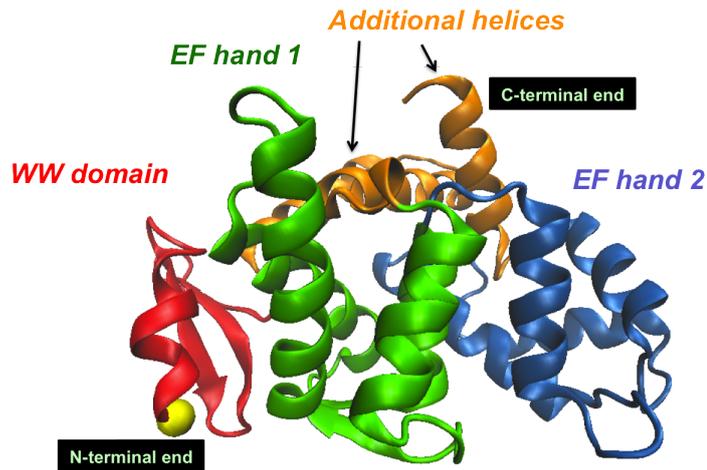


FIGURE 1.11 – Structure tridimensionnelle des domaines WW et mains EF de la dystrophine [Huang 2000].

### Le domaine C-terminal

Le dernier domaine structural de la dystrophine est le domaine C-terminal. Ce domaine est majoritairement composé de faisceaux d’hélices similaires à ceux présents dans le domaine central de la dystrophine [Blake 1995].

### 1.1.2.3 Domaines fonctionnels de la dystrophine

La structure filamenteuse de la dystrophine ainsi que ces nombreux domaines de liaison confère à la protéine un rôle charnière au sein des cellules musculaires (figure 1.12). En effet, des études ont montré qu’elle peut lier les lipides du sarcolemme et 11 protéines détaillées ci-dessous.

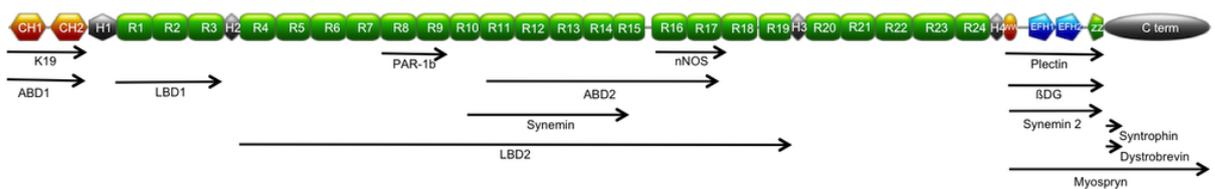


FIGURE 1.12 – Domaines de liaison de la dystrophine

Un des premiers domaines de liaison caractérisé est le domaine de liaison à l’actine en N-terminal : ABD1 [Koenig 1988] formé par deux domaines homologues à la calponine (CH1 et CH2). La liaison aux filaments d’actine se fait par l’intermédiaire d’un sillon hydrophobe identifié sur les surfaces de CH1 et CH2. Le domaine CH1 a une forte affinité pour l’actine alors que le CH2 améliore cette affinité mais n’est pas capable de lier seul l’actine [Norwood 2000, Banuelos 1998]. L’ABD1 lie également la kératine, préférentiellement la kératine 19 mais il peut aussi lier la kératine 8 [Stone 2005].

La présence de nombreux domaines de liaison dans le domaine central suggère qu'il n'a pas qu'un rôle d'absorbeur de chocs comme il a été décrit précédemment [Davies 2006]. Les plus grands domaines de liaison sont ceux assurant la liaison à la membrane plasmique des cellules musculaires, ou sarcolemme. En effet, deux domaines de liaison aux lipides ont été identifiés : le premier, de la première répétition à la troisième (LBD1) ; le deuxième, de la répétition 4 à la répétition 19 (LBD2) [Le Rumeur 2003, Legardinier 2008a, Legardinier 2009a, Vie 2010, Sarkis 2011]. Des interactions hydrophobes et électrostatiques maintiennent la protéine en contact avec le sarcolemme, proposant ainsi un nouveau modèle d'interaction de la dystrophine.

Un second domaine de liaison à l'actine, ABD2, a été identifié dans le domaine central incluant les répétitions 11 à 17. Ce domaine lie l'actine F par l'intermédiaire d'interactions électrostatiques et semble avoir moins d'affinité pour l'actine que l'ABD1 [Amann 1998, Rybakova 2006, Sarkis 2012].

La sérine/thréonine kinase Par1b interagit avec les répétitions 8 et 9 de la dystrophine en les phosphorylant [Yamashita 2010].

Le domaine central de la dystrophine peut lier les filaments intermédiaires, spécifiquement les synémines  $\alpha$  et  $\beta$  par l'intermédiaire des répétitions 11 à 14 [Bhosle 2006]. Les synémines sont de grandes protéines s'associant à la desmine et/ou à la vimentine qui font également partie de la superfamille des filaments intermédiaires.

Récemment, les répétitions 16 et 17 du domaine central ont été identifiées comme partenaires de l'Oxyde Nitrique Synthétase neuronale (nNOS) [Lai 2009]. Ainsi, conjointement à la syntrophine, la dystrophine maintient la nNOS près du sarcolemme. Cette protéine enzymatique joue un rôle clé dans l'activité musculaire notamment pendant la contraction, la régénération et l'atrophie des cellules musculaires en augmentant l'entrée du glucose dans les cellules et la perfusion du muscle [Stamler 2001, Bredt 1999, Gentil 2012].

Le domaine riche en cystéines de la dystrophine est une région impliquée dans l'interaction de plusieurs protéines notamment des protéines formant le dystrophine-glycoprotéine complexe (DGC), la dystrophine faisant elle-même partie de ce complexe. Le DGC permet d'ancrer plusieurs protéines du cytoplasme et de la matrice extracellulaire à la membrane [Winder 2001, Ervasti 2006]. La première interaction avec le DGC décrite est celle du  $\beta$ -dystroglycane ( $\beta$ DG) [Huang 2000, Hnia 2007, Ishikawa-Sakurai 2004]. Le  $\beta$ DG est une sous-unité du dystroglycane, une protéine transmembranaire. La syntrophine et la dystrobrevine se lient à la dystrophine par l'intermédiaire du domaine C-terminal [Sadoulet-Puccio 1997, Newey 2000].

Des protéines n'appartenant pas au DGC se lient également au domaine riche en cystéines comme la synémine  $\alpha$ , précédemment décrite [Bhosle 2006] et la plectine. La plectine est une très grande protéine possédant des domaines de liaison aux filaments d'actine, aux filaments intermédiaires et aux microtubules [Rezniczek 2007].

Enfin la myosprine, une protéine kinase A spécifique du muscle, se lie au domaine riche en cystéine et au domaine C-terminal de la dystrophine [Reynolds 2008].

#### 1.1.2.4 Rôle physiologique et localisation de la dystrophine

La dystrophine est localisée sous la membrane des muscles, sur la face interne du sarcolemme (figure 1.13) et fait partie d'un grand complexe appelé dystrophine-glycoprotéine complexe (DGC). Elle assure un rôle de maintien de l'architecture cellulaire lors des cycles de contraction – relaxation musculaires. Cependant, les mécanismes moléculaires mis en jeu sont encore méconnus et plusieurs hypothèses font référence à sa structure filamenteuse et à ses nombreux domaines de liaison.

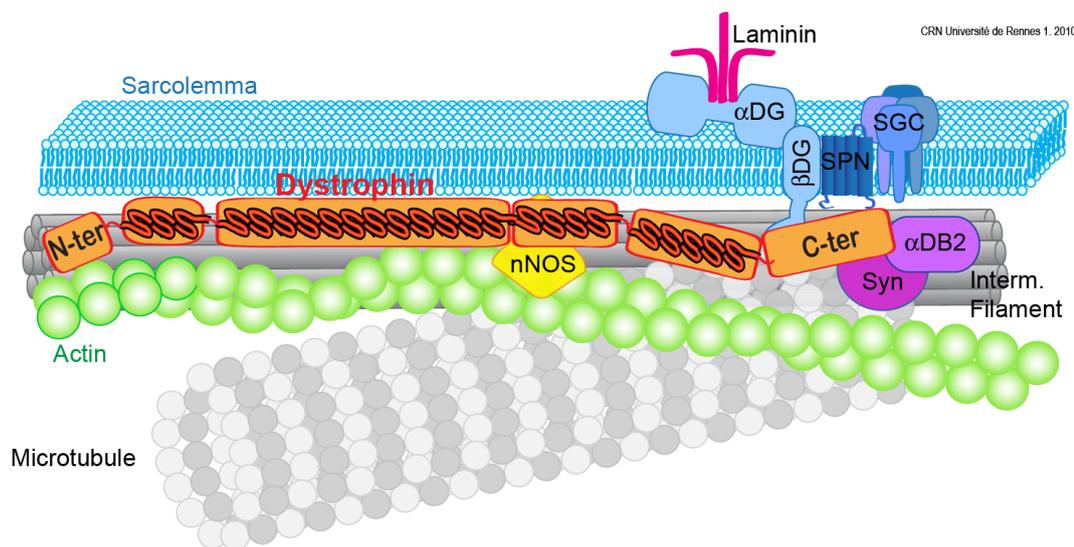


FIGURE 1.13 – Schéma des interactions de la dystrophine. La dystrophine (orange) est située sous le sarcolemme (bleu).

#### 1.1.2.5 La famille des protéines homologues à la dystrophine

En 1988, les premières analyses de la séquence de la dystrophine montrèrent la présence de répétitions similaires à celles de la spectrine [Davison 1988]. La dystrophine est, depuis, classée dans la famille des spectrines. Les protéines de la famille des spectrines sont caractérisées par la présence de répétitions homologues à la spectrine, des domaines de liaisons à l'actine et des mains EF. Dans cette famille, on retrouve l' $\alpha$ -actinine, les spectrines  $\alpha$  et  $\beta$ , "les dystrophines" (dystrophine, utrophine et DRP2), les dystrobrevines  $\alpha$ ,  $\beta$  et  $\gamma$  et la dystrotéline [Broderick 2005, Jin 2007, Roberts 2001]. L'analyse de leurs séquences protéiques a montré qu'elles sont dérivées d'une protéine ancestrale similaire à l' $\alpha$ -actinine qui a évolué suite à une série de duplications et de réarrangements génétiques [Byers 1989, Baines 2003]. Ainsi l' $\alpha$ -actinine ancestrale aurait dérivé d'un côté (1) en une  $\alpha$ -actinine allongée, ancêtre des spectrines  $\alpha$  et  $\beta$  et de l'autre côté (2) en une protéine ancêtre des dystrophines, dystrobrevines et dystrotélines (figure 1.14). Le point de divergence entre les dystrotélines,

dystrobrevines et dystrophines reste inconnu.

La branche des dystrophines est caractérisée par une élongation du domaine N-terminal, l'ajout de répétitions supplémentaires et d'un domaine WW. Chez les vertébrés, on distingue deux paralogues de la dystrophine : DRP2, la plus petite des trois, exprimée presque uniquement dans le système nerveux central [Roberts 2000] et l'utrophine, le plus proche paralogue de la dystrophine.

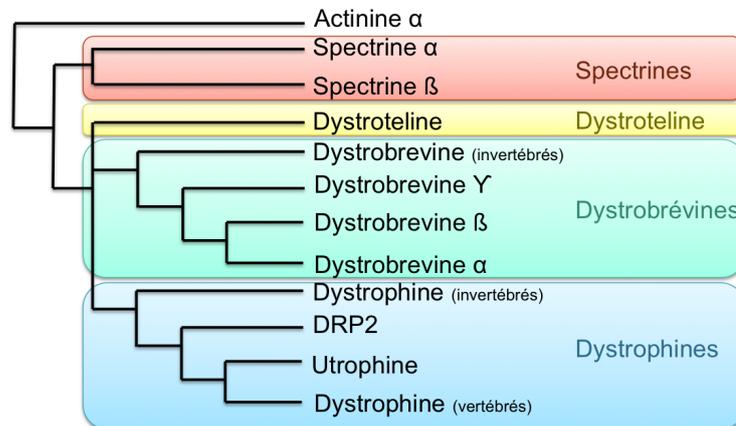


FIGURE 1.14 – Arbre phylogénétique de la famille des spectrines. Adapté de [Jin 2007] et [Roberts 2001].

L'utrophine étant le plus proche homologue de la dystrophine, elle fait l'objet de tests thérapeutiques de surexpression pour compenser l'absence de dystrophine dans les cellules des patients DMD (cf. 1.2.5.5). L'utrophine est cependant moins étudiée que la dystrophine et son rôle est encore plus méconnu. Chez l'humain, son gène est situé sur le chromosome 6 et possède deux promoteurs (A et B) ; le promoteur A est responsable de l'expression musculaire alors que le promoteur B est responsable de l'expression dans les cellules endothéliales. L'utrophine totale est composée de 3433 acides aminés et a une masse moléculaire de 395 kDa. L'expression de l'utrophine est ubiquitaire mais son expression musculaire est différente de celle de la dystrophine. En effet, l'utrophine est exprimée au niveau du sarcolemme uniquement pendant le développement foetal. Chez l'adulte, elle est exprimée au niveau des jonctions neuromusculaires et myotendineuses [Clerk 1993, Tomé 1994]. L'utrophine est observée dans les muscles en régénération ainsi que dans les muscles de patients DMD [Pons 1993, Khurana 1991].

La structure de l'utrophine présente une forte similarité avec celle de la dystrophine ; elle possède les mêmes domaines structuraux mais son domaine central ne comporte que 22 répétitions [Tinsley 1992] (figure 1.15). L'utrophine et la dystrophine partagent de nombreux domaines de liaison mais leurs modes d'interaction peuvent différer. Par exemple, les modes de liaison de l'utrophine avec le  $\beta$ DG et l'actine F diffère de celui de la dystrophine [Ishikawa-Sakurai 2004]. En effet, la dystrophine lie l'actine par l'intermédiaire de deux domaines distincts alors que l'utrophine lie l'actine par un domaine continu [Amann 1998, Rybakova 2005]. Cette différence peut suggérer que les deux protéines n'ont pas le même rôle,

la dystrophine aurait un rôle d'absorbeur de choc et l'utrophine serait comme un "guide" (molecular ruler) qui aiderait à définir la longueur des costamères pour les filaments d'actine pendant le développement musculaire [Fairclough 2011]. Un des domaines de liaison que l'utrophine ne partage pas avec la dystrophine est le domaine de liaison à la nNOS. En effet, l'utrophine ne permet pas d'assurer l'association de la nNOS à la membrane [Li 2010]. Ses propriétés d'interactions avec les lipides du sarcolemme restent encore inconnues.

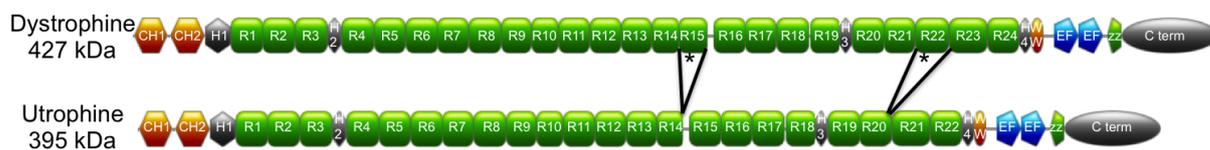


FIGURE 1.15 – *Comparaison des domaines structuraux de la dystrophine et de l'utrophine. Les étoiles représentent les répétitions de la dystrophine manquantes dans l'utrophine.*

## 1.2 Les dystrophinopathies

### 1.2.1 Aspects historiques

Les premiers descriptions de cas de myopathie de Duchenne (DMD) remontent à la première moitié du 19<sup>ème</sup> siècle (pour une revue historique cf. [Tyler 2003]). Giovanni Semmola aurait décrit le premier cas de myopathie de Duchenne en 1829 mais son article original ayant été perdu, c'est à Charles Bell qu'est attribuée la première description en 1830. Ces premières descriptions succinctes rapportent une fatigue progressive débutant dans la petite enfance et touchant d'abord les membres inférieurs sans symptômes nerveux indiquant qu'il n'y a pas d'atteinte de la moelle épinière.

Il faudra attendre 1847 et Richard Partridge pour obtenir un premier examen pathologique à l'occasion d'un séminaire à la Pathological Society of London. Il examina deux frères présentant les mêmes symptômes de faiblesse commençant par les bras et les jambes. Il releva que leurs soeurs n'étaient pas atteintes et fut le premier à examiner les muscles (à l'oeil nu). Cependant, la description qu'il fait de son patient est tellement brève qu'il est pratiquement impossible d'être sûr qu'il s'agit bien d'un cas de DMD.

C'est en 1851 que le premier examen clinique et pathologique non ambigu de DMD est décrit dans l'article d'Edward Meryon sous le nom de "progressive muscular weakness". Son premier patient était issu d'une famille de dix enfants dont les quatre garçons présentaient une faiblesse musculaire progressive. Meryon présenta dans son article la progression de la maladie chez le frère aîné. Le patient présenta des signes de faiblesse musculaire dès le plus jeune âge, perdit la marche à onze ans et décéda à 16 ans d'une insuffisance respiratoire. Meryon fut le premier à faire une autopsie sur un patient DMD et à examiner les fibres musculaires au microscope à une époque où la microscopie n'en était qu'à ses débuts. Il remarqua que les fibres musculaires étaient détruites et que des "globules d'huiles et de la matière granulaire" les remplaçait. Meryon décrivit également le cas de ses frères et d'autres fratries dans de nombreux articles jusqu'en 1870. Cependant, tous les cas qu'il décrit comme une faiblesse musculaire progressive ne sont pas des DMD. En effet, il décrivit également le premier cas de myopathie de Becker (BMD) et des cas de myopathie des ceintures. Cette hétérogénéité de patients est une des raisons pour lesquelles Duchenne accusera plus tard Meryon de ne pas avoir décrit les premiers cas de myopathie de Duchenne mais simplement des cas d'atrophie musculaire progressive.

Duchenne examina son premier cas de DMD en 1858 et le décrivit en 1861 dans son livre sous le nom de Paraplégie hypertrophique de l'enfance de cause cérébrale puis Paralysie pseudo-hypertrophique musculaire. Malgré la ressemblance avec les cas décrits par Meryon,

Duchenne insista sur le fait que la pathologie qu'il décrit est différente. Duchenne obtint la priorité de découverte de cette myopathie malgré les contestations de Meryon.

Grâce à l'emporte-pièce histologique qu'il fabriqua en 1865, Duchenne put réaliser des biopsies plus sûres et moins invasives qu'auparavant. Duchenne fut le premier à décrire les principales caractéristiques cliniques de la DMD :

- Faiblesse musculaire au début de la maladie, généralement dans les membres inférieurs
- Lordose et marche dandinante
- Hypertrophie des muscles (second stade)
- Aggravation progressive et généralisation de la maladie (troisième stade)
- Diminution ou absence de contractions musculaires aux stimuli électriques (stade avancé)
- Absence de fièvre, de troubles sensoriels, de problèmes de vessie et intestinaux durant la maladie

R. Gower contribua également significativement à établir les bases cliniques de la myopathie de Duchenne. En 1874, il examina ses premiers patients. Il s'intéressa à la transmission de la maladie et remarqua la présence d'antécédents familiaux uniquement du côté maternel. Il émit alors l'hypothèse que la maladie était transmise par l'ovule. Cependant, ce que l'on retiendra de lui est la description de la méthode particulière qu'utilisent les patients pour se relever du sol, connue aujourd'hui sous le nom de signes de Gowers.

## 1.2.2 La myopathie de Duchenne

### 1.2.2.1 Caractéristiques cliniques

La myopathie de Duchenne affecte une naissance mâle sur 3500. C'est la maladie neuromusculaire la plus fréquente chez l'enfant et la plus sévère [Brooke 1989]. Les premiers symptômes apparaissant vers 3 ans sont un retard psychomoteur, une démarche anormale et un taux de créatine kinase sérique élevé dans les muscles [Desguerre 2009]. Sans historique familial, la maladie est diagnostiquée aux environs de 5 ans [Zalaudek 1999]. Un patient est considéré comme myopathe de Duchenne s'il perd la marche avant 12 ans.

Il existe des variations dans l'expression de la maladie entre les patients mais cette hétérogénéité est moindre comparée à celle de la myopathie de Becker [Desguerre 2009]. Les patients DMD acquièrent généralement la marche plus tard que la normale (à 18 mois en moyenne) ; la moitié d'entre eux n'a jamais pu courir et un tiers n'a jamais pu monter les escaliers seul. Ils perdent la capacité à se relever du sol aux environs de 8 ans. Tous les patients développent une scoliose entre 9 et 16 ans et ont une fonction respiratoire décroissante à partir de 10 ans [Desguerre 2009]. L'atteinte cardiaque des myopathes de Duchenne

est souvent asymptomatique jusqu'à 15 ans [Connuck 2008]. Les patients DMD ont également des problèmes nutritionnels et 40% des patients présentent une obésité avant 12 ans [Willig 1993] alors que 44% des plus de 18 ans sont dénutris [Mok 2006]. Une étude d'Eagle [Eagle 2007] estime l'âge moyen de décès à 27 ans, soit dix ans de plus que les premiers patients décrits au 19<sup>ème</sup> siècle [Jeppesen 2003].

Il n'existe actuellement aucun traitement curatif de la myopathie de Duchenne mais la prise en charge des patients permet sans doute d'améliorer leur qualité et leur espérance de vie [Manzur 2009]. La kinésithérapie prévient les déformations et retarde l'apparition d'une scoliose. Une arthrodesse vertébrale peut être pratiquée et permet d'améliorer la durée et la qualité de vie des patients [Bridwell 1999]. L'insuffisance respiratoire est améliorée par l'usage d'une ventilation non invasive. La cardiomyopathie est traitée par l'usage de bêta-bloquants et d'inhibiteurs de l'enzyme de conversion de l'angiotensine, ce qui améliore les symptômes [Bushby 2003].

### 1.2.2.2 Caractéristiques histologiques

Les biopsies musculaires des patients atteints de myopathie de Duchenne montrent des amas de fibres musculaires nécrosées ou dégénérées. Ces fibres musculaires sont souvent entourées de macrophages et de lymphocytes CD4+. De petites fibres immatures avec un noyau central sont également observées indiquant une régénération des myoblastes (figure 1.16) [Schmalbruch 1984, McDouall 1990, Deconinck 2007].

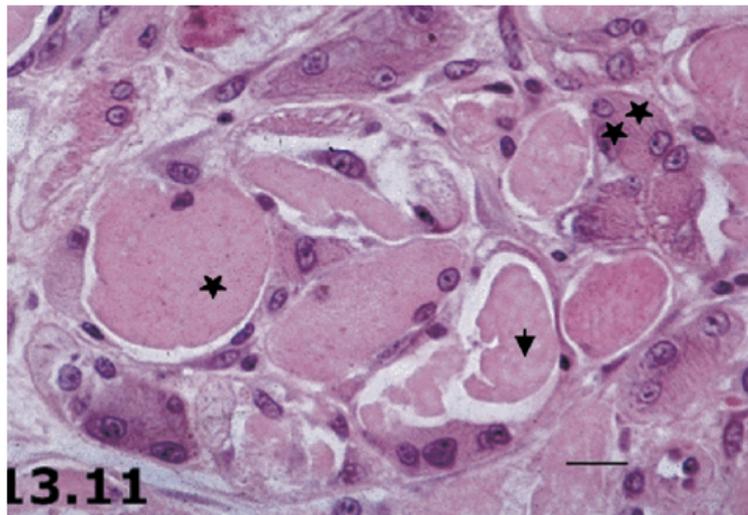


FIGURE 1.16 – *Biopsie musculaire d'un patient atteint de myopathie de Duchenne. Les fibres musculaires dégénérées ou nécrotiques (flèches) sont souvent observées en clusters entourées par des macrophages et des lymphocytes CD4+ (étoiles). De petites fibres immatures avec un noyau central sont présentes (étoile double) [Deconinck 2007].*

Les causes de cette modification des cellules musculaires des patients sont encore méconnues. Cependant plusieurs hypothèses, maintenant admises, permettent d'en comprendre

certain aspects.

La dystrophine étant un élément structural important dans les cellules musculaires, son absence entraîne une fragilité structurelle et une accumulation cytoplasmique de protéines normalement absentes des fibres musculaires [Pestronk 1982, Straub 1997]. En effet, l'intégrité de la membrane semble compromise, particulièrement pendant les contractions soutenues [Head 1994]. L'absence de dystrophine entraîne une délocalisation des protéines associées à la dystrophine de la membrane et une déstabilisation des costamères, conduisant à une fragilité de la membrane [Rybakova 2000]. Il a été démontré que les fibres de petit diamètre sont moins nécrosées que les fibres musculaires de grand diamètre [Boland 1995]. De plus, la dégénérescence continue des muscles nécessite leur régénération quasi-permanente épuisant les stocks de cellules satellites nécessaires [Heslop 2000, Snow 1978]. Une fois ces cellules épuisées, les patients perdent leur capacité de régénération et la dégénérescence musculaire ne peut être enrayée [Price 2007].

Le calcium aurait également un rôle important dans la nécrose des cellules musculaires. En effet, une accumulation de calcium a été observée dans des fibres musculaires de patients DMD [Bodensteiner 1978] et un influx massif de calcium a été démontré à travers des membranes déficientes en dystrophine [De Backer 2002]. Dans un premier temps, le mécanisme d'homéostasie calcique compense cet apport de calcium mais lorsque des micro-lésions apparaissent dans le sarcolemme, l'homéostasie ne suffit plus et la concentration intra-cellulaire de calcium augmente [Tinsley 1998]. Des protéases, dont les calpaïnes, sont alors recrutées, détruisant les protéines membranaires et conduisant à une entrée massive de  $Ca^{2+}$  et à la mort cellulaire.

## 1.2.3 La myopathie de Becker

### 1.2.3.1 Caractéristiques cliniques

La myopathie de Becker est la variante allélique de la myopathie de Duchenne. Elle est caractérisée par des symptômes similaires mais plus légers apparaissant plus tardivement que dans la DMD. Les études épidémiologiques estiment le ratio des dystrophinopathies à 1 BMD pour 3 DMD [Bushby 1991, Hughes 1996, Mostacciolo 1987]. Le phénotype clinique de la myopathie de Becker est très hétérogène comparé à la myopathie de Duchenne et les patients perdant la marche après 16 ans sont classés comme myopathes de Becker. Les premiers symptômes apparaissent à 12 ans en moyenne mais cela peut varier considérablement d'un patient à l'autre. La perte de la marche est également variable, les patients les plus gravement atteints sont en fauteuil roulant à partir de 16 ans et décèdent entre 40 et 60 ans [Emery 2002]. Certains patients ne présentent aucun symptôme malgré une mutation en

phase et sont considérés comme des BMD asymptomatiques.

Presque tous les myopathes de Becker développent une atteinte cardiaque au cours de leur vie [Finsterer 2008]. Cette atteinte peut être asymptomatique et détectable uniquement avec des examens cliniques, ou symptomatique (cardiomyopathie dilatée). Environ un tiers des atteintes asymptomatiques se transforment en atteinte cardiaque symptomatique à un âge très variable mais le plus souvent dans la trentaine [Finsterer 2003]. L'âge d'apparition des symptômes cardiaques peut précéder de plusieurs années l'affection des muscles squelettiques ce qui rend le diagnostic difficile [Yokota 2004]. Une fois que la cardiomyopathie est déclarée, elle peut rapidement évoluer et conduire au décès du patient si elle n'est pas traitée [Finsterer 1999]. Il est donc important de procéder à un examen cardiaque complet dès que le diagnostic de BMD est posé et de le refaire régulièrement. Si l'atteinte cardiaque est dépistée tôt, il est possible de proposer une thérapie adéquate par l'utilisation du même type de médicaments que ceux utilisés dans le cas de DMD et/ou par des greffes cardiaques, améliorant ainsi le pronostic vital [Doing 2002].

### 1.2.3.2 Caractéristiques histologiques

Les biopsies musculaires des patients atteints de myopathie de Becker montrent une expression plus faible que les contrôles et discontinue de la dystrophine localisée à la périphérie des fibres musculaires [Muntoni 2003] (figure 1.17). Cette expression varie entre patients ayant des mutations différentes mais également entre patients partageant la même mutation tout en restant faible. Une étude d'Anthony [Anthony 2011] sur des patients présentant des mutations dans la région hautement mutagène montre que les plus hauts niveaux d'expression de la dystrophine, du beta-dystroglycan et de la nNOS sont corrélés à des myopathies moins sévères.

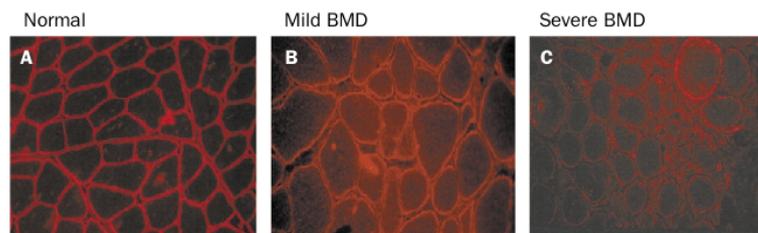


FIGURE 1.17 – Exemples de biopsies musculaires d'un muscle sain (A) et de patients atteints de myopathie de Becker (B et C). Dans le muscle sain, la dystrophine est localisée à la périphérie de chaque fibre musculaire. Chez les patients atteints de BMD, il y a une expression faible et discontinue de la dystrophine [Muntoni 2003].

Contrairement aux patients atteints de DMD, les patients atteints de BMD produisent une dystrophine mutée tronquée [Tuffery-Giraud 2009]. Cette présence de dystrophines tronquées dans des cellules musculaires de patients moins sévèrement atteints a conduit à l'hypothèse, généralement reprise dans la littérature, que les dystrophines tronquées sont par-

tiellement fonctionnelles. Cependant, la fonction de la dystrophine pleine longueur et des dystrophines tronquées reste encore largement méconnue. Selon la localisation de la mutation, différentes fonctions de la dystrophine peuvent être altérées et les caractéristiques physiopathologiques des patients BMD diffèrent donc d'une mutation à l'autre. Par exemple, il apparaît que les mutations touchant le domaine de liaison à l'actine ABD1 ou le domaine riche en cystéine sont les plus délétères [Beroud 2007, Beggs 1991]. Le rôle de maintien de l'architecture cellulaire de la dystrophine semble dans ce cas compromis par l'absence de liaison de la dystrophine aux filaments du cytosquelette ou au DGC.

### 1.2.3.3 Les mutations du gène *DMD*

Les myopathies de Duchenne et de Becker sont des maladies récessives liées à l'X. Dans 70% des cas, la mutation est transmise par la mère porteuse saine. Dans 30% des cas, il s'agit d'une mutation *de novo* [Matsuo 1996].

A cause de sa très grande taille, le taux de mutations du gène *DMD* est plus élevé que la moyenne. En effet, des milliers de mutations ont été référencées sur le gène [Aartsma-Rus 2006, Tuffery-Giraud 2009]. Les mutations les plus courantes sont des délétions intragéniques d'un ou plusieurs exons (~70%). Une région hautement mutagène (point chaud) a été identifiée entre les exons 45 et 53. Les petites mutations (petites délétions, insertions et mutations ponctuelles) représentent 20% des cas répertoriés. Plus rarement, des duplications d'un ou plusieurs exons ont été rapportées (~10%).

Les mutations sont classées dans deux catégories : les mutations en phase (in-frame en anglais) et les mutations hors-phase (out-of-frame en anglais). La séquence codante du gène présente dans les exons est lue par codon (triplet de nucléotides) lors de la traduction de l'ARN messager en protéine, chaque codon correspondant à un acide aminé. Si une délétion d'exons entraîne la suppression d'un nombre entier de codons, la traduction de la protéine se fera sans décalage du cadre de lecture (mutation en phase). La protéine produite sera écourtée et présente dans les cellules. Dans le cas contraire, c'est-à-dire lorsque les exons supprimés contiennent un codon incomplet, le cadre de lecture est décalé et la traduction de la dystrophine sera interrompue par un codon stop (mutation hors phase) et entraînera la production d'une dystrophine tronquée dans sa partie C-terminale et qui sera rapidement dégradée (figure 1.18).

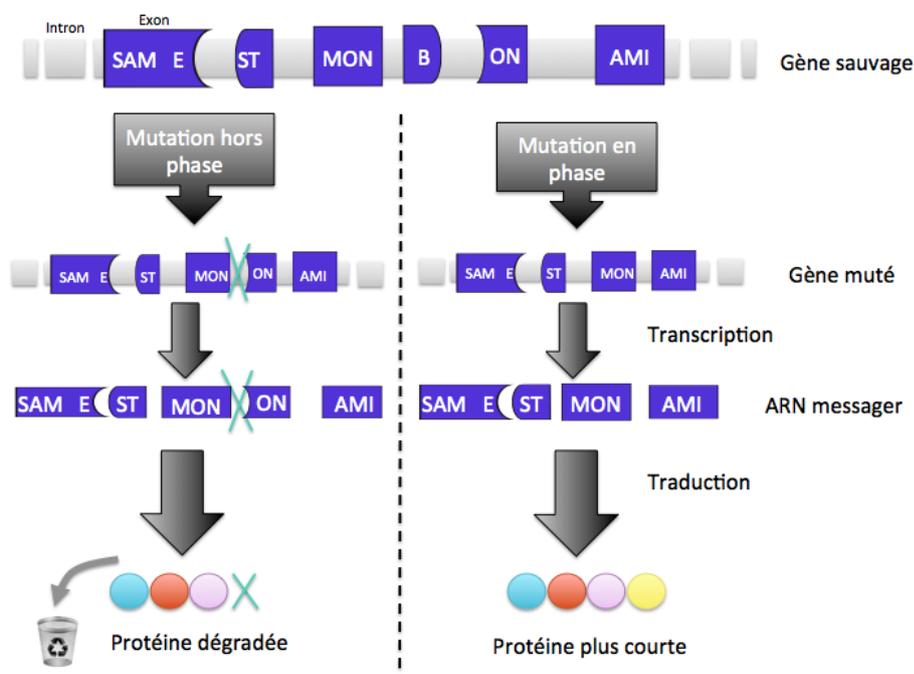


FIGURE 1.18 – Exemple de mutations hors phase et en phase (d'après [AFM 2009]).

## 1.2.4 Corrélacion génotype - phénotype ?

La première difficulté pour proposer une corrélation génotype - phénotype précise est la définition des phénotypes. Un consensus de classification des patients dans les groupes phénotypiques en fonction de l'âge de perte de la marche est utilisé. Si le patient perd la marche avant 12 ans, il a un phénotype DMD, après 16 ans, il a un phénotype BMD et entre 12 et 16 ans il est de phénotype intermédiaire (IMD) [Tuffery-Giraud 2009]. Ce consensus ne classe les patients qu'en fonction des grands groupes phénotypiques.

Cependant une règle dite "règle de Monaco" a été proposée dès 1988 et établit un lien entre les phénotypes observés et la préservation du cadre de lecture de la dystrophine [Monaco 1988]. D'un côté, les mutations décalant le cadre de lecture de la protéine font apparaître un codon STOP prématuré et conduisent à une absence de dystrophine dans les cellules et à une myopathie de Duchenne. D'un autre côté, les mutations conservant le cadre de lecture de la protéine conduisent à la production en plus faible quantité que dans les cellules saines d'une dystrophine tronquée et sont observées chez les myopathes de Becker. Cependant cette règle ne concerne que les grands groupes phénotypiques (BMD, DMD) et ne permet pas d'expliquer l'hétérogénéité des symptômes observés au sein de ces groupes. De plus, cette règle comporte des exceptions estimées entre 8 et 9% des mutations rapportées dans la Leiden Muscular Dystrophies Database [Aartsma-Rus 2006] mais peut atteindre jusqu'à 30% selon les études [Kesari 2008].

Une partie de ces exceptions peut être expliquée grâce à l'évolution des techniques de

diagnostic génétique. Les premiers diagnostics génétiques réalisés utilisaient la PCR pour rechercher la présence de certains exons sur l'ADN complémentaire ; or cette technique non exhaustive laisse un doute sur l'étendue réelle des mutations des patients [Beggs 1991]. Une délétion identifiée comme étant en phase peut, en réalité, impliquer plus d'exons que ceux détectés initialement et décaler le cadre de lecture conduisant à une absence de dystrophine caractéristique d'un phénotype DMD et inversement [Taylor 2007, Taylor 2008]. Ces dernières années, l'avancée des techniques de biologie moléculaire a permis d'identifier les mutations présentes chez les patients de façon sûre par des techniques de MLPA (Multiplex Ligation-dependant Probe Amplification), de RT-PCR de tous les exons et de séquençage d'ADN [Flanigan 2009, Kesari 2008, Deburgrave 2007]. Il est donc important de vérifier la méthode d'identification utilisée pour détecter une mutation chez un patient diagnostiqué il y a plusieurs années et, si possible, refaire des analyses.

Une autre partie de ces exceptions peut s'expliquer par des phénomènes épigénétiques. En effet, il a été observé chez des patients porteurs d'une mutation décalant le cadre de lecture, un saut d'exon "naturel" dans une partie des cellules. Dans ces cellules, le cadre de lecture de la protéine est recalé et permet la production d'une faible quantité de dystrophine tronquée conduisant à un phénotype BMD [Gurvich 2008]. Ces phénomènes sont rares, localisés à certaines cellules et difficilement prévisibles.

Enfin, certaines mutations conservant le cadre de lecture sont observées chez des patients atteints de myopathies sévères (DMD) [Muntoni 2003]. Une partie de ces mutations est observée dans les domaines de liaison à l'actine (ABD1) et au  $\beta$ -dystroglycane ( $\beta$ DG) [Beroud 2007, Tuffery-Giraud 2009]. Il semblerait donc que les mutations impliquant des modifications dans les domaines de liaison à l'actine et au  $\beta$ -dystroglycane soient plus délétères que celles modifiant le domaine central. Cependant, de nombreuses mutations conduisant aussi à un phénotype DMD ont été observées dans le domaine central. Récemment, une étude clinique montre que la troncature du domaine de liaison à la nNOS dans le domaine central délocalise la nNOS de la membrane chez certains patients et est associée à un phénotype BMD sévère [Gentil 2012].

La règle de Monaco n'est donc pas suffisante pour corrélérer de façon précise génotype et phénotype.

Récemment, des études ont été réalisées sur l'hétérogénéité clinique des patients atteints de DMD [Desguerre 2009, Humbertclaude 2012]. Ces études ont montré une hétérogénéité du phénotype DMD du point de vue des atteintes musculaires, mentales, respiratoires et cardiaques. Différents sous-groupes phénotypiques ont ainsi pu être déterminés en combinant soit, l'atteinte musculaire et respiratoire [Humbertclaude 2012], soit l'atteinte musculaire et mentale [Desguerre 2009]. De plus, une corrélation partielle entre ces sous-groupes phénotypiques et la localisation de la mutation a pu être établie concernant l'atteinte mentale.

En effet, les mutations présentes après l'exon 30 et affectant les isoformes plus courtes, sont observées chez les patients présentant une atteinte cognitive. Cependant, aucune corrélation n'a pu être réalisée entre la localisation de la mutation et l'atteinte musculaire et cardiaque. Le but principal de ces études n'est pas de proposer une corrélation génotype-phénotype mais d'aider à interpréter les résultats des essais cliniques et de guider le choix des patients à inclure dans les essais en fonction de leur atteinte clinique.

L'hétérogénéité clinique des patients BMD est plus importante que celle des patients DMD et l'évolution lentement progressive de la maladie rend plus difficile la classification des patients en sous-groupes phénotypiques. Ainsi, sont considérés comme BMD, des patients confinés en fauteuil roulant à 20 ans et décédés avant 40 ans, et des patients de 50 ans présentant uniquement des crampes. Cependant, une échelle permettant de définir la sévérité d'un BMD à partir des symptômes cliniques (GMWS) [Gardner-Medwin D 1974] est utilisée par les cliniciens mais les données sont bien souvent inaccessibles. Une difficulté supplémentaire est la description de phénotypes variables pour une même mutation [Anthony 2011]. Ceci pourrait s'expliquer par l'âge différent des patients qui sont à des stades différents de la maladie mais aussi par la différence des points de cassure intronique dans le cas de délétion d'exons [Gentil 2012].

A l'heure actuelle, il n'existe aucune corrélation claire entre phénotype observé et génotype. Cette corrélation est pourtant indispensable pour proposer de nouvelles thérapies aux patients. C'est dans ce cadre que se déroule ce travail de thèse. En effet, nous proposons d'aller plus loin dans la corrélation génotype - phénotype en étudiant la fonctionnalité des dystrophines mutées exprimées chez les patients.

### 1.2.5 Les essais thérapeutiques dans la Myopathie de Duchenne

Actuellement il n'existe aucun traitement curatif de la myopathie de Duchenne mais uniquement des traitements symptomatiques ne permettant pas de prolonger la vie des patients au delà de 30 ans. Cependant, la myopathie de Duchenne fait l'objet de nombreux essais thérapeutiques.

D'une manière générale, lorsqu'une molécule présente un intérêt thérapeutique, elle doit subir une série d'étapes avant d'être mise sur le marché. Ces étapes sont mondialement harmonisées par l'intermédiaire des Conférences Internationales d'Harmonisation (<http://www.ich.org/>). Ces étapes sont divisés en deux parties : les essais pré-cliniques et les essais cliniques (<http://clinicaltrials.gov>). Les essais pré-cliniques sont réalisés avant la première administration d'un agent thérapeutique chez l'homme et permettent notamment

de tester la toxicité de la molécule sur des modèles animaux. Les modèles animaux majoritairement utilisés dans le cas des dystrophinopathies sont la souris *mdx* et le chien *grmd*, tous deux déficitaires en dystrophine. Si ces tests montrent un rapport bénéfice / risque acceptable, les essais cliniques sur l'homme peuvent démarrer. Ils sont divisés en quatre phases répondant à différents objectifs.

- **Phase I** : première administration à l'homme. L'essai est réalisé sur un petit effectif de sujet sains et jeunes et il permet de déterminer le mode et la dose d'administration ainsi que ses effets sur le métabolisme humain.
- **Phase II** : étude pilote. L'essai est réalisé sur un petit effectif de sujet sains et/ou malades et il permet d'évaluer l'efficacité du médicament, la tolérance à court terme et les risques.
- **Phase III** : étude pivot. L'essai est réalisé sur un grand effectif de malades pour tester l'efficacité du médicament par rapport à un traitement de référence ou un placebo et pour estimer le ratio bénéfice / risque.
- **Phase IV** : post-marketing. Cette phase permet le suivi à long-terme du médicament après sa mise sur le marché afin d'apporter des informations supplémentaires sur les risques, les bénéfices et l'utilisation optimale.

### 1.2.5.1 La thérapie génique

Ces dernières années, de nombreuses recherches ont été menées afin de mettre en place une thérapie de la myopathie de Duchenne en injectant le gène *DMD* dans les cellules musculaires des patients (figure 1.19). Le principe est assez simple car il suffit d'injecter un vecteur contenant un gène fonctionnel dans les muscles des patients mais la réalisation est un vrai défi [Goyenvalle 2011].

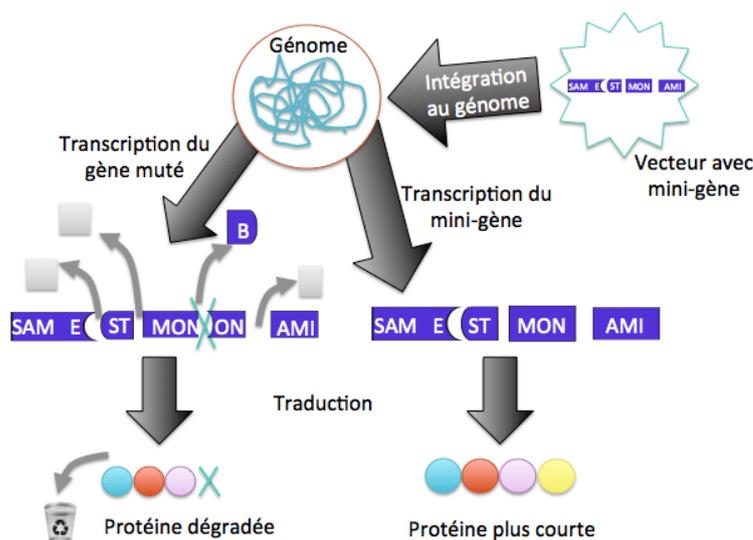


FIGURE 1.19 – Principe de la thérapie génique. Un mini-gène est encapsulé dans un vecteur et injecté dans l'organisme. Ce mini-gène s'intègre au génome et est transcrit puis traduit en une protéine plus courte. La traduction de ce mini-gène se fait en complément de la traduction du gène muté du patient.

La première problématique est le choix d'un vecteur efficace et sans risque. Trois types de vecteurs viraux ont été utilisés dans plusieurs études : les adénovirus, les virus adéno-associés (AAV) et les lentivirus. Les adénovirus et les lentivirus ont été abandonnés à cause d'une trop grande toxicité et / ou à cause d'un manque d'efficacité. Les virus adéno-associés (AAV) et leur forme recombinante (rAAV) sont donc les vecteurs viraux actuellement utilisés pour les essais thérapeutiques. Ils ont montré une bonne efficacité en produisant des gènes stables dans différents modèles animaux [Herzog 1999] ainsi qu'une faible réponse immunitaire de la part de l'organisme hôte [Manno 2003].

Cependant, de nouveaux essais thérapeutiques montrent une réponse immunitaire chez les patients [Mendell 2010, Wang 2010] et posent un nouveau problème. Des immuno-suppresseurs peuvent être utilisés pour empêcher cette réponse immunitaire [Wang 2007] mais la cause de cette réponse immunitaire est encore incertaine. Il est possible que la dystrophine elle-même déclenche cette réponse immunitaire chez les patients. En effet, l'essai clinique de phase I réalisé par Mendell et al. [Mendell 2010, Bowles 2012] met en évidence une réponse immunitaire différente et dose dépendante selon les capsides virales utilisées ainsi qu'une activation des lymphocytes T spécifique à la dystrophine chez certains patients.

Le deuxième problème majeur est la taille limitée du gène pouvant être transféré dans les vecteurs rAAV (~4,7 kb) alors que l'ARN messager du gène *DMD* avoisine les 14 kb. Par conséquent, des gènes tronqués sont utilisés aboutissant à la production de mini- et de micro-dystrophines. Ces constructions sont basées sur des observations de patients BMD présentant un phénotype léger malgré de très grandes délétions en phase dans le domaine central [Matsumura 1994, England 1990]. A partir de ces observations, des mini-dystrophines amputées d'une partie du domaine central ont été construites (les plus courantes sont décrites

dans la figure 1.20). Ces constructions ne correspondent pas à des mutations observées chez des patients puisqu'elles ne respectent pas les bornes des exons mais celles des répétitions. Des résultats encourageants ont été décrits chez la souris [Gregorevic 2008, Gregorevic 2006] mais le récent essai clinique [Mendell 2010, Bowles 2012] ne montre pas d'amélioration chez les patients traités avec un mini gène  $\Delta R3-R21+H3$  (figure 1.20). Un deuxième essai de phase I est en cours.



FIGURE 1.20 – Représentation schématique de mini- et micro-dystrophines utilisées en thérapie génique, adapté de [Goyenvallé 2011]. La mini-dystrophine  $\Delta 17-48$  a été observée chez un patient BMD présentant un phénotype léger. Les autres mini- et micro-dystrophines sont des constructions artificielles en cours de développement. La  $\Delta R3-R21+H3$  est la seule utilisée en essai clinique.

Plusieurs pistes nouvelles sont actuellement en cours d'étude pour remédier aux problèmes décrits précédemment, notamment le problème de la réponse immunitaire observée dans les derniers essais cliniques. L'utilisation d'un promoteur spécifique au muscle comme celui de la créatine kinase au lieu de promoteurs ubiquitaires, tels que le promoteur fort du cytomégalo-virus, pourrait limiter la réponse immunitaire [Cordier 2001, Hartigan-O'Connor 2001]. De plus, il est important d'optimiser la méthode de distribution des vecteurs viraux dans l'ensemble des muscles afin d'en utiliser une dose minimale pour diminuer la réponse immunitaire tout en étant efficace. Des études préconisent des injections à haut volume ou haute pression pour assurer une dissémination dans tout le corps [Su 2005, Cho 2000]. Une autre piste est l'optimisation des codons des mini-gènes en utilisant les codons optimaux permettant une traduction plus rapide des mini-dystrophines dans les cellules [Foster 2008]. Une autre voie de recherche est l'injection d'un mini-gène d'utrophine afin de restaurer une partie des fonctions de la dystrophine tout en évitant une réponse immunitaire anti-dystrophine [Odom 2008]. Cependant, l'utilisation de l'utrophine ne permet pas la restauration complète du complexe DGC, ni la liaison de nNOS au sarcolemme [Li 2010].

Le principe de la thérapie génique dans le cas des dystrophinopathies repose sur l'observation de patients présentant de grandes délétions en phase dans le domaine central et

un phénotype peu sévère [Matsumura 1994, England 1990], soulevant l'hypothèse que le domaine central de la dystrophine est superflu. Ces grandes délétions, décrites dans les années 1990, ont été depuis, très peu observées chez des patients. En 2003, Nevo [Nevo 2003] publie une étude sur trois patients présentant de grandes délétions dans le domaine central et présentant un phénotype sévère de type DMD. Ce type d'étude clinique et les propriétés de liaison du domaine central à de nombreux partenaires décrites dans plusieurs études devrait remettre en cause l'intérêt thérapeutique de la thérapie génique actuelle. En effet, un choix objectif des répétitions du domaine central à conserver dans le mini-gène de dystrophine à transférer semble essentiel à réaliser. Il est donc important de connaître le rôle de toutes les répétitions pour proposer une thérapie génique efficace.

### 1.2.5.2 Le saut d'exons

La thérapie par saut d'exons est un exemple de médecine personnalisée basée sur la maturation de l'ARNm qui pourrait guérir 80% des patients DMD [Aartsma-Rus 2009]. Elle utilise de courtes séquences d'acides nucléiques synthétiques : les oligonucléotides anti-sens (AON). Ces AON sont capables d'altérer l'épissage des ARN messagers en bloquant les sites amplificateurs (enhancers) ou en altérant la structuration de l'ARNm afin de sauter un exon pour recalibrer le cadre de lecture de la dystrophine (figure 1.21). Ainsi une mutation hors phase ne permettant pas la synthèse correcte d'une dystrophine se transformerait en mutation en phase qui permettrait de produire une dystrophine écourtée mais fonctionnelle comme dans les mutations observées chez les patients atteints de myopathie de Becker [Bloetzer 2012, Goyenvalle 2011, Fairclough 2011].

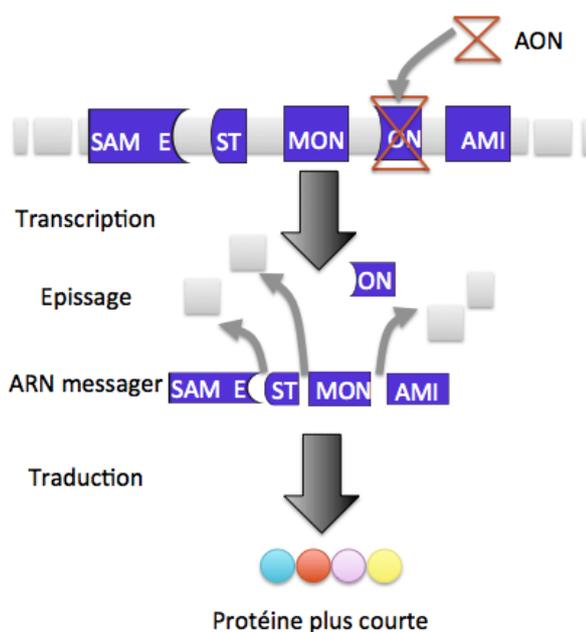


FIGURE 1.21 – *Principe du saut d'exon. Un oligonucléotide anti-sens (AON) est inséré dans les cellules du patient. Cet AON vient empêcher la transcription de l'exon cible. L'absence de l'exon cible dans l'ARN messenger permet de recalibrer le cadre de lecture de la protéine et ainsi produire une protéine plus courte non dégradée.*

Plusieurs essais cliniques de phases I et II ont eu lieu et un essai de phase III est actuellement en cours. Les essais se concentrent principalement sur le saut de l'exon 51 qui cible, à lui seul, 13% des patients DMD. En effet, ce saut pourrait corriger les délétions hors phase d'exons 45-50, 47-50, 49-50, 50, 52 et 52-63. Deux récents essais cliniques de phase I et II publiés en 2011 ont montré une augmentation de l'expression de la dystrophine de manière dose-dépendante dans les muscles squelettiques sans effet secondaire important. Par contre, ils ont montré une faible efficacité dans le muscle cardiaque et l'absence d'amélioration significative de la fonction motrice [Goemans 2011, Cirak 2011].

Les essais cliniques réalisés ont démontré la faisabilité de la thérapie par saut d'exon mais plusieurs obstacles demeurent. L'un des problèmes majeurs est l'efficacité variable du saut d'exon selon les cellules musculaires, notamment sa faible efficacité dans le muscle cardiaque et l'autre problème majeur est une rapide dégradation des AON. Des tests pré-cliniques sur les animaux ont montré que même avec de très fortes doses d'AON (au moins 10 fois les doses utilisées en essai clinique), il n'y a pas d'expression uniforme de dystrophine dans les muscles squelettiques et il n'y a qu'une faible expression dans le cœur [Lu 2005, Yokota 2009]. Récemment, de nouveaux AON couplés à des peptides ont été développés, les PPMO [Moulton 2010, Yin 2008]. Ces PPMO semblent présenter une meilleure efficacité de pénétration notamment dans le cœur tout en étant utilisés à des doses et des fréquences d'injection réduites [Jearawiriyapaisarn 2010].

Une deuxième approche est l'encapsulation d'un oligonucléotide anti-sens dans un vecteur

viral de la même manière que pour la thérapie génique [Goyenvalle 2004]. Cette approche permettrait une dissémination durable des AON au corps entier avec une seule injection [Denti 2006]. Cette approche est pour l'instant à un stade pré-clinique et l'utilisation d'un vecteur viral pose les mêmes problèmes que pour la thérapie génique.

Malgré les avancées majeures que la thérapie par saut d'exons a réalisé, l'applicabilité de cette thérapie personnalisée demeure un problème majeur. En effet, même si le saut d'exons peut théoriquement être appliqué à 83% des mutations observées chez les patients DMD, il nécessite le développement de nombreux AON ciblant différents exons correspondant aux mutations. Outre les problèmes de développement de ces AON et le surplus de coût, le principal problème est la mise sur le marché de ces AON. Chaque AON, considéré comme un nouveau médicament, doit subir le long et coûteux processus d'autorisation de mise sur le marché. En outre, plusieurs AON développés pour corriger des mutations rares ne pourraient pas passer les différentes phases d'essais cliniques faute d'un nombre suffisant de patients et seraient trop chers à produire [Bloetzer 2012, Goyenvalle 2011, Fairclough 2011].

Pour contrer ce problème, certaines études préconisent d'utiliser un saut d'exon multiple, comme par exemple, le saut des exons 45 à 55 qui permettrait de corriger à lui seul près de 60% des patients DMD [Beroud 2007]. Cependant, cette mutation pose problème puisqu'elle tronque le domaine de liaison à la nNOS et est associée chez certains patients à un phénotype BMD sévère [Gentil 2012].

D'une manière générale, le choix des exons à masquer dans le cas des thérapies par saut d'exons dépend uniquement du nombre de patients DMD que ce saut pourrait potentiellement guérir et de la facilité de mise au point de l'AON correspondant. La sévérité du phénotype de certains patients BMD présentant la mutation en phase ciblée et la fonctionnalité de la dystrophine tronquée produite ne sont donc jamais pris en compte dans ce choix. Ceci soulève donc la question de l'efficacité potentielle de la thérapie par saut d'exons.

### 1.2.5.3 La thérapie cellulaire

La thérapie cellulaire utilise des cellules souches pour corriger le phénotype des patients DMD [Farini 2009, Blau 2008, Fairclough 2011]. Une fois transplantées chez le patient, ces cellules doivent fusionner avec les myotubes existant ou en former de nouveaux conduisant à la production d'une dystrophine fonctionnelle [Price 2007]. Actuellement, il existe deux stratégies. La première utilise les cellules du patient modifiées génétiquement pour produire une dystrophine fonctionnelle puis réinjectées chez le patient [Mendell 2006, Chakkalakal 2005]. La deuxième stratégie utilise les cellules souches d'un donneur sain et injectées chez le patient [Farini 2009, Partridge 2004]. Les deux méthodes ont leurs avantages et leurs inconvénients. L'utilisation de cellules du patient permet d'éviter une réponse immunitaire mais leur modification génétique présente une difficulté supplémentaire ainsi que des effets

indésirables tels que la transformation des cellules souches pouvant conduire à leur mort [Rivière 2006, Hacein-Bey-Abina 2003]. L'utilisation de cellules souches d'un donneur sain ne nécessite pas de modification génétique mais elle présente un risque de rejet et nécessite donc une bonne compatibilité immunitaire entre le donneur et le receveur ainsi que l'utilisation d'immunosuppresseurs.

En théorie, la thérapie cellulaire est très prometteuse. En effet, un petit nombre de cellules couplées à un facteur de croissance est nécessaire pour obtenir un effet thérapeutique, mais en pratique, elle présente encore de nombreux obstacles [Price 2007]. Actuellement, plusieurs recherches portent sur différentes lignées de cellules souches avec des degrés de pluripotence et des potentiels myogéniques différents [Rouger 2011, Farini 2009, Price 2007]. Quelque soit le type de cellule étudié, les trois principaux critères pour obtenir une thérapie efficace sont la survie, la capacité migratoire et la faible immunogénicité des cellules souches étudiées. Actuellement les cellules présentant les meilleures chances de succès sont les mesangioblastes [Price 2007] et les cellules CD133+ [Torrente 2007].

#### 1.2.5.4 La translecture du codon STOP

La translecture du codon STOP peut être appliquée aux patients présentant une mutation non-sens (mutation conduisant à la génération d'un codon STOP) qui représentent environ 13% des patients DMD [Fairclough 2011, Bloetzer 2012].

Les antibiotiques amino-glycosides permettent aux ribosomes d'ignorer les codon STOP et de continuer la synthèse de la protéine. Un essai clinique avec la gentamicine a montré une augmentation de l'expression de la dystrophine et une baisse du taux de créatine kinase dans le muscle squelettique mais il n'a été démontré aucun bénéfice clinique et des questions sur sa toxicité ont été soulevées [Malik 2010]. Une deuxième molécule est à l'essai : l'ataluren (PTC124). Cette molécule agit exclusivement sur les codon STOP prématurés. Plusieurs essais cliniques ont été réalisés (dont un en phase IIb) mais les premières analyses n'ont pas montré d'amélioration significative [Welch 2007, Finkel 2010]. Tous les essais cliniques ont donc été arrêtés mais devraient reprendre prochainement.

#### 1.2.5.5 La surexpression de l'utrophine

La surexpression de l'utrophine est une thérapie visant à compenser l'absence de dystrophine par une surexpression de son paralogue : l'utrophine. L'utrophine et la dystrophine ont une taille et une structure proche et partagent de nombreux domaines de liaisons [Fairclough 2011, Tinsley 1992, Winder 1995b]. Cependant, l'utrophine est très peu présente dans les muscles squelettiques adultes et est incapable d'assurer toutes les fonctions de liaison de la dystrophine (voir 1.1.2.5). Différentes molécules sont étudiées afin de sur-activer le promoteur A responsable de l'expression musculaire de l'utrophine [Fairclough 2011].

Des études chez la souris ont montré une amélioration du phénotype dans des conditions de vie sédentaires par la restauration du complexe associé à la dystrophine au sarcolemme [Tinsley 1996, Gilbert 1999]. Le même type d'étude réalisé chez le chien a également montré une amélioration du phénotype [Cerletti 2003]. Cependant en 2010, Li *et al.*, ont démontré que l'utrophine ne compensait pas complètement la dystrophine chez les souris faisant de l'exercice [Li 2010]. Un essai de phase I a été réalisé sur une molécule ayant démontré son potentiel thérapeutique chez la souris en diminuant le taux de membranes lésées [Tinsley 2011].

## 1.3 L'apport de la bioinformatique

Depuis les années 1970, les avancées en biologie moléculaire et en génomique ont conduit à une explosion de la quantité d'information générée nécessitant l'aide de l'informatique pour les traiter. C'est ainsi que naquit la bioinformatique. Il est cependant difficile de dater précisément son apparition mais le terme de bioinformatique apparaît pour la première fois en 1977 lorsqu'il est utilisé par Paulien Hogeweg pour décrire la nouvelle équipe qu'il souhaite former à l'université de Utrecht [Hogeweg 1978]. C'est dans les années 1990 que la bioinformatique prend son essor avec la création du National Center for Biotechnology Information (NCBI) aux USA et la conférence "Bioinformatics in the 90s" en Europe [Attwood 2011]. A l'origine cantonnée à la création de bases de données pour stocker essentiellement des séquences nucléotidiques et protéiques, la bioinformatique est maintenant un champ de recherche multidisciplinaire aux frontières de la biologie, de l'informatique, des mathématiques, de la physique et de la chimie. Le champ d'application de la bioinformatique est vaste et comprend, la génomique, l'écologie, la biologie structurale et la biologie cellulaire. La bioinformatique est difficile à définir, mais le NCBI en propose la définition suivante : " la bioinformatique est le domaine de la science où la biologie, l'informatique et la technologie de l'information se rassemblent pour former une seule discipline" (<http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>). Par analogie avec les études *in vivo* et *in vitro*, on parle alors d'études *in silico* (expression formée à partir du mot silicium, un composant majeur des ordinateurs).

### 1.3.1 Dates importantes dans l'histoire de la bioinformatique

- **1951** Pauling et Corey proposent la structure de l'hélice  $\alpha$  et du feuillet  $\beta$ .
- **1953** Watson et Crick proposent le modèle de la double hélice d'ADN basé sur des données de rayons X obtenues par Franklin et Wilkins.
- **1955** Sanger propose la première séquence d'une protéine : l'insuline bovine.
- **1960** Découverte d'un lien entre séquence ADN et séquence protéique.  
Lien entre séquence et structure (Globines, Perutz).
- **1965** L' "Atlas of Protein Sequences and Structures" est publié par Dayhoff *et. al.*
- **1966** Découverte du code génétique.
- **1970** L'algorithme d'alignement de séquences de Needleman-Wunsch est publié.
- **1972** La première molécule d'ADN recombinant est créée par l'équipe de Paul Berg.
- **1974** "Prediction of Protein Conformation" (Chou et Fasman)
- **1977** La Protein Data Bank est publiée.  
Maxam, Gilbert et Sanger publient la méthode de séquençage de l'ADN.  
Première simulation de dynamique moléculaire d'une protéine.

- **1980** La première séquence d'un génome entier est publiée (FX174).
- **1981** L'algorithme d'alignement de séquences de Smith-Waterman est publié.
- **1985** La réaction de PCR est décrite par Mullis *et. al.*
- **1986** Apparition du terme "Génomique".  
La base de données SWISS-PROT est créée.
- **1988** Création du National Center for Biotechnology Information (NCBI).  
Développement de CLUSTAL.
- **1990** Le programme BLAST est implémenté.
- **1991** Création des séquences EST (expressed sequence tag).  
Création de l'European Bioinformatics Institute (EMBL-EBI) à Cambridge.
- **1995** Séquençage des premiers génomes bactériens.
- **1996** Séquençage de la levure *Saccharomyces cerevisiae* (12,1 Mb).
- **1998** Séquençage du premier génome d'organisme multicellulaire.  
La fondation Swiss Institute of Bioinformatics est créée.
- **2001** Publication de la séquence génome humain.

Depuis les années 2000, l'acquisition de nouvelles données s'accroît de façon exponentielle et la bioinformatique est devenue indispensable pour analyser cette masse de données.

## 1.3.2 Les bases de données

### 1.3.2.1 Une nécessité pour la recherche actuelle

Les bases de données biologiques représentent une importante section de la bioinformatique dont les plus connues sont GenBank, Swissprot et la Protein Data Bank (PDB). Ce sont des systèmes structurés de données persistantes, généralement associées à un programme permettant de faire des recherches et de mettre à jour les données du système [Thampi 2009]. En bioinformatique, les bases de données relationnelles sont le standard actuel. Elles peuvent être représentées comme un ensemble de tableaux contenant le même type de données et associés entre eux par des relations. En recherche, les deux principaux intérêts de l'utilisation des bases de données sont (1) un accès facile et rapide aux informations et (2) une extraction simple des informations nécessaires pour répondre à la question biologique. Depuis l'avènement d'internet, les bases de données sont généralement accessibles grâce à un site internet permettant l'accès aux données par l'intermédiaire de formulaires à remplir. L'accès aux données est donc facilité puisqu'il ne nécessite pas de connaissances en langage de programmation mais en contre partie il est restreint à quelques questions. Il n'existe pas de chiffres précis concernant le nombre de bases de données biologiques mais l'apparition récente de journaux dédiés montre l'importance grandissante de cette branche de la bioinformatique.

A l'origine, les premières bases de données bioinformatiques (EMBL, Genbank, DDBJ,

Swissprot, TrEMBL, PIR-NBRF, UniProt) étaient des bases généralistes, conçues pour héberger des séquences nucléiques ou protéiques. Ces banques s'échangent systématiquement leur contenu depuis 1987 et adoptent un système de conventions communes (The DDBJ/EMBL/GenBank Feature Table Definition). Aujourd'hui, les bases de données contiennent toutes sortes de données générées au fur et à mesure des avancées de la recherche en biologie, comme par exemple les annotations de génomes, les données d'interactome, les structures de protéines et les sites de coupure d'enzyme de restriction... La quantité de données évolue également et est en croissance exponentielle, comme par exemple pour Genbank [Botstein 1997, de Brem 2012]. Les bases de données sont donc devenues un outil d'analyse important en biologie. C'est le cas, par exemple, de Pubmed, qui permet l'accès à la base de données bibliographique MEDLINE, rassemblant des citations et des résumés d'articles de recherche biologie et en médecine.

### 1.3.2.2 Les bases de données générales et les base de données spécifiques

Alors que les bases de données généralistes citées précédemment se concentrent sur un seul type de données et tendent à être exhaustives. Les bases de données dites "spécialisées" sont créées en fonction des besoins spécifiques liés à l'activité d'un groupe de chercheurs et sont souvent créées au sein des laboratoires. Elles sont d'intérêt divers et la masse des données qu'elles contiennent varient d'une base à une autre. Ces bases correspondent à des améliorations ou à des regroupements par rapport aux données issues des bases généralistes.

Les bases de données spécifiques sont créées pour répondre à une question spécialisée, comme par exemple les bases de données qui analysent un gène responsable d'une maladie : les bases de données gène-spécifique (LSBD). Ces bases de données sont standardisées car elles sont réalisées à partir de logiciels développés spécifiquement pour ce type de base de données (ex : LOVD [Fokkema 2005, Fokkema 2011], UMD [Bérout 2005]). Les bases de données gène-spécifiques rassemblent des données sur les mutations et les polymorphismes observés chez des patients pour une pathologie donnée. Elles peuvent également contenir des informations cliniques et biologiques détaillées utiles pour le diagnostic [Humbertclaude 2010]. Ces bases de données ont pour but d'étudier les corrélations génotypes - phénotypes mais sont bien souvent confrontées au problème de l'analyse des données phénotypiques. En effet, ce type de données très hétérogène est souvent renseigné sous forme simple et non standardisée, par exemple, sous forme de texte avec peu de détails [Patrinos 2005]. Ce type de données est très difficile, voire impossible à analyser de façon automatisée dans une base de données. Cependant, depuis quelques années, un effort de standardisation a été réalisé à l'aide d'experts afin de créer un langage commun pour décrire, collecter, analyser les allèles, les génotypes et les phénotypes à l'échelle nationale et internationale [Humbertclaude 2010, Claustres 2002]. Ce travail nécessite de reprendre les dossiers des patients un par un, parfois de les compléter

en réalisant de nouvelles analyses ou en réinterrogeant les médecins qui ont posé le diagnostic. Ces données en cours de réannotation sont, bien souvent, publiquement inaccessibles.

### 1.3.2.3 L'intérêt d'une base de données spécifique à la dystrophine

Il existe deux bases de données publiques spécifiques du gène *DMD* : la Leiden Muscular Dystrophy database [White 2006] et l'UMD-DMD [Tuffery-Giraud 2009]. La Leiden Muscular Dystrophy database collecte l'ensemble des mutations du gène *DMD* publiées dans la littérature ou soumis par des contributeurs et inclut quelques détails phénotypiques et biochimiques. L'UMD-DMD fournit des informations moléculaires et cliniques de patients français ayant une mutation sur le gène *DMD*. Ces bases de données contiennent des mutations en phase et hors-phase du gène et ciblent les informations génétiques. Elles visent à établir une corrélation génotype - phénotype. Cependant, elles n'apportent pas d'information sur la dystrophine produite. Or la connaissance de la structure et la fonction de la dystrophine native ainsi que des dystrophines tronquées produites suite à des mutations en phase est primordiale pour proposer une telle corrélation génotype - phénotype. Par conséquent, nous proposons d'établir une base de données spécifique de la dystrophine ciblée sur la protéine en complément des bases de données gène-spécifique existantes.

## 1.3.3 La modélisation moléculaire

La modélisation moléculaire des protéines est une méthode théorique à la frontière entre la biologie, la chimie et la physique utilisant des moyens informatiques. Elle permet de prédire la structure tridimensionnelle de molécules ainsi que leurs interactions les unes avec les autres. Elle est utilisée à des fins prédictives et explicatives. La plupart des méthodes de prédiction *in silico* de structure de protéines repose sur l'hypothèse thermodynamique de Anfinsen [Anfinsen 1973]. Selon cette hypothèse, la structure tridimensionnelle d'une protéine dépend de sa séquence en acides aminés et la prédiction consiste à trouver la structure de plus faible énergie.

### 1.3.3.1 L'intérêt de la modélisation par homologie

La fonction des protéines dépend de leur structure tridimensionnelle, c'est-à-dire de la manière dont les acides aminés de la chaîne polypeptidique s'agencent les uns par rapport aux autres dans l'espace. En général, une protéine est fonctionnelle dans son état replié (natif). La connaissance de la structure des protéines est donc essentielle pour comprendre leur mécanisme de fonctionnement.

Il existe plusieurs méthodes expérimentales pour obtenir la structure tridimensionnelle à haute-résolution des protéines : la cristallographie et la diffraction à rayons X et la Résonance

Magnétique Nucléaire (RMN). Lorsque la structure d'une protéine est résolue expérimentalement, elle est déposée dans la Protein Data Bank (PDB). Ces techniques sont coûteuses, difficiles et longues à mettre en place. Au regard du nombre de protéines découvertes (536 000 séquences référencées dans UniProtKB/Swiss-Prot en 2012), il est impossible de proposer une approche expérimentale pour caractériser la structure de toutes ces protéines (77 000 structures de protéines dans la PDB en 2012). De plus, les techniques expérimentales se heurtent souvent à des problèmes de faisabilité, comme c'est le cas pour les répétitions de la dystrophine. Dans ce type de cas, des méthodes basse résolution comme la cryo-microscopie et la diffusion des rayons X aux petits angles (SAXS) peuvent être mises en place.

Lorsqu'aucune structure n'est disponible, la méthode de prédiction de structure tridimensionnelle la plus utilisée est la modélisation par homologie. Elle est basée sur un concept simple : des protéines (ou des domaines protéiques) qui sont proches d'un point de vue évolutif ont tendance à adopter une structure tridimensionnelle proche [Venclovas 2012]. Par conséquent, une ou plusieurs structures de protéines homologues sont utilisées comme patron pour réaliser un modèle par homologie. Les structures 3D sont considérées comme la caractéristique la plus conservée au cours de l'évolution et c'est particulièrement vrai pour les faisceaux d'hélices comme dans la dystrophine qui sont, par conséquent, de bons candidats pour la modélisation par homologie [Parry 2008]. Lorsqu'aucune structure de protéine homologue n'est disponible des méthodes de prédiction de repliement *ab-initio* (folding) sont utilisées.

L'approche *in silico* de modélisation par homologie comporte quatre étapes principales (figure 1.22) [Venclovas 2012, Venselaar 2010] :

- **Identification de protéines homologues** dont les structures expérimentales sont disponibles pour être utilisées comme patron (étape 1 de la figure 1.22). Ces protéine-patrons sont généralement recherchées à l'aide d'outils de recherche d'homologie de séquences interrogeant des bases de données, comme par exemple la Protein Data Bank [Cozzetto 2009]. L'étape initiale de choix du patron est cruciale pour obtenir un modèle par homologie de qualité. Par conséquent, quand plusieurs patrons sont disponibles, ils sont utilisés pour proposer un modèle plus robuste.
- **Alignement de la séquence d'acides aminés** de la protéine à modéliser et de celle du (des) patron(s) choisi(s) (étape 2 de la figure 1.22). Afin d'obtenir le meilleur alignement possible, les logiciels utilisent une combinaison de plusieurs algorithmes. La séquence alignée est alors séparée en régions conservées et régions variables (boucles) afin d'optimiser l'alignement de ces régions variables.
- **Création du modèle 3D** à partir de l'alignement de séquences précédemment réalisé (étapes 3 à 6 de la figure 1.22). Les coordonnées des acides aminés identiques sont reprises pour créer le modèle de la protéine d'intérêt alors que seules les coordonnées

du squelette (backbone) des acides aminés similaires sont reprises puis complétées en ajoutant les chaînes latérales. La modélisation des boucles est une étape délicate et nécessite un traitement particulier basé sur des alignements structuraux. Le modèle est ensuite optimisé et minimisé. Il est enfin vérifié pour minimiser les angles de torsion non permis, minimiser le nombre de mauvais contacts, maximiser les liaisons hydrogènes.

- **Vérification de la qualité** du modèle créé (calcul d'un score de qualité) (étapes 7 et 8 de la figure 1.22). Si la qualité est satisfaisante, l'ensemble des coordonnées des atomes de la protéine sous format pdb est retourné; dans le cas contraire, les étapes précédentes sont réitérées.

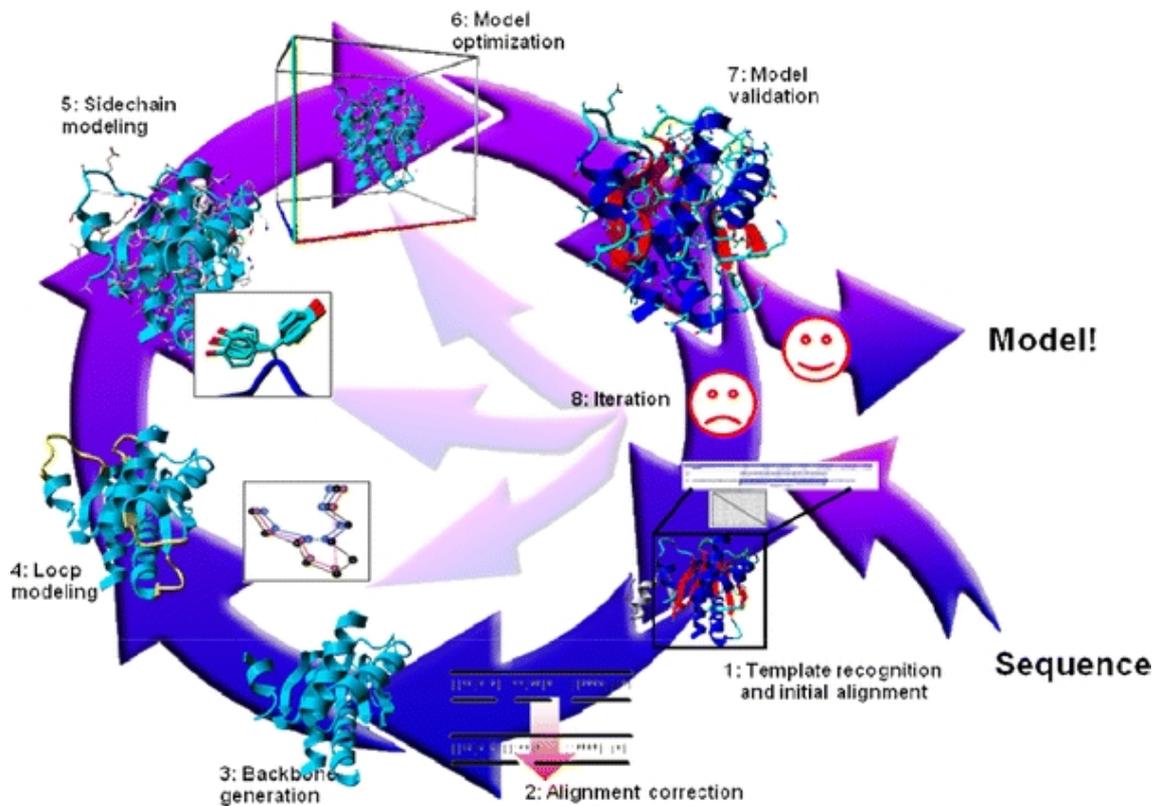


FIGURE 1.22 – Description des étapes de modélisation moléculaire par homologie [Venselaar 2010].

Les méthodes de prédiction de structure tridimensionnelle de protéines sont en amélioration permanente, notamment grâce à l'impulsion du "concours" CASP (Critical Assessment of Techniques for Protein Structure Prediction) qui est organisé tous les deux ans depuis 1994 [Kryshtafovych 2011b]. Ce concours permet à des groupes de recherche de tester leurs méthodes de prédiction de manière objective en prédisant la structure de plusieurs protéines. Les structures de ces protéines ont été résolues expérimentalement mais les concurrents n'ont accès qu'à la séquence protéique. Les modèles obtenus sont alors comparés aux structures expérimentales afin de proposer un classement des méthodes les plus efficaces [Kryshtafovych 2011a].

### 1.3.3.2 Les champs de force

Afin de calculer les forces nécessaires pour caractériser les interactions intra- et extramoléculaires qui participent au repliement des molécules, il est nécessaire d'utiliser un champ de force [Bordner 2012]. Un champ de force est défini comme l'ensemble des équations et des paramètres qui permettent de décrire quantitativement l'énergie d'un système chimique. Les champs de force sont utilisés en modélisation par homologie pendant l'étape de construction du modèle 3D. Ils corrigent la géométrie de la protéine, évitent les conflits stériques et les recouvrements atomiques, sélectionnent le modèle le plus proche de la structure native et évaluent la qualité du modèle final. Ils sont également utilisés en dynamique moléculaire (cf. section 1.3.3.3).

On distingue notamment des champs de force tout-atome et gros grain. Les champs de force tout-atome, comme par exemple AMBER [Duan 2003], CHARMM27 [Zhu 2012] ou GROMOS [Oostenbrink 2004] décrivent les protéines au niveau atomique. Ils permettent de caractériser chaque atome par des liaisons, angles, dièdres, interactions d'attraction - répulsion et des charges électrostatiques. Les champs de force gros grains comme par exemple, MARTINI [Monticelli 2008] utilisent des "pseudo-atomes" représentant un ensemble d'atomes diminuant ainsi les calculs afin d'accélérer l'échantillonnage conformationnel et d'accélérer la simulation.

L'utilisation des champs de force pour étudier les protéines permet également de se rapprocher des conditions physiologiques en prenant en compte la présence d'un solvant (en général, eau et ions). Cette solvatation peut être explicite en incluant des molécules d'eau et des ions ou implicite où le solvant est représenté par une fonction mathématique. Les champs de force en solvant implicite sont plus approximatifs mais diminuent considérablement les coûts de calcul.

Le choix du champ de force dépend donc de ce que l'on souhaite modéliser (protéine, ligand...), de la durée de simulation et de la précision que l'on souhaite obtenir et de la puissance de calcul dont on dispose.

### 1.3.3.3 La dynamique moléculaire

La dynamique moléculaire (DM) est la méthode *in silico* la plus utilisée pour étudier les propriétés dynamiques des protéines à l'échelle atomique dans le cas d'une simulation tout-atome [Cramer 2004]. Elle permet, par exemple, d'étudier la stabilité d'une protéine, les changements conformationnels et les interactions ligands-récepteurs. La première simulation de dynamique moléculaire d'une protéine a été réalisée en 1977 sur une petite protéine de 500 atomes pendant 10 picosecondes [McCammon 1977]. De nos jours, grâce à l'augmentation de la puissance des ordinateurs (calculateurs) et l'optimisation des programmes, on peut couramment réaliser des simulations de petites enzymes pendant quelques microsecondes ou simuler des systèmes contenant des millions d'atomes. Cependant, ce type de simulations

nécessite plusieurs semaines de calculs.

L'amélioration de modèles par homologie grâce à des simulations de DM est une méthode très largement utilisée [Nurisso 2012]. Les simulations de DM permettent de pallier un affinement insuffisant des modèles qui est une des principales sources d'erreurs en modélisation par homologie.

### Principe général

La dynamique moléculaire évalue l'évolution d'un système (ici la structure ou l'agencement des atomes d'une protéine) en intégrant numériquement la seconde équation de mouvement de Newton (Eq. 1.1) [Leach 2001]:

$$a_i(t) = \frac{d^2x_i}{dt^2} = \frac{F(x_i)}{m_i} \quad (1.1)$$

où  $a_i$  est l'accélération de la particule  $i$  au temps  $t$  déterminée par la force  $F(x_i)$  agissant sur une particule  $i$  de masse  $m_i$  à la position  $x_i$ . La force  $F(x_i)$  est calculée à partir de l'équation du champ de force choisi en dérivant l'expression du potentiel d'énergie  $V(x_i)$ . NAMD utilise une fonction de potentiel d'énergie paramétrable suivant le champ de force choisi (CHARMM ou AMBER) et décrite ci-dessous (Eq. 1.2 [Phillips 2005]):

$$V_{total} = \sum_{i=1}^{N_{atom}} V_{bond}(i) + V_{angle}(i) + V_{dihedral}(i) + V_{vdW}(i) + V_{Coulomb}(i) \quad (1.2)$$

où  $V_{total}$  est le potentiel d'énergie totale,  $V_{bond}$  décrit les liaisons covalentes,  $V_{angle}$  décrit les angles entre deux liaisons peptidiques,  $V_{dihedral}$  décrit les angles de torsion,  $V_{vdW}$  décrit les forces de van der Waal (interactions à courte distance) et  $V_{Coulomb}$  décrit les interactions électrostatiques.

La position de chaque atome au temps  $t$  dépend de sa position au temps  $t - 1$  et de sa vitesse durant  $dt$ . La vitesse d'un atome au temps  $t$  peut être calculée en intégrant l'équation de vitesse pour chaque atome à chaque pas de temps  $dt$ . Ainsi, il est possible d'évaluer la position de chaque atome du système en fonction du temps. Le coût des calculs en dynamique moléculaire dépendent notamment du pas de temps. En théorie, le pas de temps est déterminé par la vibration la plus rapide au sein de la molécule; dans le cas des protéines, il s'agit de la vibration des liaisons hydrogènes. Le pas de temps maximum théorique pour une protéine est donc de 0,5 fs ( $10^{-15}$  s), mais son utilisation serait trop coûteuse. Par conséquent, des algorithmes particuliers (ex : SHAKE) sont utilisés pour augmenter ce pas de temps à 2 fs afin de diminuer la complexité des calculs.

Ainsi, au cours de la dynamique moléculaire d'une protéine, une structure est produite toutes les 2 fs. Chaque structure produite dépend de la structure précédente et on parle donc de trajectoire. Le point de départ d'une trajectoire est aléatoire mais la trajectoire converge,

quelque soit le point de départ, vers des conformations représentatives de la protéine. Cependant, un incident peut se produire sur une structure et se répercuter tout au long de la trajectoire. Pour écarter ce biais, une simulation de DM doit être reproduite plusieurs fois avec un point de départ différent pour confirmer la convergence des trajectoires vers des structures représentatives.

### Déroulement d'une simulation de dynamique moléculaire

En pratique, les simulations de dynamique moléculaire comportent cinq étapes [Nurisso 2012] :

- **Préparation du système.** La structure initiale (expérimentale ou modélisée) est préparée en fonction du champ de force et des paramètres choisis. Généralement, il s'agit d'ajouter, dans le fichier de coordonnées, les atomes manquants (par exemple les atomes d'hydrogène), d'attribuer les charges et de renommer les atomes en fonction du champ de force. Dans le cas d'une simulation en solvant explicite, la protéine est placée dans une boîte où des molécules d'eau et des ions sont ajoutés en fonction des paramètres du champ de force choisi (figure 1.23).
- **Minimisation.** La minimisation permet de relaxer le système et d'éliminer les mauvais contacts afin d'éviter les changements de conformations trop brusques dès le début de la DM. La structure minimisée obtenue correspond généralement à un minimum local proche de la structure de départ plutôt qu'au minimum global d'énergie.
- **Chauffage.** Le système est lentement chauffé jusqu'à la température désirée afin de lui apporter de l'énergie.
- **Equilibration.** L'équilibration permet au système de se relâcher et d'atteindre un équilibre. Il est très important de vérifier que le système converge vers un état d'équilibre avant de passer à l'étape suivante. Plusieurs analyses de la simulation peuvent être réalisées pour déterminer l'état d'équilibre comme, par exemple, le calcul de la déviation quadratique moyenne (RMSD) par rapport à la structure initiale.
- **Production.** La durée de cette phase est prédéterminée, elle est généralement la plus longue. Durant cette phase, les données structurales et énergétiques du système sont collectées à intervalle régulier afin d'être analysées. Le but de cette phase est, généralement, de collecter suffisamment de conformations représentatives de la protéine.

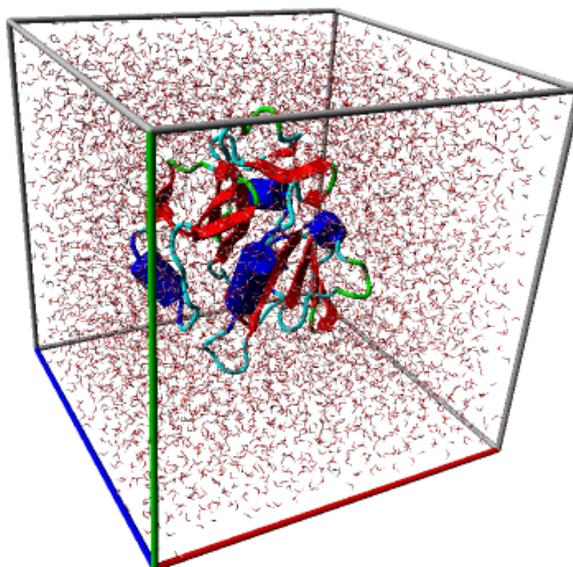


FIGURE 1.23 – Exemple d'une protéine (la dihydrofolate réductase) en présence de molécules d'eau utilisée pour une simulation de dynamique moléculaire en solvant explicite.

#### 1.3.3.4 Quelques succès en modélisation moléculaire

Lorsque la structure 3D d'une protéine n'est pas résolue mais qu'il existe suffisamment d'informations sur sa structure, les approches de modélisation par homologie sont généralement un succès [Nurisso 2012]. Plusieurs études de modélisation par homologie couplés à de la dynamique moléculaire ont montré des résultats significatifs notamment dans le domaine de la conception de médicaments (drug design en anglais) [Wyss 2003].

Ainsi, les récepteurs couplés aux protéines G sont très étudiés car ils sont impliqués dans de nombreux processus physiologiques et sont les cibles de nombreux médicaments. Dans cette famille, la structure tridimensionnelle du récepteur aux opioïdes  $\mu$  a été modélisée par homologie puis de courtes simulations de DM ont été réalisées dans un environnement membranaire et aqueux. Le naltrexone, un médicament décrit comme antagoniste potentiel [Metzger 2001], a ensuite pu être arrimé (docké) sur le récepteur et les acides aminés impliqués ont pu être identifiés [Zhang 2005]. La modélisation de ce récepteur capable de lier un antagoniste connu est la première étape pour construire un complexe d'interaction comprenant le récepteur, un agoniste et une protéine G dans une matrice membranaire.

Les cytochromes P450 (CYP450) impliqués dans la dégradation des xénobiotiques dont les médicaments sont également très étudiés. Les structures de deux CYP450 murins (CYP2C38 et CYP2C39) présentant des profils de sélectivité de substrats différents et une très forte identité de séquence (91,8%) ont été modélisées par homologie. Les modèles structuraux affinés par de courtes DM ont ensuite permis de réaliser des amarrages de ligands. Ainsi, l'étude a montré que le tolbutamide, un substrat commun aux deux enzymes, s'arrime dans des orientations différentes dans leurs site actifs [Meng 2009]. Dans un deuxième temps, la spécificité de l'acide rétinoïque all-trans pour le CYP2C39 a été étudiée. La comparaison

des sites actifs des CYP450, a permis d'émettre plusieurs hypothèses afin d'expliquer cette différence d'affinité. Cette étude a ainsi mis en évidence les acides aminés essentiels à la liaison aux substrats.

La modélisation moléculaire est également utilisée pour étudier les vecteurs de maladies tropicales telle que la malaria causée par des parasites du genre *Plasmodium*. La structure d'un fragment de la protéine de surface du mérozoïte de *Plasmodium vivax* (PvMSP-1) a pu être modélisée par homologie et affinée par de courtes DM [Serrano 2006]. Le modèle ainsi obtenu a permis de décrire une poche d'interactions protéines-protéines. Ce modèle ouvre de nouvelles perspectives dans l'étude des modes d'action du parasite et dans la mise au point d'agent anti-malaria.

### 1.3.3.5 L'intérêt de la modélisation pour l'étude de la dystrophine

L'étude des répétitions de la dystrophine est un enjeu important pour comprendre la corrélation génotype-phénotype des dystrophinopathies. En effet, la majorité des mutations observées chez les patients se situent dans ce domaine central. Peu d'études sont réalisées sur la dystrophine native et très peu d'études existent sur les dystrophines tronquées produites. Or, il est primordial de connaître la structure de la dystrophine et en particulier de ses répétitions pour connaître la fonctionnalité des dystrophines tronquées observées chez les patients et des dystrophines créées pour les thérapies. Les études récentes décrivent le domaine central de la dystrophine comme impliqué dans de nombreuses interactions au sein de la cellule musculaire. La structure des répétitions du domaine central n'ayant pas été résolue expérimentalement, il n'existe pas d'étude des interactions au niveau moléculaire de ce domaine central.

Ainsi, nous proposons une étude de la structure du domaine central par une approche de modélisation par homologie couplée à de la dynamique moléculaire. Cette approche permettra d'étudier le comportement des dystrophines tronquées et de les comparer entre elles pour aider à comprendre pourquoi certaines dystrophines tronquées sont associées à des phénotypes plus ou moins sévères.

## 1.4 Objectifs du travail

La dystrophine est une protéine impliquée dans les myopathies de Duchenne (DMD) et de Becker (BMD). Malgré les nombreuses recherches cliniques et thérapeutiques effectuées sur la DMD, le rôle moléculaire précis de la dystrophine est largement méconnu rendant la corrélation entre génotype et phénotype difficile à établir. Cette corrélation est pourtant indispensable pour proposer de nouvelles thérapies efficaces aux patients. C'est dans ce contexte que se déroule ce travail de thèse.

Nous proposons d'aller plus loin dans la corrélation génotype-phénotype en étudiant la fonctionnalité des dystrophines mutées exprimées chez les patients afin de répondre aux trois questions suivantes :

- Comment la structure de la dystrophine permet-elle de comprendre sa fonction au sein des cellules musculaires ?
- Que peut-on déduire des mutations en phase identifiées chez les patients ?
- Comment l'étude de la structure de la dystrophine et de ses mutations permet-elle de corréler génotype et phénotype ?

Une des difficultés majeures pour l'étude de la dystrophine est sa très grande taille (3685 acides aminés) qui rend très difficile les approches expérimentales. C'est pour cela qu'au sein de l'équipe, nous étudions cette protéine par fragments. Ces fragments sont proposés en fonction de la présence d'un domaine de liaison ou d'une région présentant des mutations d'intérêt. Afin de proposer des études multi-disciplinaires de ces fragments, la même stratégie est appliquée pour les différentes approches développées au sein de l'équipe dont celles développées dans ce travail de thèse.

Deux approches *in silico* ont été mises en place au cours de ce travail de thèse, l'une étudie précisément la structure des répétitions du domaine central et l'autre propose une vision plus intégrative de la dystrophine et de ses mutations.

Le chapitre 2 exposera l'utilisation couplée de la modélisation et de la dynamique moléculaire effectuée sur une région précise de la dystrophine : le domaine central. Cette approche permet de pallier à l'absence de structure expérimentale. Dans un premier temps, l'étude des 24 répétitions natives de la dystrophine sera décrite. L'article correspondant est disponible en annexe C. Dans un deuxième temps, l'étude des répétitions tronquées sera présentée. Cette étude porte sur un fragment de dystrophine allant de la répétition 16 à 21 où la région correspondante du gène a été identifiée comme hautement mutagène. La modélisation puis la dynamique moléculaire permettent de comparer le fragment de protéine native avec quatre fragments de protéines tronquées identifiées chez des patients.

Le chapitre 3 présente une approche plus globale des conséquences des mutations en phase sur la fonctionnalité des dystrophines mutées. La base de données eDystrophin permet l'analyse d'une collection de plus de 200 mutations en phase. En effet, nous proposons une analyse automatisée et prédictive des conséquences d'une mutation sur les domaines de liaison et la structure tridimensionnelle de la protéine produite suite à ces mutations en phase. De plus, ces données sont mises en relation avec les données cliniques fournies par le Laboratoire de Biochimie et de Génétique de l'Institut Cochin. La base de données est librement accessible à l'adresse : <http://edystrophin.genouest.org/> et l'article correspondant est disponible en annexe D.

# Chapitre 2

## La modélisation des répétitions de la dystrophine

L'étude du domaine central de la dystrophine par des approches de modélisation moléculaire s'est déroulée en deux temps.

Dans un premier temps, une étude, à laquelle j'ai participé, a été mise en place sur les répétitions de la dystrophine native. Cette étude a permis de mieux caractériser les répétitions et de mettre en évidence leur hétérogénéité. Elle fait l'objet d'un article [Legrand 2011].

Dans un deuxième temps, nous avons étudié des répétitions tronquées observées chez des patients. Nous nous sommes alors tout particulièrement intéressés à la région comprise entre les répétitions 16 à 21 qui comprend le point chaud de mutations codé par les exons 45 à 53. Nous avons ainsi comparé le fragment de protéine native à quatre fragments de protéines tronquées correspondant à des mutations identifiées chez des patients. J'ai réalisé ce travail en collaboration avec Olivier Delalande et Emmanuel Giudice.

### 2.1 Matériel et Méthodes

#### 2.1.1 Prédiction de la structure 3D des répétitions

Le serveur I-TASSER [Roy 2010] a été utilisé pour générer les modèles par homologie. I-TASSER, un des meilleurs programmes de modélisation au concours CASP, combine différentes approches telles que l'alignement structural, la prédiction *ab initio* et l'échange de répliques (replicat exchange) par des simulations de Monte-Carlo (figure 2.1).

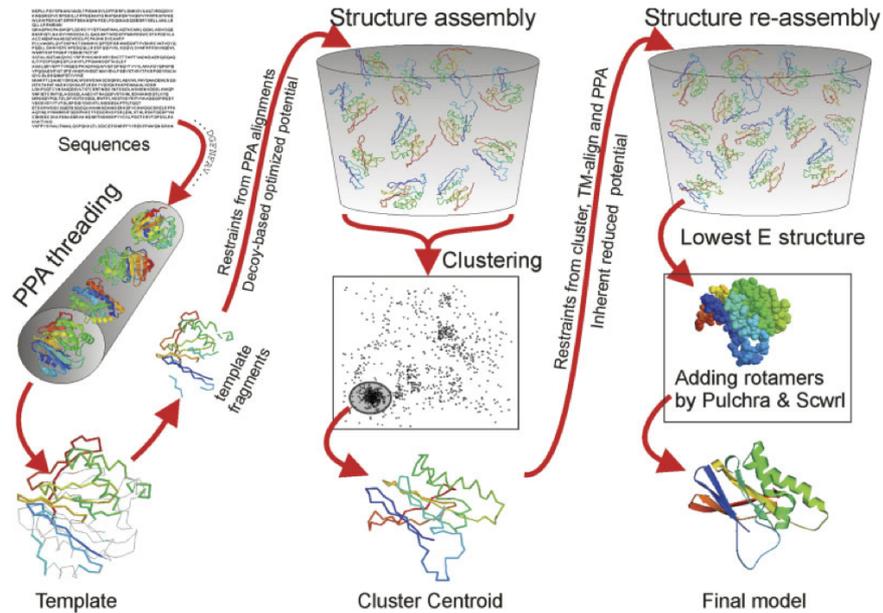


FIGURE 2.1 – Protocole d’I-TASSER pour prédire la structure d’une protéine [Zhang 2007]. La première étape est l’alignement structural (threading) sous la forme d’alignements profil à profil (PPA) de la séquence cible contre une bibliothèque représentative de la PDB. Cet alignement est réalisé avec différentes combinaisons de modèles de Markov, de profils PSI-BLAST et d’algorithmes d’alignement afin d’obtenir des structures-patrons. Les régions de la protéine cible s’alignant avec des structures-patrons sont modélisées. Les régions non alignées (principalement des boucles) sont modélisées par des méthodes *ab initio*. Les modèles des régions sont assemblés et subissent des échanges de répliques par simulations de Monte-Carlo. Les structures sont ensuite regroupées (clustering) et un modèle correspondant au centre du groupe est extrait. Ce modèle est ensuite affiné par la même méthode que lors de l’assemblage précédent. Le modèle de plus basse énergie est alors choisi comme modèle final.

I-TASSER propose un à cinq modèles structuraux par séquence soumise (voir annexe A pour un exemple de résultats). Ces modèles sont accompagnés d’un score qualitatif, le C-score (Eq. 2.1 [Zhang 2008]), qui varie de -5 à 2 : plus ce score est élevé, plus le modèle proposé est de qualité.

$$C - score = \ln \frac{M}{M_{tot}} * \frac{1}{RMSD} * \frac{\prod_{i=1}^4 Z(i)}{\prod_{i=1}^4 Z_0(i)} \quad (2.1)$$

où  $M$  est la multiplicité des structures dans le groupe (cluster) ;  $M_{tot}$  est le nombre total de structures I-TASSER du groupe ;  $RMSD$  est l’écart quadratique moyen entre le modèle final et la structure centrale du groupe ;  $Z(i)$  est le meilleur score de l’algorithme PPA et  $Z_0$  est un score spécifique au programme PPA qui distingue les bonnes et les mauvaises structures-patrons ( $Z_0(1) = 7,0$  ;  $Z_0(2) = 8,5$  ;  $Z_0(3) = 8,0$  ;  $Z_0(4) = 10,5$ ).

Les deux premiers termes de l’équation prennent en compte le degré de convergence du regroupement et le troisième terme prend en compte la qualité de l’alignement structural. L’étalonnage du serveur I-TASSER a été réalisé sur un lot de 300 protéines à un seul domaine, non homologues et représentatives de la PDB.

Bien qu’I-TASSER soit un des meilleurs serveurs aux concours CASP, il possède quelques

points faibles. En effet, il est optimisé pour des protéines comprenant entre 100 et 400 acides aminés et constituées d'un seul domaine. Nous avons donc pris en considération ces limitations pour générer nos modèles.

Afin d'identifier les séquences des répétitions de la dystrophine, trois outils de recherche de domaines ont été testés : Conserved domains [Marchler-Bauer 2009], Pfam [Finn 2010] et SMART [Letunic 2009]. Ces outils n'ont pas pu identifier l'ensemble des 24 répétitions de la dystrophine. En effet, Pfam, l'outil ayant identifié le plus de répétitions n'en a identifié que 16. Bien que les répétitions de la dystrophine soient homologues à la spectrine, leur motif est plus permissif que celui de la spectrine et peut expliquer le faible taux de réussite de Pfam. Cependant, ces outils ont permis d'identifier les autres domaines structuraux de la dystrophine et leurs résultats sont intégrés dans la base de données eDystrophin décrite dans le chapitre suivant.

Nous avons donc choisi d'utiliser l'alignement des répétitions réalisé par Winder [Winder 1995a] (figure 1.8) pour borner les séquences primaires correspondant aux répétitions natives. De plus, l'utilisation de cet alignement permet d'être en accord avec les expérimentations réalisées au sein de l'équipe. Les répétitions ont été soumises en tandem au logiciel de modélisation afin de conserver l'hélice commune à deux répétitions consécutives (linker) et ainsi chaque répétition est présente dans deux modèles, une fois en position N-terminale et une fois en position C-terminale. Ainsi, nous avons réalisé des modèles pour le tandem de répétitions R1-2, puis le tandem R2-3, etc... Les deux exceptions sont les répétitions entourant des charnières, les tandems R3-4 et R19-20 que nous avons volontairement omis, ces répétitions étant séparées respectivement par les charnières H2 et H3. Nous avons ainsi obtenu 21 modèles. Cette stratégie permet de modéliser chaque répétition et l'ensemble des linkers entre les répétitions.

Nous avons généré les modèles des répétitions tronquées afin d'étudier les jonctions des délétions. Nous avons donc choisi de modéliser les deux répétitions tronquées par la délétion ainsi que les répétitions précédente et suivante. Soit, pour la délétion des exons 45 à 53, on modélise la répétition nouvellement formée constituée des répétitions tronquées 17 et 21 ainsi que les répétitions 16 et 22. La séquence du fragment de protéine correspondante a été soumise à I-TASSER afin de générer le modèle par homologie.

Nous avons remarqué que lorsque les modèles comprennent une charnière (H1 à H4), la charnière est modélisée en hélices. Il s'agit là d'une faiblesse d'I-TASSER qui modélise de façon imparfaite les protéines multi-domaines. Ceci est particulièrement remarquable dans notre cas, puisqu'il existe peu de structures patrons de charnières alors que les structure patrons des répétitions de spectrine ont un bon score de similarité avec nos protéines. Nous avons donc choisi de modéliser les charnières séparément afin de ne pas biaiser le choix d'une structure patron. Ces charnières ont ensuite été incorporées au modèle des protéines

tronquées selon le protocole décrit dans la figure 2.2. Ces modèles composés de deux blocs sont ensuite minimisés avec le logiciel YASARA afin d'éviter les conflits stériques et les angles de torsion non permis pour obtenir un modèle relaxé.

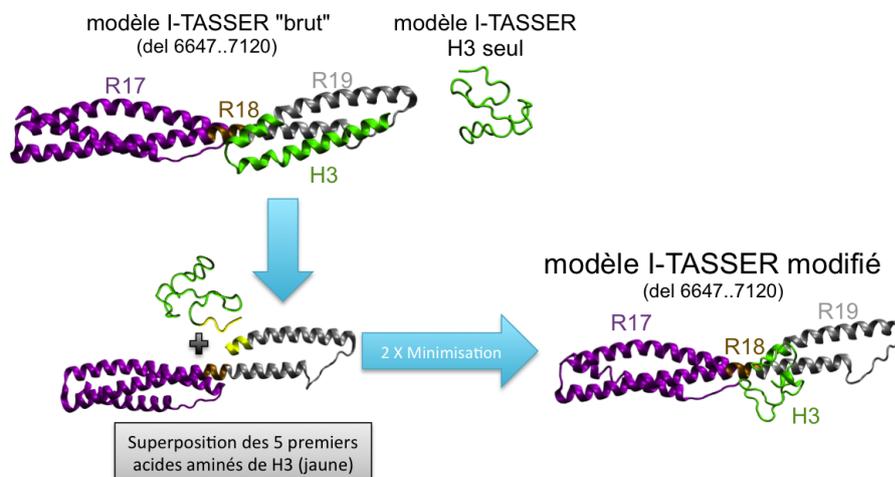


FIGURE 2.2 – *Protocole de modification d'un modèle contenant une charnière. Le fragment de protéine soumis à I-TASSER est notamment composé de la charnière H3 modélisée en hélices. La charnière est également soumise à I-TASSER séparément. Dans un deuxième temps, H3 est ôtée de la protéine tronquée à l'exception de ses 5 premiers acides aminés. Le modèle isolé de H3 est superposé avec la protéine tronquée sur les 5 premiers acides aminés puis les deux modèles sont rassemblés dans un même fichier de coordonnées PDB. Le nouveau modèle généré est ensuite minimisé deux fois avec le logiciel YASARA.*

Les modèles structuraux ont ensuite été soumis à plusieurs programmes de vérification de la qualité des modèles (PROCHECK [Laskowski 1993], ProSA-web [Wiederstein 2007] et Verify3D [Luthy 1992]).

L'ensemble des modèles générés est disponible dans la base de données eDystrophin décrite dans le chapitre 3.

## 2.1.2 La dynamique moléculaire

La dynamique moléculaire a été utilisée pour relaxer plusieurs modèles de répétitions natives et tronquées. Plus précisément, les modèles structuraux ont été optimisés en permettant des changements conformationnels autorisés. Je n'ai pas participé aux simulations de dynamiques moléculaires des répétitions natives, je ne détaillerai donc pas le protocole ici mais il est disponible dans l'article en annexe C. Le protocole d'analyse des répétitions tronquées que j'ai mis en place s'inspire cependant de celui-ci. Les simulations tout-atome et gros-grains ont été réalisées en présence d'un solvant explicite. Seules les protéines  $\Delta 45-49$  et  $\Delta 45-51$ , ont été, sur la période de cette thèse, soumises à des simulations gros-grains.

**Par convention dans cette thèse, le terme de structure sera employé pour parler des conformères générés par la DM bien que la structure de départ soit**

un modèle.

### 2.1.2.1 Les modèles de dystrophines tronquées choisis

La dystrophine étant trop grande pour être simulée en entier (3685 acides aminés), nous avons choisi de nous intéresser à la région comprenant le point chaud de mutation de la dystrophine : la région allant de la répétition 16 à 21 du domaine central. Trois régions d'interactions ont été caractérisées dans le fragment de protéine étudié ici : la fin du deuxième domaine de liaison à l'actine, le domaine de liaison à la nNOS au niveau des répétitions 16 et 17 et une partie du domaine de liaison aux lipides entre les répétitions 16 et 19.

Nous avons effectué les simulations de DM sur cinq protéines : la dystrophine native R16-21 et quatre dystrophines tronquées observées chez des patients et nommées en fonction des exons supprimés (figure 2.3 et tableau 2.1). L'alignement des séquences primaires des cinq protéines est en annexe B. Les délétions de ces quatre protéines commencent toutes à l'exon 45 mais se terminent différemment. Les répétitions tronquées formées au point de jonction commencent donc toutes par la région N-terminale de la répétition 17. Ainsi nous avons choisi de noter ces répétitions R17//X où X est la répétition tronquée se situant après la délétion.

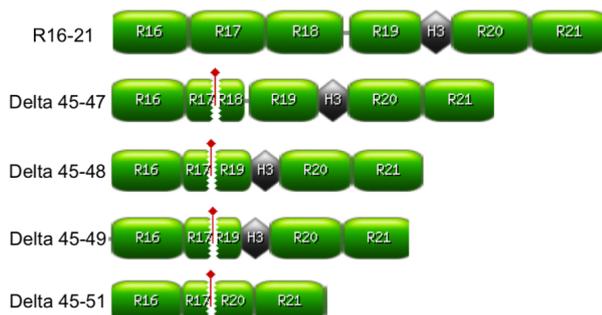


FIGURE 2.3 – Collection de protéines simulées en dynamique moléculaire. Les spots rouges indiquent la zone de troncature suite à des délétions d'exons.

Protéine	Description	Nb résidus	Nb atomes
R16-21	R16 - R17 - R18 - R19 - H3 - R20 - R21	694	11419
$\Delta$ 45-47	R16 - R17//18 - R19 - H3 - R20 - R21	536	8827
$\Delta$ 45-48	R16 - R17//19 - H3 - R20 - R21	474	7810
$\Delta$ 45-49	R16 - R17//19 - H3 - R20 - R21	440	7265
$\Delta$ 45-51	R16 - R17//20 -R21	326	5435

Tableau 2.1 – Description des modèles de dystrophines simulées en dynamiques moléculaires.

Les modèles par homologie décrits dans la 2.1.1 ont été rallongés pour réaliser les simulations de DM. I-TASSER est optimisé pour modéliser des protéines de 300 à 400 acides aminés, nous avons donc choisi de reconstruire les protéines simulées en DM à partir de petits fragments de protéines décrits dans la section 2.1.1. Ainsi, nous avons construit les cinq protéines d'intérêt en combinant les modèles par homologie des tandems de répétitions, des répétitions tronquées et de la charnière H3, tous soumis séparément au serveur I-TASSER.

Afin de reconstruire ces modèles, nous avons superposé les hélices B et C des répétitions communes de chaque tandem de répétition puis nous avons réuni les deux modèles en coupant au niveau de la boucle BC comme décrit dans la figure 2.4. Nous coupons dans les régions les plus flexibles afin de maintenir les hélices intactes ainsi que les linkers. Ainsi, la première étape de la construction de la protéine R16-21 a été la construction de R16-18 à partir des tandems de répétitions R16-17 et R17-18. Les deux tandems de répétitions successifs ont alors été superposés sur les hélices B et C de R17 puis les protéines ont été coupées au niveau de la boucle BC pour réunir les deux tandems de répétition et former un modèle à 3 répétitions. Le modèle ainsi formé a subi une courte minimisation dans le logiciel YASARA afin de l'optimiser.

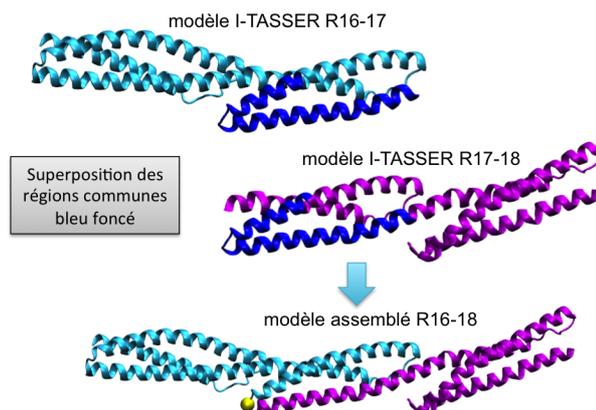


FIGURE 2.4 – Protocole de construction d'un modèle avec trois répétitions à partir de deux modèles de tandems de répétitions. La moitié de l'hélice B et l'hélice C de R17 communes au deux tandems de répétitions R16-17 et R17-18 sont colorées en bleu foncé. Les deux modèles sont superposés sur cette région. Chaque modèle est alors coupé au niveau de la boucle BC (en jaune) entre les deux hélices afin de réunir les deux tandems. Ainsi, un modèle à trois répétitions est construit.

Le même protocole a ensuite été appliqué pour ajouter R19 à partir du tandem R18-19. Ensuite, la charnière H3 a été ajoutée à R16-19 en utilisant une protéine tronquée possédant la charnière H3 et l'hélice C de R19 dont la méthode de construction est décrite dans la section 2.1.1. Les quatre derniers résidus de l'hélice C de R19 provenant de R16-19 et de la protéine tronquée ont été superposés puis les coordonnées de la charnière ont été copiées dans le modèle de R16-19 afin de créer le modèle R16-H3. De la même manière, la charnière a été ajoutée au tandem de répétition R20-21. Enfin, les deux charnières H3 des modèles R19-H3 et H3-R21 ont été superposées sur les quatre premiers acides aminés de H3 afin de réunir les deux modèles dans le même fichier de coordonnées pour la protéine R16-21.

Les modèles des protéines tronqués ont été construits sur le même principe, à l'exception de  $\Delta 45-51$  qui, du fait de sa petite taille, a été soumise directement à I-TASSER. Les modèles proposés par I-TASSER ont donc été allongés en reprenant les régions manquantes dans le modèle de R16-21 nouvellement créé. Ainsi, le modèle initial de  $\Delta 45-47$  comprenant les répétitions R16, R17//18 et R19 a été allongé en ajoutant les régions H3, R20 et R21 en superposant les quatre derniers acides aminés des deux modèles de R19. Le même principe a été utilisé pour construire les modèles initiaux de  $\Delta 45-48$  et  $\Delta 45-49$ .

### 2.1.2.2 La dynamique moléculaire tout-atome avec NAMD

#### Les calculateurs utilisés

Nous disposons de quatre calculateurs pour réaliser les simulations de dynamiques moléculaires : deux clusters de calculs locaux (Dolmen et Menhir) utilisés pour réaliser les simulations de DM de la plus courte protéine tronquée ( $\Delta 45-51$ ) et certaines analyses post-productions, et deux clusters de calculs nationaux (Jade et Vargas) du GENCI-[CINES/IDRIS] utilisés pour simuler les quatre autres protéines (tableau 2.2). L'utilisation de ces calculateurs a permis de paralléliser les calculs.

Caractéristiques \ Calculateur	Dolmen	Menhir	Jade	Vargas
Noeuds	12	8	2880	112
CPU / noeud	4 / 8	12	8	32
Fréquence (GHz)	2,6 / 2,1	2,4	2,8 / 3	4,7
Total CPU	64	96	23040	3584
Parallélisation	12 CPU	16 CPU	512 CPU	256 CPU

Tableau 2.2 – Description des calculateurs utilisés pour la dynamique moléculaire.

#### La préparation des systèmes

Les protéines ont été simulées en présence de molécules d'eau et d'ions en utilisant le programme NAMD (2.7b2 ou 2.8) [Phillips 2005] et le champ de force CHARMM27 [Zhu 2012]. Les modèles initiaux ont été solubilisés dans une boîte d'eau rectangulaire générée grâce à l'extension *Solvate* du logiciel VMD [Humphrey 1996]. Les modèles ont tous été alignés selon l'axe x. Des molécules d'eau de type TIP3P (modèle à trois sites d'interactions correspondant aux trois atomes de la molécule) ont été ajoutées en veillant à conserver au moins 25 Å d'épaisseur autour de la molécule selon l'axe principal x et 35 Å d'épaisseur autour de la molécule selon les axes y et z (tableau 2.3). Afin de s'affranchir des effets de bord, les simulations sont réalisées en conditions de limites périodiques, c'est-à-dire que la boîte d'eau contenant le modèle est virtuellement répliquée dans les trois dimensions. Ainsi, un atome sortant de la boîte par une face est réintroduit dans la boîte par la face opposée avec la même vitesse.

Protéine	Nb de résidus	Nb de $Na^+$	Nb total d'atomes	Taille initiale de la boîte (en Å)
R16-21	694	18	652 306	170,22 X 117,48 X 335,71
$\Delta$ 45-47	536	13	464 972	133,25 X 114,22 X 315,13
$\Delta$ 45-48	474	7	391 994	135,55 X 113,16 X 264,01
$\Delta$ 45-49	440	6	328 751	92,13 X 169,23 X 218,66
$\Delta$ 45-51	326	3	223 727	113,25 X 109,93 X 186,185

Tableau 2.3 – Description des modèles de dystrophines simulés en dynamique moléculaire.

Les modèles ont ensuite été ionisés par l'intermédiaire de l'extension *Autoionize* de VMD afin de neutraliser le système en ajoutant un minimum d'ions réduisant ainsi les probabilités de conflits entre les protéines et les ions. Toutes nos molécules étant chargées négativement, seuls des ions  $Na^+$  sont ajoutés.

### L'étape de minimisation

Chaque système a été minimisé par la méthode du gradient conjugué en deux étapes de 10 000 pas chacune. La première étape a été effectuée en fixant les positions des atomes de la protéine afin d'ajuster les positions des molécules d'eau et des ions autour de la protéine. La seconde étape a été effectuée en laissant libre l'ensemble des atomes afin de minimiser l'ensemble du système.

**La structure minimisée obtenue est la structure considérée comme initiale dans l'ensemble de ce chapitre.**

### Les étapes de chauffage et de production

Les systèmes sont ensuite chauffés pendant 50 ps pour atteindre la température de 310K. Ensuite des trajectoires d'au moins 30 ns ont été générées (tableau 2.5) avec un pas de temps de 2 fs grâce à l'algorithme SHAKE. Une structure est donc calculée toutes les 2 fs mais on ne sauvegarde qu'une structure toutes les 500 calculées, soit une structure toutes les picosecondes. Les interactions de van der Waals sont calculées avec un rayon de coupure de 12 Å mais ramenées à 0 entre 10 et 12 Å. Les interactions électrostatiques à longue distance sont calculées avec la méthode du Particule Mesh Ewald (PME) [Cheatham 1995] avec un maillage de  $1 \text{ \AA}^{-3}$ . La température et la pression sont maintenues constantes, à 310 K et 1 atm, durant toute cette étape grâce à la méthode de Berendsen. Le volume est quant à lui variable.

Protéine	Nb total d'atomes	Temps de simulation (ns)	Nb de structures enregistrées	Temps de calcul sur 1 CPU (heures)
R16-21	652 306	40	40 000	72 000
$\Delta$ 45-47	464 972	30	30 000	63 000
$\Delta$ 45-48	391 994	35	35 000	32 000
$\Delta$ 45-49	328 751	30	30 000	32 000
$\Delta$ 45-51	223 727	30	30 000	27 000

Tableau 2.4 – *Caractéristiques des trajectoires de DM.*

### 2.1.2.3 La dynamique moléculaire gros grain avec GROMACS

#### La préparation et la minimisation des systèmes

Les simulations de dynamique moléculaire gros-grain ont été réalisées avec le logiciel GROMACS [Hess 2008] en utilisant le champ de force MARTINI [Marrink 2007] sur les protéines  $\Delta$ 45-49 et  $\Delta$ 45-51. Le fichier de départ utilisé pour lancer une dynamique gros-grain est le fichier de sortie de la minimisation NAMD présenté ci-dessus.

Contrairement, aux simulations tout-atome, les structures secondaires sont calculées par l'outil DSSP puis fixées au départ dans le fichier de topologie [Frishman 1995]. Les fichiers ont été préparés par l'intermédiaire des extensions de GROMACS.

Dans un premier temps, pour  $\Delta$ 45-49, une boîte cubique a été ajoutée autour de la protéine en veillant à conserver un écart de 25 nm (selon les 3 axes) entre la protéine et les bords de la boîte. Une courte minimisation de la protéine dans le vide par la méthode du gradient descendant a été réalisée pendant 10 000 pas maximum.

Ensuite, des molécules d'eau gros-grain (soit l'équivalent de quatre molécule d'eau tout-atome) ont été ajoutées à la boîte. Puis l'ensemble a été minimisé par la méthode du gradient descendant pendant 25 000 pas maximum.

Protéine	Nb total d'atomes	Temps de simulation ( $\mu$ s)	Nb de structures enregistrées	Taille de la boîte (en Å)
$\Delta$ 45-49	126866	5	100 000	247 X 247 X 247
$\Delta$ 45-51	128965	5	100 000	250 X 250 X 250

Tableau 2.5 – *Caractéristiques des trajectoires de DM.*

#### Les étapes de chauffage et de production

Le système est chauffé jusqu'à 303 K. Les simulations sont réalisées en conditions périodiques avec un pas de temps de 25 fs pendant 5  $\mu$ s. Une structure est donc calculée toutes les 25 fs mais on ne sauvegarde qu'une structure sur 2000, soit une structure toutes les 50

ps. La température et la pression sont maintenues constantes, à 303 K et 1 atm, durant toute la simulation grâce à la méthode de Berendsen. A l'issue de l'étape de production, les trajectoires contiennent 100 000 structures.

## 2.1.3 L'analyse des modèles des répétitions

### 2.1.3.1 Les tandems de répétitions

Les propriétés hydrophobes et électrostatiques des répétitions natives ont été analysées et une relaxation de quatre de ces modèles (R1-2, R10-11, R15-16 et R23-24) a été réalisée par dynamique moléculaire (cf. annexe C). Les modèles ont été analysés graphiquement en utilisant PyMOL et VMD. J'ai participé à cette étude en réalisant l'analyse des propriétés électrostatiques des modèles.

### 2.1.3.2 Les répétitions tronquées

Les écarts quadratiques moyens (RMSD) des carbones  $\alpha$  de chaque structure extraite de la trajectoire de DM par rapport à la structure de son modèle initial ont été calculés grâce à l'outil *g\_rms* de GROMACS. Par la suite, les analyses des trajectoires ont été réalisées sur les 22 dernières nanosecondes des simulations, qui apparait comme la période de stabilité la plus longue, commune à l'ensemble des trajectoires après analyse des tracés de RMSD. La fluctuation des structures au cours des DM a été calculée par l'intermédiaire de l'extension *g\_rmsf* de GROMACS.

Les analyses visuelles ont été réalisées avec VMD et les structures secondaires ont été calculées grâce à l'extension STRIDE [Frishman 1995].

Le regroupement des DM tout-atome en structures représentatives (cluster) de chacune des trajectoires de DM a été réalisé grâce à l'outil Ptraj inclut dans AmberTools 12. Les trajectoires ont ainsi été réduites en utilisant l'algorithme des k-moyennes (k-means) basé sur un critère de distance entre les structures de façon à réduire la trajectoire en trois groupes et à extraire une structure représentative moyenne par groupe (soit la structure de la trajectoire la plus proche en RMSD de la structure moyenne calculée à partir de l'ensemble des structures constituant le groupe). Le regroupement des DM gros-grains en structure représentatives a été réalisé par l'extension *g\_cluster* de GROMACS en utilisant l'algorithme de Jarvis Patrick. Cet algorithme utilise l'approche des plus proches voisins pour créer plusieurs groupes (non prédéterminés à l'avance contrairement à Ptraj). Les trajectoires gros-grain balayent plus de conformations possibles que les trajectoires tout-atome et il est donc nécessaire d'utiliser un algorithme plus élaboré que celui des k-moyennes pour les regrouper.

L'analyse en composante principale qui a permis d'extraire les mouvements principaux des protéines a été réalisée par l'intermédiaire des outils *g\_covar* et *g\_anaeig* de GROMACS.

Les potentiels de surface électrostatiques et hydrophobes ont été calculés par l'intermédiaire des outils APBS [Baker 2001] et PLATINUM [Pyrkov 2009], respectivement.

## 2.2 L'étude des répétitions natives

### 2.2.1 L'analyse des séquences primaires

Les séquences des 24 répétitions de la dystrophine ont un faible score d'identité mais leur alignement révèle la présence de motifs en heptade où les résidus en positions a et d, correspondant aux résidus hydrophobes, sont généralement conservés. Cependant, on observe une permissivité dans ces motifs entraînant le décalage d'un résidu hydrophobe en position +1 ou -1 et/ou la présence d'un résidu polaire en position a ou d.

Contrairement à ce qui est observé dans les répétitions de spectrines, on remarque la présence de prolines dans la région entre deux répétitions successives. Sur les 21 modèles de répétitions en tandem, sept structures conservent le motif en heptade au niveau de cette jonction.

Le logiciel de prédiction de structure secondaire PSIPRED prédit que les répétitions en tandem présentent entre 79 et 91% d'hélices  $\alpha$ . Cette prédiction est en accord avec les données expérimentales de dichroïsme circulaire précédemment réalisées [Le Rumeur 2003, Legardinier 2009a, Mirza 2010]. Cependant, PSIPRED prédit une baisse du taux d'hélicité au milieu des hélices B de chaque répétition.

### 2.2.2 L'analyse des modèles structuraux

I-TASSER a identifié des structures de répétitions de spectrine comme patron pour générer les modèles de répétitions de dystrophine. Les C-score fournis par I-TASSER oscillent entre -0.72 et 0.76 indiquant un repliement correct des modèles. De plus, les modèles ont été soumis à plusieurs logiciels qui évaluent leur qualité et les résultats confirment que les modèles sont de haute qualité.

Les modèles forment des faisceaux de trois hélices allongés d'environ 100 X 50 Å (figure 2.5). Chaque répétition est composée de 3 hélices  $\alpha$  A, B et C reliées par deux boucles AB et BC, le tout organisé en faisceau de trois hélices. Ces répétitions s'enchaînent et sont liées par une hélice commune formée par l'hélice C de la première répétition et l'hélice A de la suivante. La région au centre de l'hélice commune est appelée le linker. Le coude dans les hélices B prédit par PSIPRED est présent dans toutes les répétitions à l'exception de R9 (modèle R9-10) et R14 (modèle R14-15). Ce coude est également présent dans la structure cristallographique très récemment obtenue de R1 [Muthu 2012]<sup>1</sup>.

---

1. Je n'ai pas pu analyser la structure cristallographique en détail car elle n'est pas encore publiquement accessible.

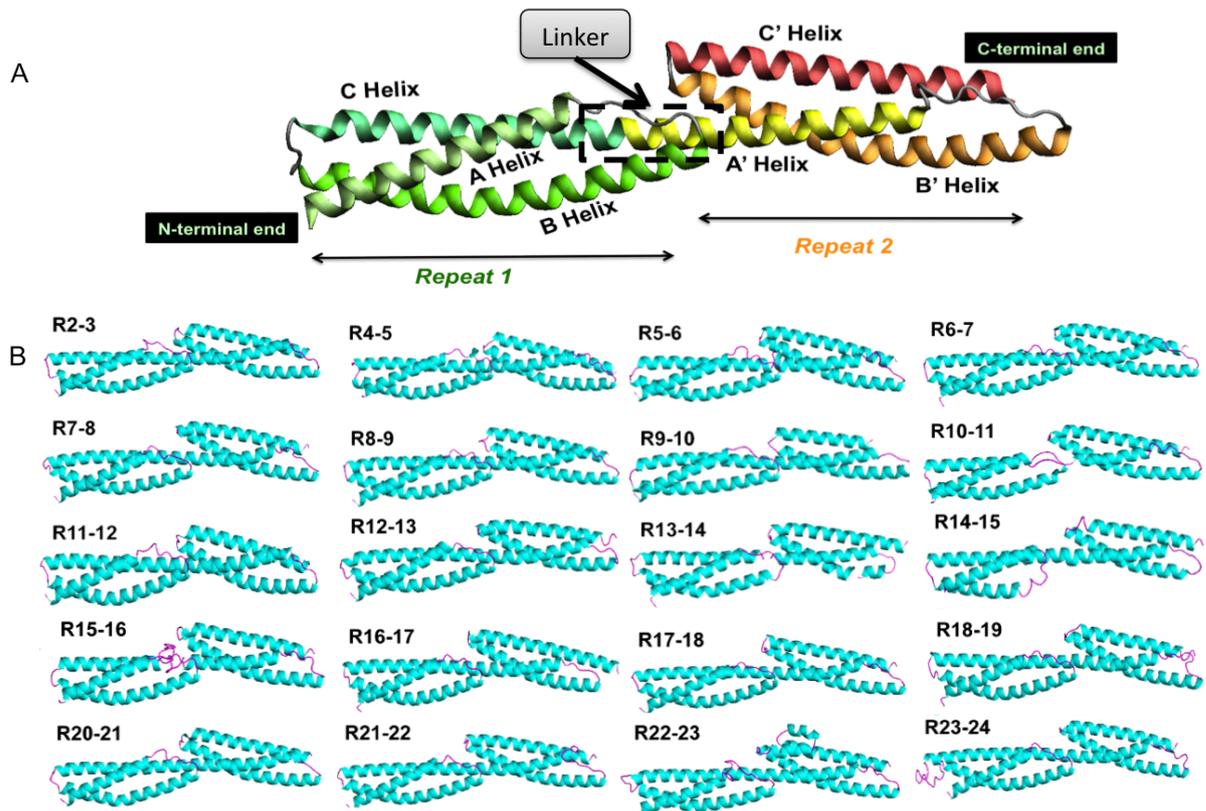


FIGURE 2.5 – A. Organisation de la structure tertiaire du modèle de répétition en tandem R1-2 obtenu par I-TASSER . B. Modèles par homologie de la structure tertiaire des 20 autres modèles de répétition en tandem de la dystrophine.

Contrairement à ce qu'on aurait pu penser en se référant aux études sur la spectrine [Mirijanian 2007], les linker de dystrophine ne sont pas tous hélicaux et rigides. En effet, un linker hélicale a été observé dans 17 des 21 modèles. Les quatre linker non hélicaux sont, soit une boucle (R5-6 et R10-11), soit une longue boucle (R15-16), soit une cassure (R13-14). La présence d'une ou deux prolines dans les linker de R5-6, R10-11 et R13-14 peut expliquer l'absence d'un linker hélicale. Une insertion, précédemment décrite [Menhart 2006], de 20 acides aminés entre les répétitions 15 et 16 peut expliquer la formation d'une boucle dans notre modèle.

Les linker sont situés entre deux faisceaux d'hélices et interagissent avec la boucle AB du premier faisceau et avec la boucle BC du deuxième. Cependant, on observe que ces boucles peuvent être à des distances variables du linker en fonction de la taille des hélices B. On observe, par exemple, que l'hélice B de R14 (modèle R14-15) est plus courte que les autres et place la boucle AB loin du linker diminuant le nombre de contacts entre le linker et la boucle. A l'opposé, on remarque que dans le modèle R4-5 la boucle AB de R4 est longue et recouvre parfaitement le linker, suggérant des interactions entre le linker et la boucle AB mais aussi avec la fin de l'hélice B. Cette hétérogénéité des linker de la dystrophine pouvant suggérer des propriétés dynamiques différentes.

### 2.2.3 Les propriétés de surface des modèles structuraux

La surface des répétitions en tandem de dystrophine est majoritairement hydrophile ( $66,4 \pm 3,5$  %) comme c'est le cas pour la plupart des protéines cytoplasmiques (figure 2.6). Cependant, des différences de potentiel hydrophobe ont été observées entre les différents tandems. Ainsi, les tandems R13-14 (27% d'hydrophobicité) et R14-15 (27,5% d'hydrophobicité) sont les plus hydrophiles alors que le tandem R18-19 (40% d'hydrophobicité) est le plus hydrophobe. Plusieurs patches et sillons hydrophobes pouvant être des sites d'interaction ont été observés dans tous les tandems, particulièrement dans R2-3, R5-6 et R10-11.

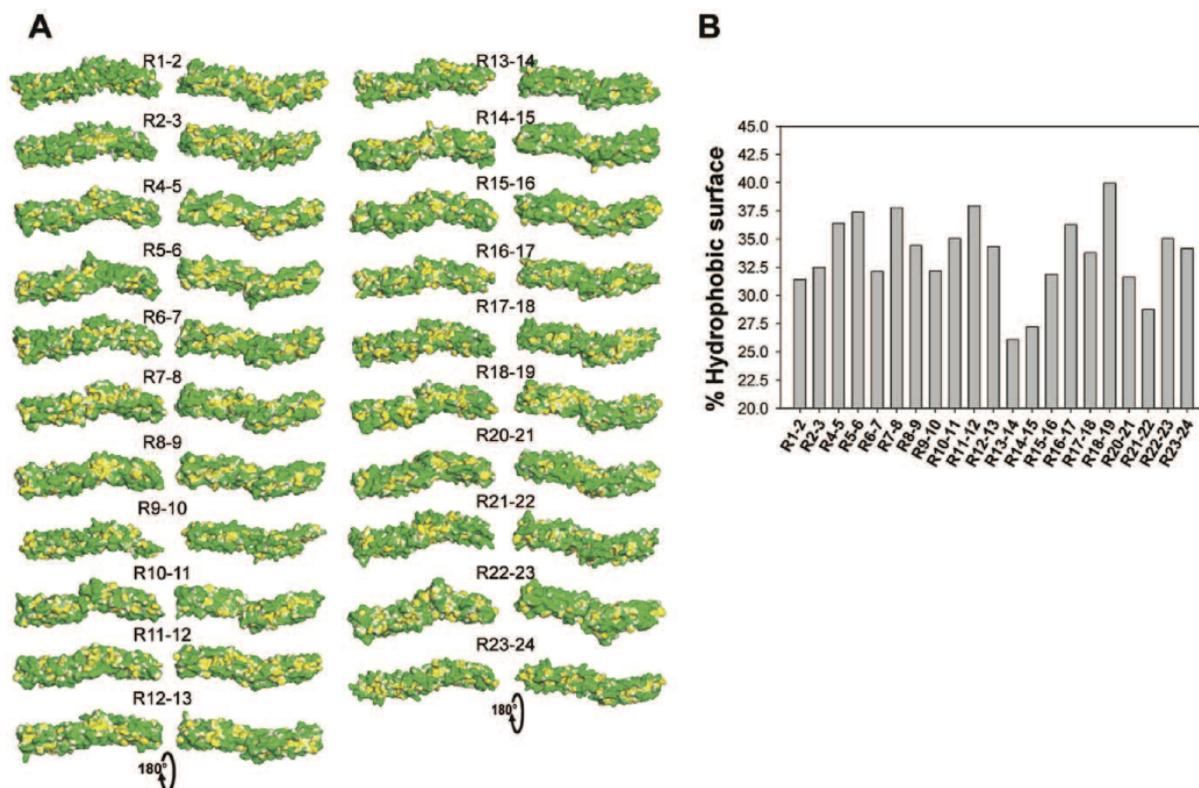


FIGURE 2.6 – Potentiel d'hydrophobicité de surface des répétitions en tandems. Les potentiels ont été calculés par l'intermédiaire du logiciel PLATINUM. A. L'échelle d'hydrophobicité utilisée est vert - blanc - jaune, où le vert représente les régions les plus hydrophiles. Les tandems de répétition sont représentés avec l'extrémité N-terminale à gauche. B. Histogramme du pourcentage de surfaces hydrophobes de chaque tandem de répétition.

L'analyse par APBS [Baker 2001] des propriétés électrostatiques des répétitions met également en évidence une grande hétérogénéité entre les différentes répétitions. L'utilisation de PIPSA [Richter 2008] a permis de grouper les répétitions en fonction de leur potentiel électrostatique de surface (cf. annexe C). Ainsi, on peut distinguer 4 groupes de répétitions en tandem avec des propriétés électrostatiques différentes (figure 2.7). On remarque que les groupes ne sont pas formés de répétitions successives.

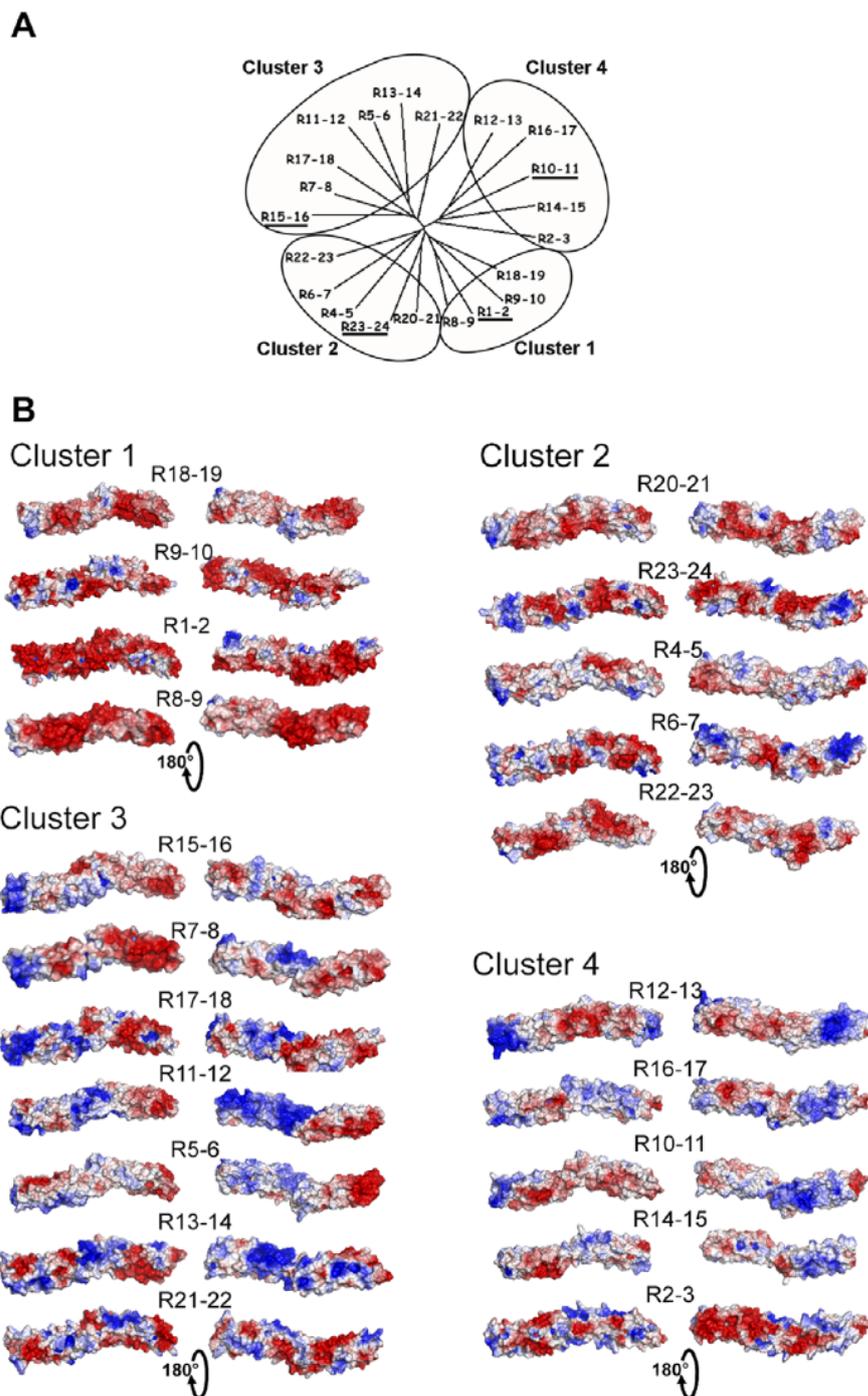


FIGURE 2.7 – *Potentiel de surface électrostatique des répétitions en tandems et regroupement des tandems de répétitions par PIPSA. A. Dendrogramme généré par PIPSA. Les tandems ont été regroupés en quatre groupes distincts en fonction de leurs potentiels de surface électrostatique. Les quatre répétitions en tandem simulés en dynamique moléculaire sont soulignés. B. Représentation des potentiels électrostatiques projetés sur la surface accessible au solvant de chaque tandem de répétition. Les modèles sont colorés en fonction du potentiel électrostatique calculé par APBS en présence de 50mM d'ions. Le gradient de couleur utilisé est rouge (négatif) - blanc - bleu (positif).*

La définition de ces régions avec des propriétés différentes est en accord avec les différentes zones d'interactions aux lipides et aux protéines partenaires définies antérieurement à l'exception des régions fortement chargées négativement R1-3 et R18-19 qui, curieusement, présentent de fortes propriétés de liaisons aux lipides anioniques [Legardinier 2009b]. La ré-

gion R10-17 possède de grands patchs chargés positivement et négativement et a la capacité de lier à la fois les lipides anioniques et zwitterioniques (caractérisés par des charges opposées). Les deux régions les moins chargées peuvent, soit lier les lipides dans le cas de R4-7, soit ne pas les lier dans le cas de R20-24 [Legardinier 2008b].

## 2.2.4 La relaxation par dynamique moléculaire des modèles

Quatre modèles présentant une structuration différente des linkers et appartenant à différents groupes définis par PIPSA (R1-2, R10-11, R15-16 et R23-24) ont été relaxés par une approche de dynamique moléculaire. La DM permet de s'éloigner du biais lié à l'utilisation de patrons et de proposer une structure affinée. Elle permet également d'étudier l'évolution des propriétés dynamiques des répétitions de dystrophine au cours du temps. Une telle approche avait déjà été utilisée pour analyser les structures de spectrines  $\alpha$  et  $\beta$  [Mirijanian 2007].

Les quatre modèles ont été simulés en solvant explicite en présence de 150 mM d'ions. Les modèles ont tout d'abord subi une minimisation par gradient conjugué puis ont subi une étape de chauffage à 310 K pendant 50 ps. Enfin, la phase de production a été effectuée pendant 31 ns en condition périodique avec un pas de temps de 2 fs.

La structure en faisceau de trois hélices de chaque répétition des tandems est conservée au cours de la DM. Cependant, trois des quatre modèles maintiennent une structure globale relativement similaire aux modèles initiaux alors que le modèle R10-11 forme rapidement un important coude au niveau du linker non hélical. Les modèles R1-2 et R23-24 conservent leur linker hélical au cours de la DM et le modèle R15-16 montre une légère différence de géométrie de la répétition 16 par rapport à la répétition 15 (figure 2.8).

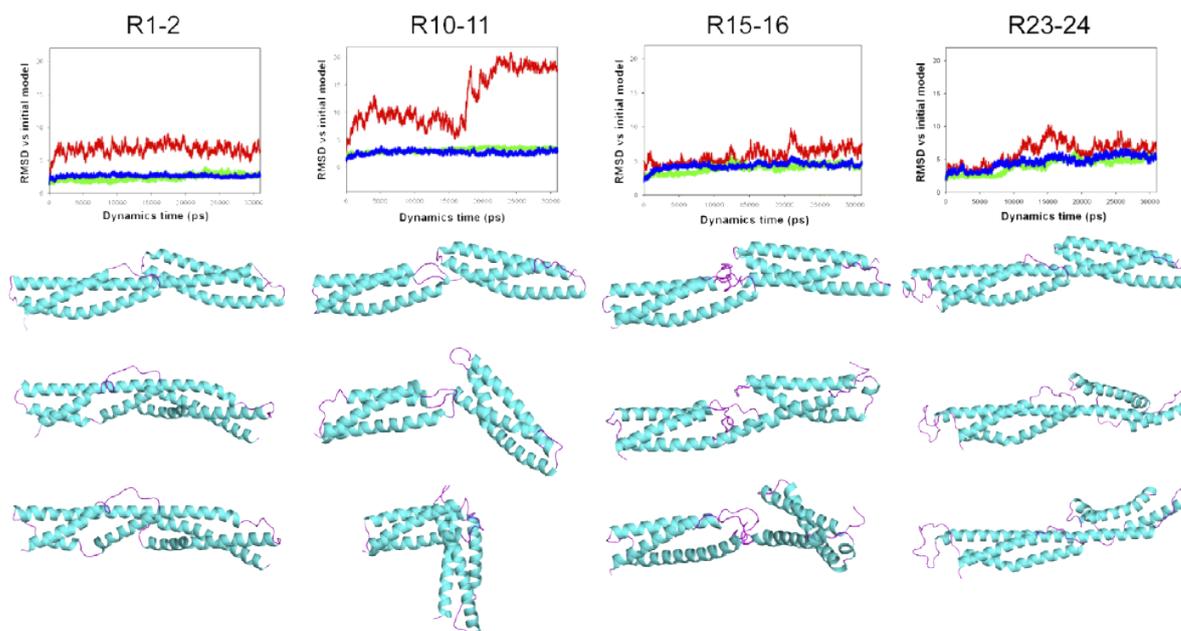


FIGURE 2.8 – *Changements structuraux observés dans les quatre tandems de répétitions pendant les simulation de DM. La fluctuation quadratique moyenne (RMSD) est calculée entre les  $C\alpha$  au cours des DM. La courbe rouge représente le tandem en entier, la courbe verte représente la répétition N-terminale et la courbe bleu représente la répétition C-terminale. Deux structures représentatives (C1 et C2) ont été extraites de chaque DM et sont représentées sous les graphes de RMSD avec les modèles initiaux.*

Nous observons dans la figure 2.9 que les régions de plus haute flexibilité dans les modèles se situent au niveau des boucles AB et BC situées entre les hélices et au niveau des linkers, hélicaux ou non. Dans les modèles R1-2 et R15-16, les hélices A et B de la première répétition les hélices B et C de la seconde se retrouvent du même côté du linker par un effet de torsion des deux répétitions le long de l'axe du linker.

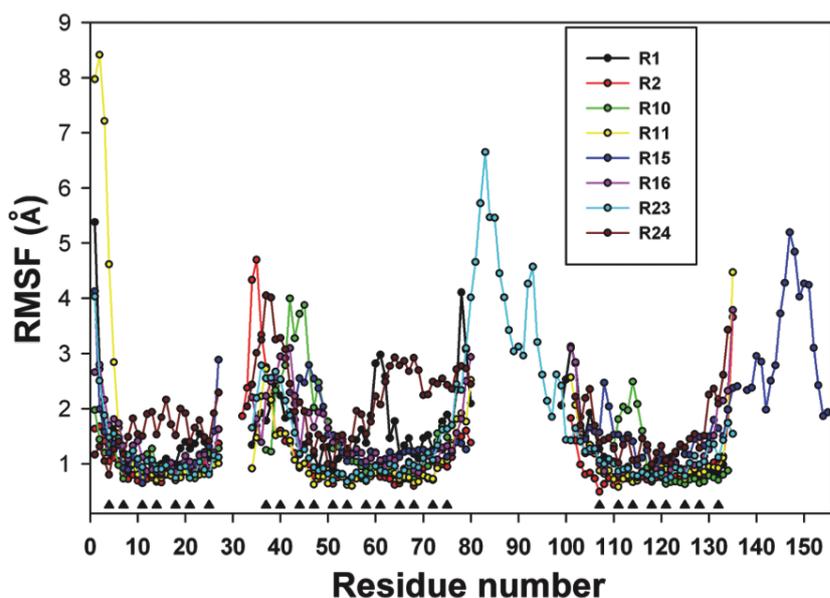


FIGURE 2.9 – Fluctuation des carbones  $\alpha$  de chaque résidu au cours des DM. La fluctuation des 8 répétitions issues des 4 tandems de répétition est superposée en fonction de la position des résidus a et d des heptades indiquée par des triangles noirs.

De manière intéressante, ces propriétés dynamiques différentes peuvent être directement reliées à la présence connue de domaines de liaison spécifiques. Ainsi, dans R1, il existerait un motif senseur de courbure (amphipathic lipid packing sensor ALPS) à la fin de l'hélice B et la rotation rapide de R2 dans le tandem R1-2 permettrait de le placer [Drin 2007] vis à vis des lipides de la membrane suggérant un contrôle dynamique de l'interaction avec les membranes courbées.

Le même type de coude que celui observé dans le tandem R10-11 a été observé dans le domaine de liaison à l'ankyrine des répétitions 14 et 15 de la spectrine [Stabach 2009]. Les auteurs ont émis l'hypothèse que ce coude permet de lier de façon dynamique l'ankyrine à la spectrine et ainsi de réguler la liaison. Nous pouvons imaginer le même type de mécanisme pour R10-11 qui pourrait ainsi contrôler dynamiquement la liaison de la région R11-17 à l'actine F.

Ainsi, le linker flexible dans R10-11 et le long linker de R15-16 peuvent être considérés comme de petites "jonctions" séparant la région R11-15 du reste du domaine central. Or, cette région correspond à la description initiale de l'ABD2 [Amann 1998] qui avait ensuite été étendue jusqu'à R17 [Rybakova 2006]. Ces jonctions pourraient isoler le motif R16-17 pour permettre la liaison de la nNOS et ainsi réguler l'interaction de R16-17 à la nNOS et les interactions de R11-15 avec les lipides et l'actine.

## 2.3 L'étude des répétitions tronquées

### 2.3.1 Les répétitions hybrides et les répétitions fracturées

L'analyse graphique des modèles de dystrophines tronquées nous a permis de distinguer deux types de structures à la jonction des répétitions : les répétitions hybrides qui conservent la succession des motifs en heptades des hélices et ont une structure proche des répétitions de la dystrophine native (ex : délétion d'exon 45-55) [Menhart 2006] et les répétitions fracturées qui ne conservent pas la succession des motifs en heptade et ne reconstituent pas une répétition complète (ex : délétion d'exon 45-47) (figure 2.10). Nous avons donc classé l'ensemble des délétions du domaine central selon ce critère.

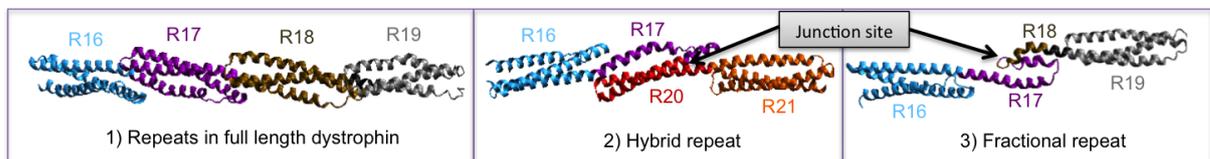


FIGURE 2.10 – 1. Répétitions 16 à 19 de la dystrophine native. 2. Répétition hybride dans la délétion d'exons 45-51. 3. Répétition fracturée dans la délétion d'exons 45-47.

La mise en évidence des répétitions hybrides et fracturées est une des innovations de la base de données eDystrophin présentée dans le chapitre 3.

### 2.3.2 L'analyse des modèles relaxés par DM

#### 2.3.2.1 La stabilité des protéines

##### Les modèles initiaux

Dans cette étude, la protéine native R16-21 est un fragment de la dystrophine correspondant au point chaud de la mutation. Comme attendu et comme observé dans l'étude des répétitions en tandem, son modèle par homologie est composé de six répétitions structurées en faisceau de trois hélices numérotées de 16 à 21. La plus courte répétition est R19 composée de 101 résidus puis R17 avec 107 résidus, R21 avec 109 résidus et enfin R16, R18 et R20, toutes les trois composées de 110 acides aminés. Les répétitions 19 et 20 sont séparées par une charnière intrinsèquement désordonnée (H3) composée de 41 acides aminés et une mini-charnière hélicoïdale de 7 résidus est située entre les répétitions 18 et 19. Nous constatons que l'ensemble des linkers de la protéine est structurée en hélice  $\alpha$ . Le modèle présente une forme de filament (figure 2.11).

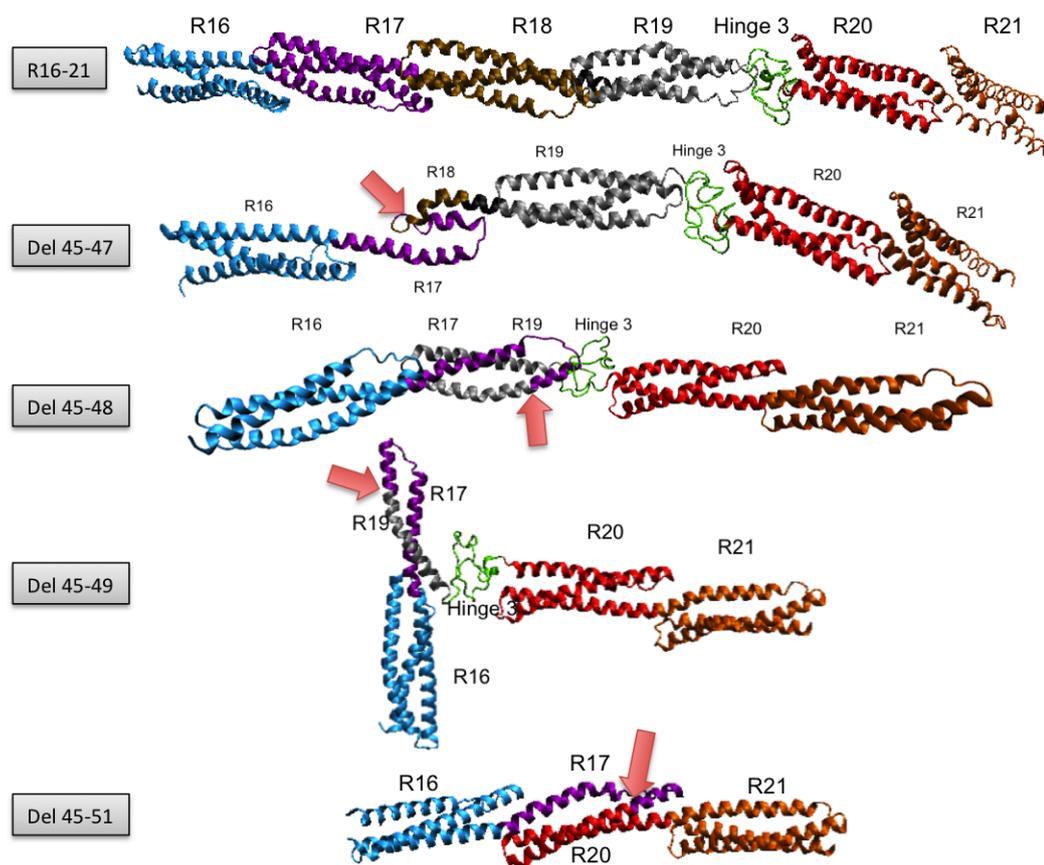


FIGURE 2.11 – Modèles par homologie initiaux utilisés pour effectuer les simulations de DM. La protéine R16-21 est la protéine native. Les quatre autres protéines correspondent à des délétions d'exons identifiées chez les patients. Les flèches rouges indiquent les jonctions des délétions. Le code couleur est le même pour les cinq protéines : R16 est en bleu, R17 en violet, R18 en ocre, la mini-charnière en noir, R19 en gris, H3 en vert, R20 en rouge et R21 en orange.

Le modèle de la protéine tronquée  $\Delta 45-47$  est composée de quatre répétitions entières (R16, R19, R20 et R21), de la charnière 3 et d'une répétition fracturée R17//18 constituée de 44 acides aminés issus de R17 et 15 acides aminés issus de R18 (cf. séquences en annexe B). Cette répétition fracturée de 59 acides aminés rompant la succession des motifs en heptade est modélisée par I-TASSER par trois petites hélices et deux boucles. Le modèle de la protéine tronquée présente une forme filamenteuse malgré la répétition fracturée.

Le modèle de la protéine tronquée  $\Delta 45-48$  est composée de trois répétitions entières (R16, R20 et R21), de la charnière 3 et d'une répétition hybride R17//19 constituée par l'hélice A et le début de l'hélice B de R17 (44 résidus) et de la fin de l'hélice B et l'hélice C de R19 (61 résidus), soit un total de 101 acides aminés comme la répétition R19 native. La répétition hybride conserve la succession des motifs en heptade et reconstitue une répétition entière. Le modèle présente une forme de filament.

Le modèle de la protéine tronquée  $\Delta 45-49$  est composée de trois répétitions entières (R16, R20 et R21), de la charnière 3 et d'une répétition fracturée R17//19 constituée par l'hélice A et le début de l'hélice B de R17 (44 résidus) et la fin de l'hélice C de R19 (27 résidus). Cette répétition fracturée de 71 acides aminés est composée de deux hélices  $\alpha$  et d'une boucle

se structurant en un faisceau de deux hélices. Ce faisceau d'hélice contraint la protéine à adopter une structure tertiaire en "T" très éloignée de la structure en filament de la protéine native.

Le modèle de la protéine tronquée  $\Delta 45-51$  est composée de deux répétitions entières (R16 et R21) et d'une répétition hybride R17//20 constituée par l'hélice A et le début de l'hélice B de R17 (44 résidus) et de la fin de l'hélice B et l'hélice C de R20 (64 résidus), soit un total de 108 acides aminés. La répétition hybride conserve la succession des motifs en heptade et reconstitue une répétition entière. La charnière 3 est absente de la protéine. Le modèle présente une forme de filament.

### **L'écart quadratique moyen au cours des trajectoires des DM**

Le tracé des écarts quadratiques moyens (RMSD) permet de suivre l'évolution des structures au cours des trajectoires en mesurant l'écart qui existe entre chaque structure de la trajectoire et la structure initiale. Ces courbes sont calculées après l'étape de chauffage qui apporte de l'énergie au système, favorisant les changements conformationnels. Après cette étape, les systèmes s'équilibrent durant une période plus ou moins longue, déterminée grâce au tracé du RMSD. Une fois l'équilibre atteint, on entre dans la phase de production qui est celle analysée dans la suite de ce chapitre ( section 1.3.3.3).

Nous observons que les structures s'éloignent rapidement de leurs conformations de départ (figure 2.12) et nous observons de grands mouvements de fluctuation au début des trajectoires de R16-21 et  $\Delta 45-48$  correspondant à de grands réarrangements structuraux. Nous avons donc choisi de prolonger ces trajectoires. En effet, en DM, les systèmes ne mettent pas tous le même temps pour atteindre un équilibre et nous remarquons que les protéines les plus courtes  $\Delta 45-49$  et  $\Delta 45-51$  l'atteignent plus rapidement. De plus, on observe que la structure de la protéine la plus courte et ne possédant pas de charnière ( $\Delta 45-51$ ) est celle qui reste la plus proche de sa structure initiale suggérant que les réarrangements structuraux observés dans les autres protéines sont causés en grande partie par la charnière.

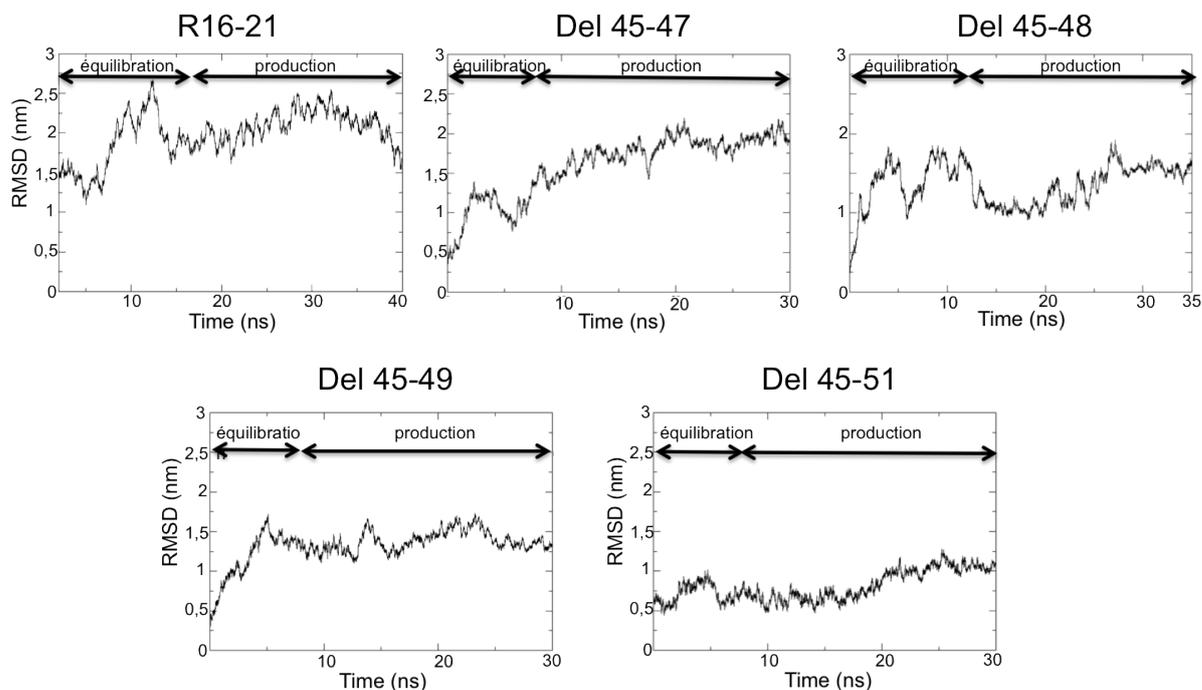


FIGURE 2.12 – *Ecart quadratique moyen entre les carbones  $\alpha$  au cours des DM.*

A partir des tracés de RMSD, nous avons évalué des temps d'équilibration propres à chaque protéine. Ainsi, les temps d'équilibration sont de 8 ns pour  $\Delta 45-47$ ,  $\Delta 45-49$  et  $\Delta 45-51$ , 13 ns pour  $\Delta 45-48$  et 18 ns pour R16-21. Ces temps ont été définis afin que tous les systèmes soient équilibrés en un temps minimum et que toutes les phases de production aient la même durée : 22 ns. Nous avons choisi de sauvegarder une structure toutes les picosecondes, chaque protéine est donc représentée, à ce stade, par 22 000 structures.

### La fluctuation des structures secondaires au cours des DM

L'étude de la fluctuation des structures permet de déterminer quelles régions des protéines sont les plus mobiles et de vérifier, dans notre cas, que les faisceaux d'hélices sont stables.

Comme attendu, les régions fluctuant le plus dans les cinq protéines se situent dans les boucles entre les hélices (figure 2.13). Par contre, la charnière (H3), essentiellement composée de boucles, est moins mobile particulièrement dans la trajectoire de R16-21 où la structure compacte proposée par I-TASSER semble limiter les réarrangements structuraux de H3. Globalement, on observe une fluctuation plus importante au niveau de la répétition R21, particulièrement dans R16-21. Cette fluctuation peut être artificielle et causée par l'absence de la répétition suivante dans la simulation. Toutefois, des données expérimentales indiquent que la répétition R21 serait peu stable [Mirza 2010].

Un des points de flexibilité important dans R16-21 est situé au niveau de la boucle BC de R19 où il est couplé avec une hausse de fluctuation de la mini-charnière hélicoïdale située entre R18 et R19. Or, la présence de cette mini-charnière augmente la taille du linker et l'éloigne du coeur du faisceau d'hélices, modifiant ainsi les interactions entre le linker et

la boucle BC. Il est donc raisonnable de supposer que ce linker allongé est plus mobile et entraîne la mobilité de la boucle BC de R19 qui se situe en face dans la structure 3D. La boucle AB de R19 est également mobile à cause de la charnière H3 se situant en face. La répétition 19 entourée de la mini-charnière et de la charnière H3 semble donc très mobile.

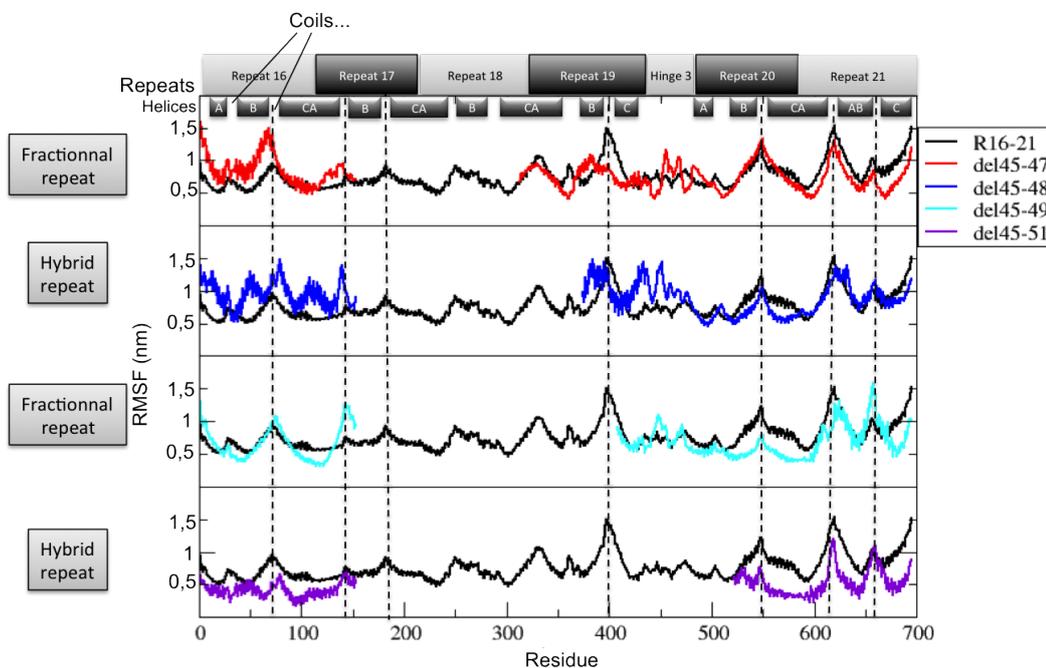


FIGURE 2.13 – *Fluctuation des carbones  $\alpha$  de chaque résidu au cours des DM. Les courbes de fluctuation des dystrophines tronquées sont découpées en fonction de la délétion et superposées sur la courbe de la dystrophine native R16-21 (en noir) dans chaque cadre. Les positions des répétitions et de la charnière (Hinge 3) sont indiquées en haut ainsi que les positions théoriques des hélices par des rectangles noirs. Les espaces entre les rectangles noirs sont les boucles reliant les hélices. Les lignes pointillées indiquent les régions les plus fluctuantes de la dystrophine native.*

En comparant les trajectoires de  $\Delta 45-47$  et R16-21, on observe que R16 et H3 sont plus flexibles dans la protéine tronquée alors que R20 et R21 ont des profils de flexibilité semblables. La répétition fracturée R17//18 formée au niveau de la jonction est un peu plus flexible du côté N-terminal. De plus, R19 de  $\Delta 45-47$  présente un profil de flexibilité différent de R16-21 avec une hélice B plus flexible que les boucles l'entourant. Il semblerait donc que la répétition fracturée perturbe la flexibilité de R19 et de R16.

$\Delta 45-48$  présente une fluctuation plus importante que R16-21 entre R16 et H3 mais similaire pour R20 et R21. Étrangement, on observe que les points de flexibilité sont décalés au niveau de R16. De plus, on observe et que la flexibilité de H3 augmente et que la répétition hybride formée au niveau de la jonction présente un profil de flexibilité plus élevée que les répétitions correspondantes dans R16-21.

Les intensités de fluctuation de R16-21 et  $\Delta 45-49$  sont du même ordre de grandeur mais des différences de fluctuation localisées apparaissent. Ainsi, la fluctuation augmente au niveau de la jonction, probablement car la répétition fracturée formée est peu stable. De plus, on observe une intensité de fluctuation légèrement augmentée dans H3 alors que R20

est globalement stabilisée.

En comparant  $\Delta 45-51$  avec R16-21, on observe que, globalement la protéine tronquée a une plus faible intensité de fluctuation et que les points de flexibilité sont situés dans les mêmes régions. On observe un décalage de flexibilité au niveau de la deuxième boucle de R16 et une hausse de flexibilité de  $\Delta 45-51$  au niveau de la jonction formant une répétition hybride.

D'une manière générale, on ne peut pas distinguer une protéine avec une répétition hybride d'une protéine avec une répétition fracturée grâce à leurs profils de fluctuation.

### **L'analyse en composante principale des mouvements des protéines**

L'analyse en composante principale permet d'extraire un à un les mouvements globaux des protéines au cours des trajectoires de DM par ordre décroissant d'occurrence au cours de la trajectoire. Cette analyse décompose la fluctuation globale présentée dans le paragraphe précédent et permet d'étudier uniquement les grands mouvements des protéines. Nous avons choisi d'analyser les quatre premiers mouvements de chaque protéine qui représentent au total plus de 80 % de l'intensité de mouvement de notre lot de protéines. Il est possible de décomposer chaque mouvement en sous-mouvements qui impliquent généralement des paires de répétitions. Ainsi, pour faciliter l'analyse, nous avons subdivisé chaque mouvement en sous-mouvements localisés dans les régions N-terminale, centrale et C-terminale, à l'exception des mouvements de la protéine  $\Delta 45-51$  plus courte, qui sont subdivisés en deux sous-mouvements.

Dans l'ensemble du jeu de protéines, nous pouvons distinguer quatre types de sous-mouvements bidirectionnels décrits dans la figure 2.14 : la dilatation d'un faisceau d'hélices, la formation d'un coude, l'élongation de la région et la torsion autour de l'axe du faisceau d'hélice. Bien que nous ayons choisi de séparer ces mouvements, il est important de noter qu'ils sont liés. Ainsi, un mouvement localisé à une extrémité de la protéine peut se répercuter sur toute la protéine. Par exemple, la formation d'un coude entraîne soit une torsion des faisceaux d'hélices avoisinants, soit un mouvement d'élongation-contraction d'une partie de la protéine.

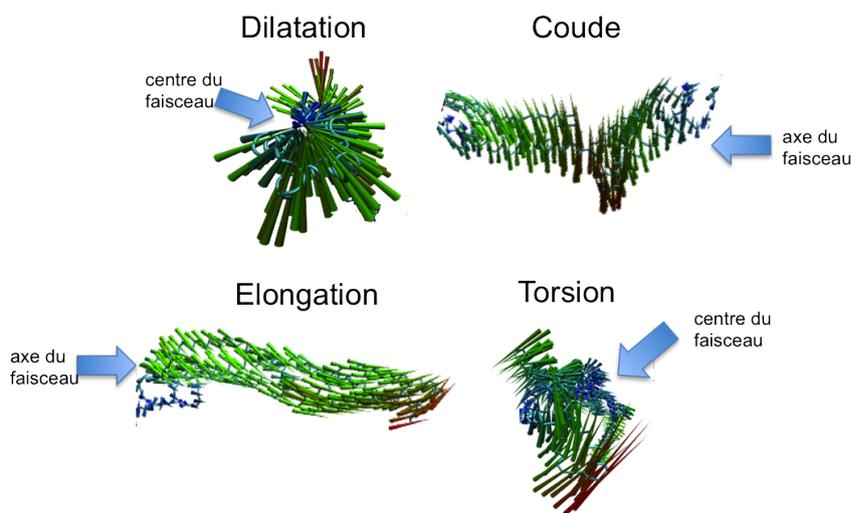


FIGURE 2.14 – Exemple des quatre sous-mouvements observés. Les mouvements sont représentés par des vecteurs indiquant l'intensité du mouvement par un dégradé croissant de bleu - vert - rouge.

L'ensemble de ces mouvements est décrit dans le tableau 2.6. D'un point de vue général, les principaux mouvements diffèrent d'une protéine à l'autre. Cependant, l'étude des sous-mouvements met en évidence des similitudes entre les protéines qui présentent toutes des coudes majoritairement situés en position centrale. Dans le cas de la protéine native R16-21 les points de flexion se situent au niveau des linkers entre R18 et R19 et entre R20 et R21 dans les deux composantes majoritaires, alors que dans les protéines tronquées, ils se situent au niveau de la charnière H3. Les délétions entraîneraient donc un décalage du point de flexibilité des protéines. De façon prévisible, un point de flexion apparaît au niveau de la répétition fracturée R17//18 de  $\Delta 45-47$  qui ne présente pas de structure tertiaire.

Ces coudes peuvent entraîner un mouvement de torsion des faisceaux d'hélices voisins comme c'est le cas pour le deuxième mouvement de  $\Delta 45-48$  et le premier mouvement de  $\Delta 45-49$ . On observe également un mouvement de dilatation de faisceaux d'hélices dans les protéines tronquées notamment pour  $\Delta 45-48$ . Cette dilatation, observée uniquement dans les répétitions tronquées, semble être causée par la fluctuation importante de H3 et de la répétition R17//18 qui n'aident pas à maintenir la structure des faisceaux d'hélices voisins. Dans le cas de  $\Delta 45-48$ , ce phénomène semble accentué par la présence de la répétition hybride.

On note que le mouvement d'élongation n'est observé que dans deux mouvements de la protéine native R16-21 et n'est jamais observé dans les protéines tronquées. Ces mouvements font suite à la formation de coudes au niveau du linker entre R18 et R19 dans le mouvement 1 et au niveau du linker entre R17 et R18 dans le mouvement 4 et ne résultent donc pas d'une élongation des répétitions elles-mêmes mais font suite à une invagination de la protéine.

La protéine  $\Delta 45-51$  est un cas particulier. En effet, comme présenté dans le graphe de fluctuation 2.13, elle fluctue peu. Elle adopte principalement un mouvement de torsion. Ceci peut s'expliquer par l'absence de la charnière H3 qui est flexible dans les protéines tronquées.

Protéine	Composante (mouvement)	N-term	Centre	C-term
R16-21	1 (47 %)	Torsion	Coude (Linker R18-19)	Elongation
	2 (17 %)	Torsion	Coude (Linker R18-19)	Coude (Linker R20-21)
	3 (13 %)	Torsion	Torsion	Torsion
	4 (7 %)	Coude (Linker R17-18)	Elongation	-
$\Delta$ 45-47	1 (45 %)	Torsion	Coude (H3)	-
	2 (18 %)	Coude & Dilatation (R17//18)	Coude (H3)	Torsion
	3 (12 %)	Coude (R17//18)	Torsion	Torsion
	4 (9 %)	Torsion	-	Torsion
$\Delta$ 45-48	1 (52 %)	Dilatation	Coude (H3)	Torsion
	2 (26 %)	Torsion	Coude (H3)	Torsion
	3 (8 %)	Dilatation	Coude (H3)	Dilatation
	4 (4 %)	Dilatation	Coude (H3)	-
$\Delta$ 45-49	1 (42 %)	Torsion	Coude (H3)	Torsion
	2 (26 %)	Torsion	Coude (H3)	-
	3 (11 %)	Torsion	-	Dilatation
	4 (4 %)	Torsion	Coude & Torsion (H3)	Dilatation
$\Delta$ 45-51	1 (51 %)	Torsion		Torsion
	2 (26 %)	-		Torsion
	3 (8 %)	-		Coude (Linker R20-21)
	4 (4 %)	-		Dilatation

Tableau 2.6 – Description des quatre mouvements principaux des protéines au cours des simulations de DM. Chaque composante est indiquée avec son pourcentage d’occurrence au cours de la trajectoire de la protéine. Les mouvements principaux sont découpés en trois sous-mouvements localisés dans les parties N-terminales (R16-17), centrales (R18-19-H3) et C-terminales (R20-21) des protéines, excepté pour la plus courte protéine,  $\Delta$ 45-51, dont les mouvements sont découpés en deux sous-mouvements. Dans le cas des coudes, le point de flexion du coude est indiqué entre parenthèses. Lorsqu’aucun sous-mouvement significatif n’est présent, la case est vide.

## La conservation de la structure secondaire au cours des DM

L'étude de la conservation de la structure secondaire est un moyen d'étudier la stabilité des protéines. Dans notre cas, les fragments de dystrophine sont majoritairement composés de faisceaux de trois hélices théoriquement stables, et nous nous intéressons à la conservation des hélices.

Globalement, la structuration en hélices des répétitions hybrides est stable au cours de nos simulations.

Les modèles initiaux sont majoritairement composés d'hélices  $\alpha$ , mais on observe que pendant les trajectoires de DM la structuration en hélices peut fluctuer notamment au niveau des extrémités des hélices (figure 2.15 - tableau 2.7). Les acides aminés structurés en hélices pendant au moins 80% du temps des trajectoires de DM sont considérés comme faisant partie d'une hélice.

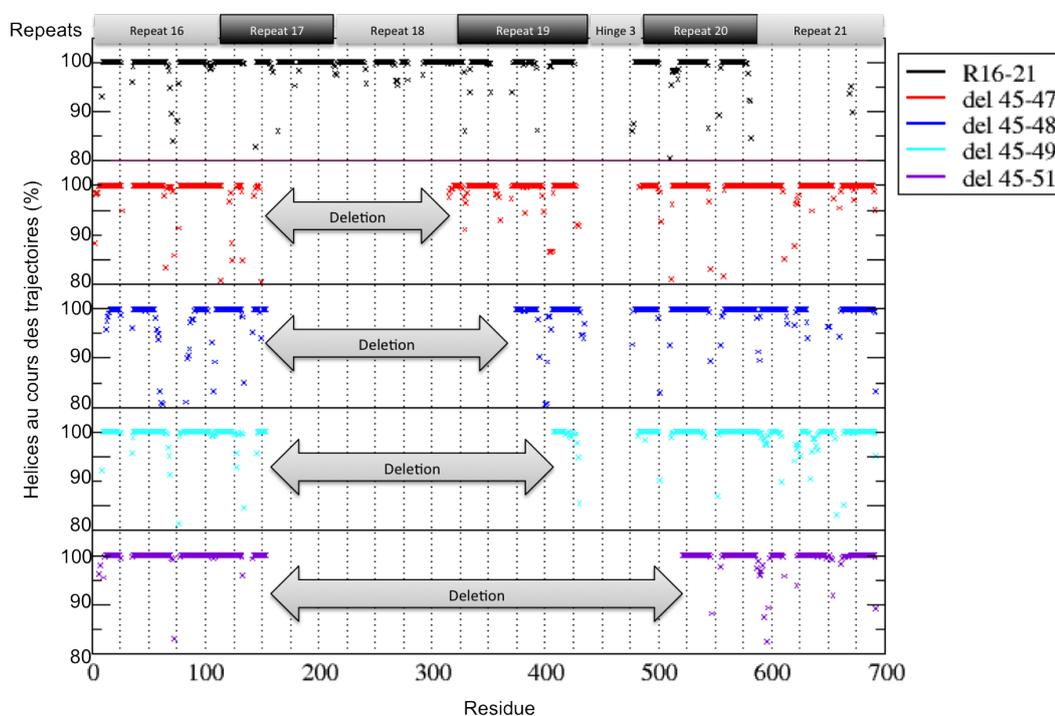


FIGURE 2.15 – Conservation de la structuration en hélice de chaque acide aminé au cours des trajectoires de DM. Chaque acide aminé est représenté par une croix. Les traits sont formés par la superposition de croix consécutives. Les acides aminés structurés en hélices pendant au moins 80% du temps des trajectoires de DM sont considérés comme faisant partie d'une hélice. Par clarté, seuls les acides aminés structurés en hélices pendant au moins 80% du temps de DM sont représentés. Les intervalles entre les hélices sont des boucles.

En accord avec la figure 2.13, nous observons une déstructuration totale des hélices de la répétition 21 dans la protéine R16-21 et partielle dans les protéines  $\Delta 45-48$  et  $\Delta 45-49$ . Une analyse plus précise de sa structure secondaire nous indique qu'elle est alors essentiellement constituée de tours.

Dans R16-21, nous observons des "points de faiblesse" au niveau des linkers et au milieu des hélices B de R17 et R18. La perte d'hélicité peut être expliquée par la présence d'acides

aminés cycliques possédant des chaînes latérales gênant le repliement en faisceaux d'hélices. Ces "points de faiblesse" sont également observés au niveau des linkers entre R16 et R17 pour  $\Delta 45-48$  et entre R20 et R21 pour  $\Delta 45-48$ ,  $\Delta 45-49$  et  $\Delta 45-51$ . De plus, on observe, en accord avec l'analyse de la fluctuation que la mini-charnière entre R18 et R19 n'est plus structurée en hélice, ce qui explique sa flexibilité.

Nous observons que la délimitation des hélices peut varier, généralement d'un ou deux acides aminés, soit moins d'un tour d'hélice (tableau 2.7). Cependant certaines régions hélicoïdales sont moins bien définies, comme par exemple les hélices B et C de R16 dans  $\Delta 45-48$  qui diminuent. Cette déformation est certainement dû à l'absence de la répétition en position N-terminal R15. On note également ce type d'évènement au niveau des hélices R19C et R20A entourant H3.

Au niveau des jonctions des répétitions, seule la répétition fracturée R17//18 de  $\Delta 45-47$  présente une perte importante d'hélicité.

Les répétitions hybrides se structurent en faisceaux de trois hélices similaires aux faisceaux d'hélices natifs bien que nous observions des points de faiblesse au niveau de la répétition hybride de  $\Delta 45-48$ .

	R16			R17			R18		
	A	B	C	A	B	C	A	B	C
R16-21	9.....25	35.....71	75..	..134	144....159 164....179	182..	..242	250....270 273....279	292....329
Δ45-47	2.....26	36.....72	76..	..114 121....124 128....133	143-149	DEL	DEL	DEL	315..
Δ45-48	12...25	35...62	83...102	106...134	142..	DEL	DEL	DEL	DEL
Δ45-49	9....25	35....69	76..	..134	145..	DEL	DEL	DEL	DEL
Δ45-51	6....7 10.....25	35....72	78..	..133	142..	DEL	DEL	DEL	DEL

	R19			R20			R21		
	A	B	C	A	B	C	A	B	C
R16-21	333....352	371....393	404....426	477....500	510....544	554....582	T	T	T
Δ45-47	..360	370....399	404....429	483....502	511....546	557..	..611	620....655	660....691
Δ45-48	DEL	..395	400....435	477....501	510....546	555..	..614	620....632 650....651	660....692
Δ45-49	DEL	DEL	..430	482....501	511....541	552..	..609	620....657	663....692
Δ45-51	DEL	DEL	DEL	DEL	..547	555..	..593	622....654	661....680
							596....611		

Tableau 2.7 – Positions des hélices des répétitions. Les hélices sont définies comme telles si elles adoptent une conformation hélicoïdale pendant au moins 80% du temps de la DM. T : Tour; DEL : hélice tronquée.

### Le regroupement des DM en structures représentatives

Le regroupement des DM en structures représentatives permet de réduire les 22 000 structures collectées par protéines à trois structures représentatives. Elles permettent de réaliser une analyse graphique de ces structures et de réaliser certaines analyses qui ne peuvent pas être mises en place sur l'ensemble de la trajectoire.

Les structures représentatives reflètent les conclusions des analyses réalisées précédemment. En effet, les structures représentatives de R16-21 mettent en évidence les points de flexibilité au niveau de la mini-charnière entre R18 et R19 et au niveau de R19. Bien que la répétition 21 ne soit pas composée d'hélices, elle conserve son agencement en faisceau (figure 2.16).

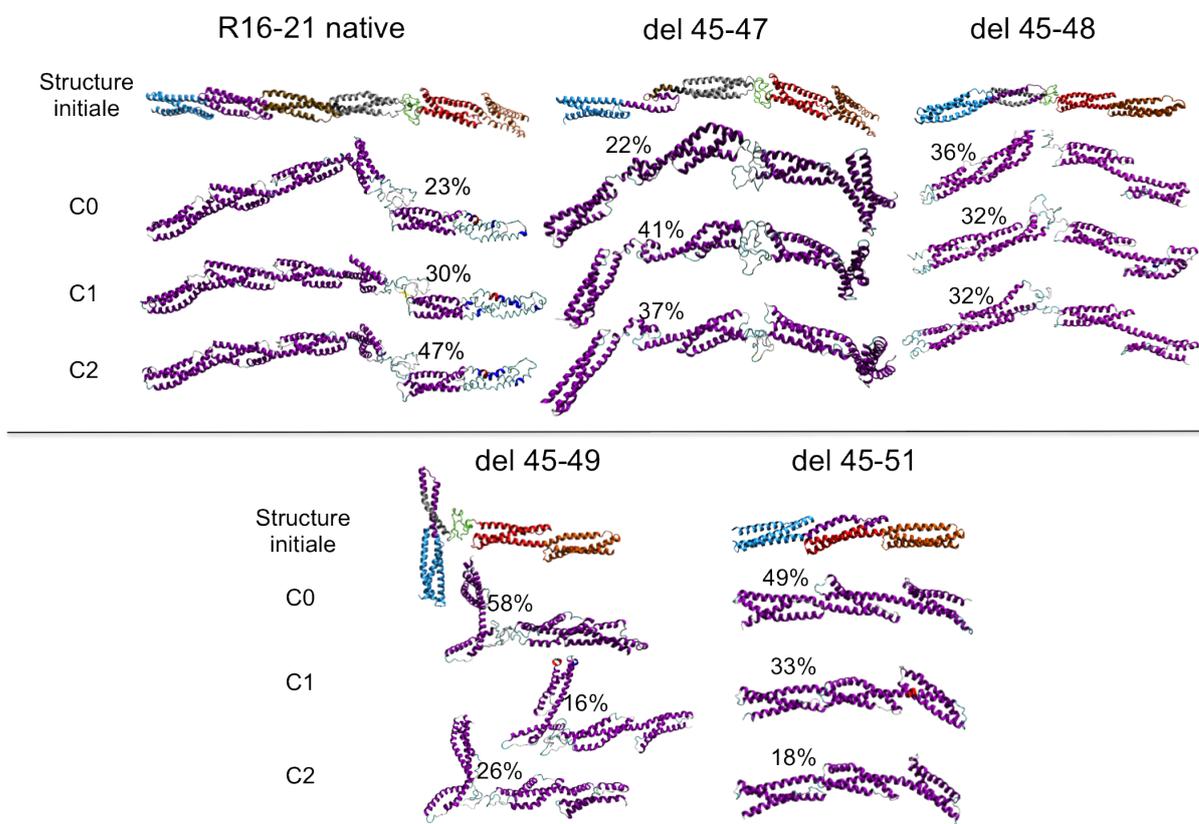


FIGURE 2.16 – Structures représentatives de chaque protéine au cours des 22 dernières nanosecondes de DM. Pour chaque protéine, sont représentées les structures initiales colorées en fonction des répétitions puis les 3 structures représentatives C0 à C2. Les structures représentatives sont colorées en fonction de la structure secondaire, les hélices alpha sont en violet, les boucles en blanc, les tours en vert, les hélices 3.10 en bleu et les hélices Pi en rouge. Pour chaque structure, le pourcentage qu'elle représente dans l'ensemble des structures de la trajectoire est indiqué.

	C0	C1	C2
C0	0	13,6	9,7
C1	-	0	8,4
C2	-	-	0

(a) R16-21

	C0	C1	C2
C0	0	14,4	9,1
C1	-	0	10
C2	-	-	0

(b)  $\Delta$ 45-47

	C0	C1	C2
C0	0	12,1	11,8
C1	-	0	13,8
C2	-	-	0

(c)  $\Delta$ 45-48

	C0	C1	C2
C0	0	9,4	9,9
C1	-	0	8,1
C2	-	-	0

(d)  $\Delta$ 45-49

	C0	C1	C2
C0	0	8,8	6,1
C1	-	0	5,4
C2	-	-	0

(e)  $\Delta$ 45-51

Tableau 2.8 – *Ecart quadratique moyen des  $C\alpha$  (en Å) entre les structures représentatives de chaque groupe.*

En accord avec la figure 2.13, la protéine  $\Delta$ 45-47 semble posséder de nombreux points de flexibilité, notamment au niveau de la répétition fracturée R17//18.

La répétition hybride de  $\Delta$ 45-48 et la charnière H3 apparaissent comme les éléments les plus flexibles de la protéine.

Les deux hélices de la répétition fracturée de  $\Delta$ 45-49 pivotent l'une par rapport à l'autre permettant à la répétition de s'aligner dans le prolongement de R20 et R21. La protéine passe ainsi d'une forme en T à une forme en L. Les boucles de la charnière s'adaptent alors à cette nouvelle conformation.

La répétition hybride de  $\Delta$ 45-51 conserve sa structure en faisceau de trois hélices. Les structures représentatives sont proches comme le confirme le calcul des écarts quadratiques moyens (figure 2.8). Cependant, on observe un point de flexibilité au milieu de l'hélice B de la répétition hybride correspondant à la présence d'une proline.

### 2.3.2.2 Les propriétés de surface

Comme nous l'avons vu dans l'étude des répétitions natives (section 2.2), les propriétés de surface hydrophobes et électrostatiques peuvent être corrélées avec les propriétés de liaison des répétitions de la dystrophine. Nous avons donc entrepris des études similaires pour les dystrophines tronquées.

#### Les propriétés électrostatiques

Le potentiel de surface électrostatique de chaque protéine est représenté dans la figure 2.17. La figure 2.18 présente le potentiel électrostatique moyen pour chaque acide aminé du jeu de protéines étudiées au cours des trajectoires de DM. Il s'agit d'une moyenne calculée sur les structures représentatives décrites dans la figure 2.16 puis lissée sur une fenêtre de neuf résidus. Cette approche est donc plus précise que la description qualitative des propriétés de surface que nous avons effectuée sur les tandems de répétitions de la protéine native.

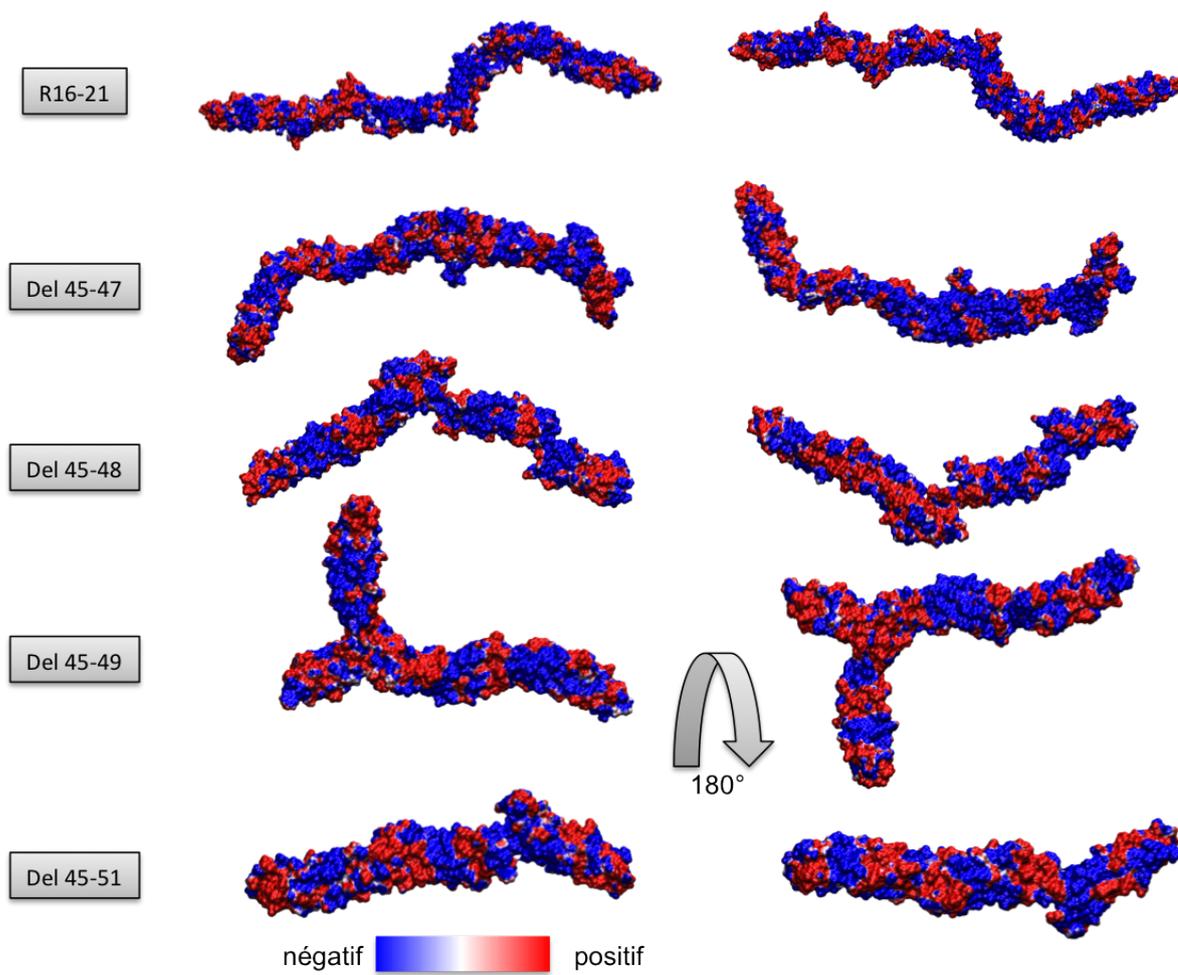


FIGURE 2.17 – Potentiels électrostatiques projetés sur les surfaces des structures représentatives des groupes les plus importants de chaque protéine. Un gradient de couleur bleu - blanc - rouge indique le potentiel électrostatique où les régions négatives sont colorées en bleu et les régions positives en rouge.

En accord avec l'étude que nous avons réalisée précédemment (section 2.2), nous observons que la protéine R16-21 est composée de deux répétitions (R16 et R17) présentant une alternance de patches positifs et négatifs puis de deux répétitions chargées négativement (R18 et R19), d'une charnière globalement neutre et de deux répétitions 20 et 21 possèdent des patches fortement électronégatifs (figures 2.17 et 2.18).

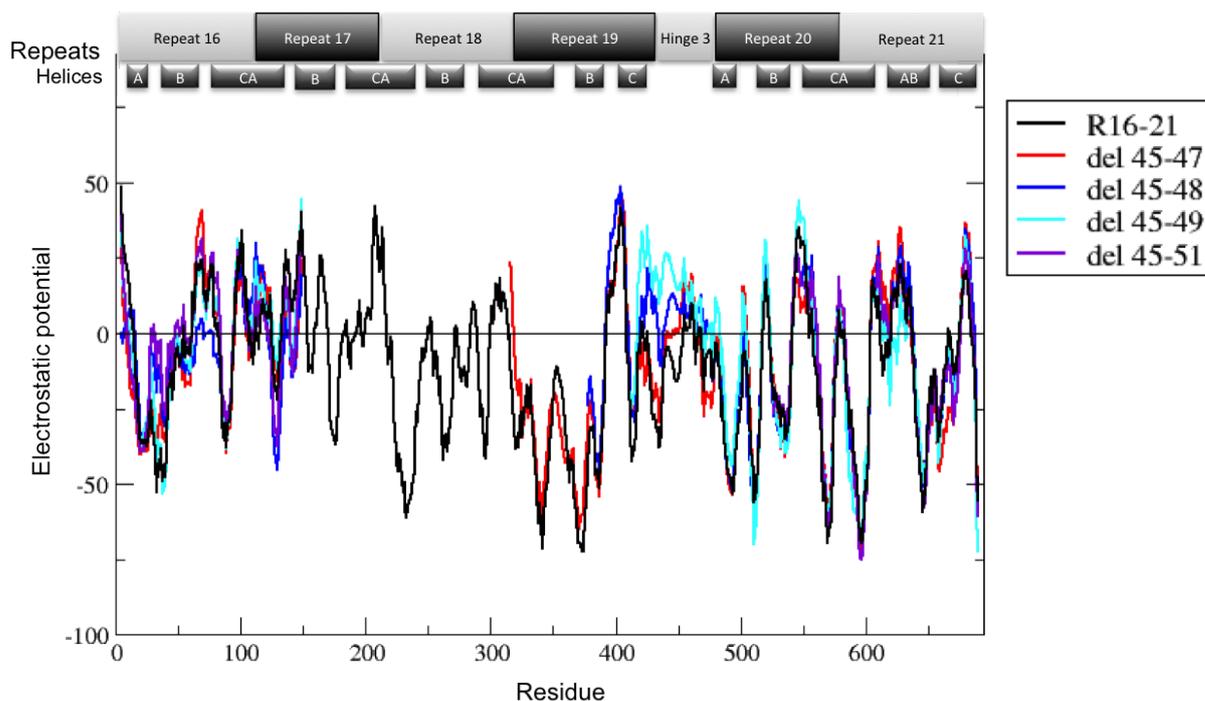


FIGURE 2.18 – Potentiels électrostatiques des protéines en présence de 50 mM de NaCl. Les potentiels électrostatiques (en kTe) ont été calculés par APBS sur les structures représentatives de la trajectoire de DM et décrits dans la figure 2.16 puis moyennés par résidu. Les courbes correspondent à ces potentiels moyens et sont lissés sur une fenêtre de 9 résidus. Les courbes de potentiel des dystrophines tronquées sont découpées en fonction de la délétion et superposées sur la courbe de la dystrophine native R16-21.

On observe que la charnière H3 globalement neutre dans R16-21 se charge positivement notamment dans  $\Delta 45-49$ . Cette charnière essentiellement composée de boucles désordonnées peut facilement se réarranger et exposer au solvant plus de résidus chargés positivement. La structure non filamenteuse de  $\Delta 45-49$  semble imposer un changement conformationnel de H3 plus important que les autres protéines tronquées et expose plus de résidus chargés positivement en surface.

Dans un deuxième temps, nous nous sommes tout particulièrement intéressés aux propriétés électrostatiques de la jonction entourant les délétions. Ainsi, nous avons superposé les potentiels électrostatiques de la répétition 17 de R16-21 avec les répétitions hybrides et fracturées des quatre protéines tronquées (figure 2.19). Dans la partie N-terminale de R17, commune à toutes les protéines, nous remarquons le même profil mais avec des différences d'intensité.

Après la délétion, nous observons des profils électrostatiques différents entre R17 native et les répétitions reconstituées. Ceci est particulièrement vrai pour les répétitions fracturées R17//18 de  $\Delta 45-47$  et R17//19 de  $\Delta 45-49$  qui présentent, respectivement, des potentiels entièrement négatifs et entièrement positifs. Les répétitions hybrides R17//18 de  $\Delta 45-48$  et R17//20 de  $\Delta 45-51$  présentent des profils hétérogènes avec des patches négatifs et positifs comme R17 de R16-21 mais avec une plus forte intensité et un profil inversé. En effet, les

profils de R18 ( $\Delta 45-48$ ) et de R20 ( $\Delta 45-51$ ) présentent une alternance de patches négatif - positif - négatif alors que la région correspondante dans R17 présente un alternance de patches positif - négatif - positif.

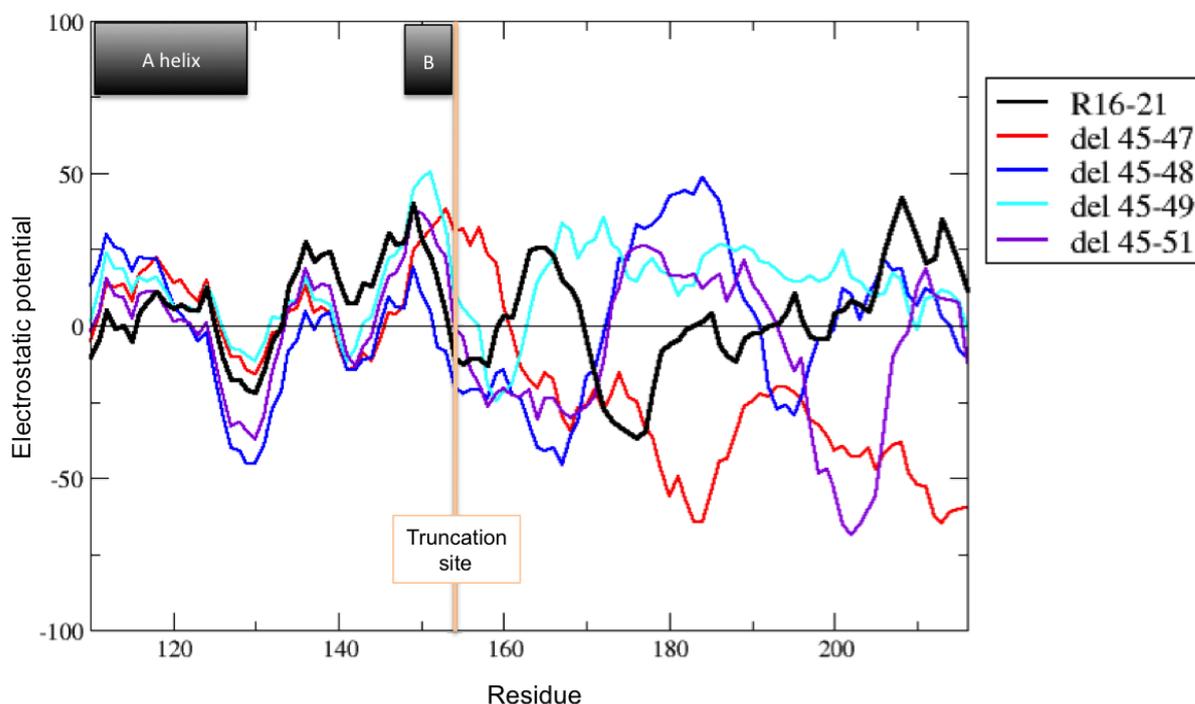


FIGURE 2.19 – Potentiels électrostatiques (en  $kTe$ ) de la répétition 17 de R16-21 et des répétitions formées suite aux délétions à partir de la partie N-terminale de R17. Les potentiels électrostatiques ont été calculés par APBS sur les structures représentatives de la trajectoire de DM et décrits dans la figure 2.16 puis moyennés par résidu. Les courbes correspondent à ces potentiels moyens et sont lissées sur une fenêtre de 9 résidus.

### Les propriétés hydrophobes

Les potentiels hydrophobes ont été calculés par le serveur PLATINUM sur les 3 structures représentatives de chaque protéine. En accord avec l'étude des répétitions natives (section 2.2), les protéines étudiées présentent une surface largement hydrophile avec des patches et sillons hydrophobes (figure 2.20). Les cinq protéines ont un pourcentage de surface hydrophobe (i.e. le pourcentage de la surface moléculaire ayant des propriétés hydrophobes) similaire : 21 % pour R16-21 et  $\Delta 45-51$  et 22 % pour les trois autres.

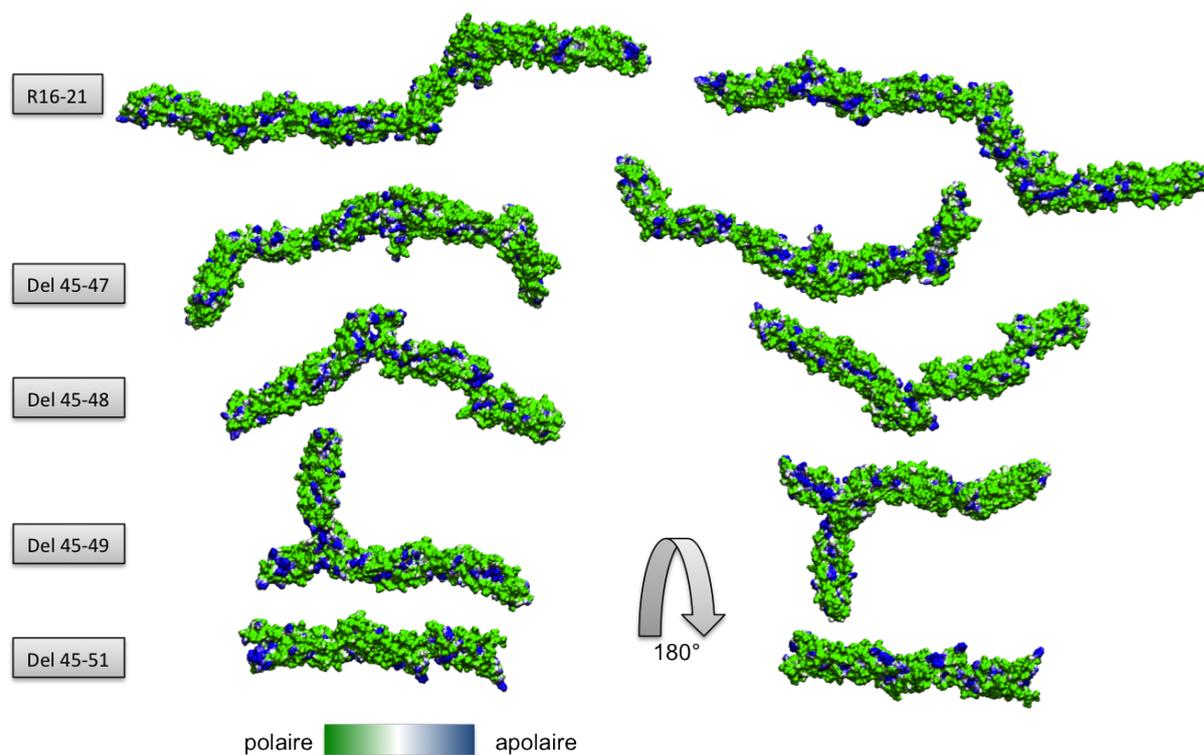


FIGURE 2.20 – Potentiels hydrophobes projetés sur les surfaces des structures représentatives des groupes les plus importants de chaque protéine. Un gradient de couleur vert - blanc - bleu indique le potentiel hydrophobe avec les régions polaires colorées en vert et les régions apolaires en bleu.

Comme pour le potentiel électrostatique, le potentiel hydrophobe a ensuite été moyenné pour chaque résidu. La courbe résultant de ce calcul a été lissée sur 7 acides aminés. De façon attendue, nous ne remarquons pas de différences sur l'hydrophobicité globale entre les cinq protéines (figure 2.21).

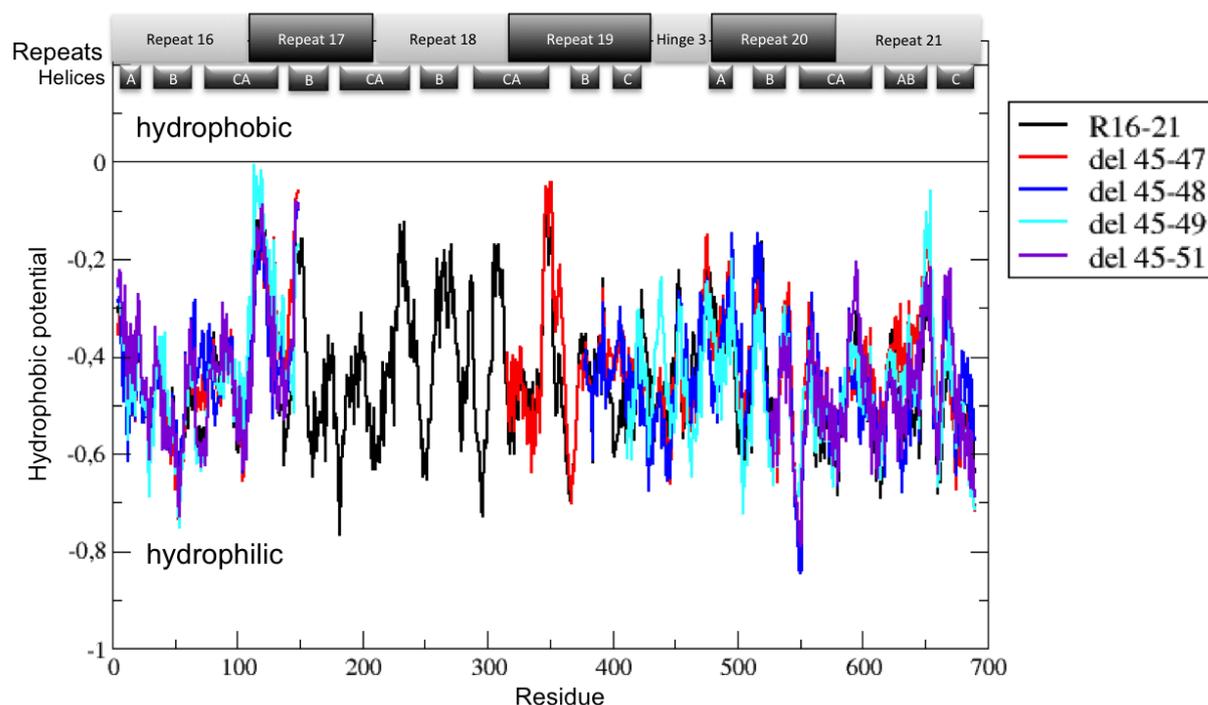


FIGURE 2.21 – Potentiels hydrophobes des protéines. Les potentiels hydrophobes ont été calculés par PLATINUM sur les structures représentatives de la trajectoire de DM et décrits dans la figure 2.16 puis moyennés par résidu. Les courbes correspondent à ces potentiels moyens et sont lissées sur une fenêtre de 7 résidus correspondant à une heptade. Dans une fenêtre, les résidus hydrophobes des positions a et d des heptades sont répartis entre plusieurs résidus hydrophiles diminuant artificiellement le potentiel hydrophobe. Les courbes de potentiel des dystrophines tronquées sont découpées en fonction de la délétion et superposées sur la courbe de la dystrophine native R16-21.

Nous nous sommes alors plus particulièrement intéressés à la région entourant le point de jonction des délétions. Comme pour l'étude du potentiel électrostatique, nous avons ainsi superposé le potentiel hydrophobe de R17 de R16-21 avec les répétitions hybrides et fracturées des quatre protéines tronquées (figure 2.22).

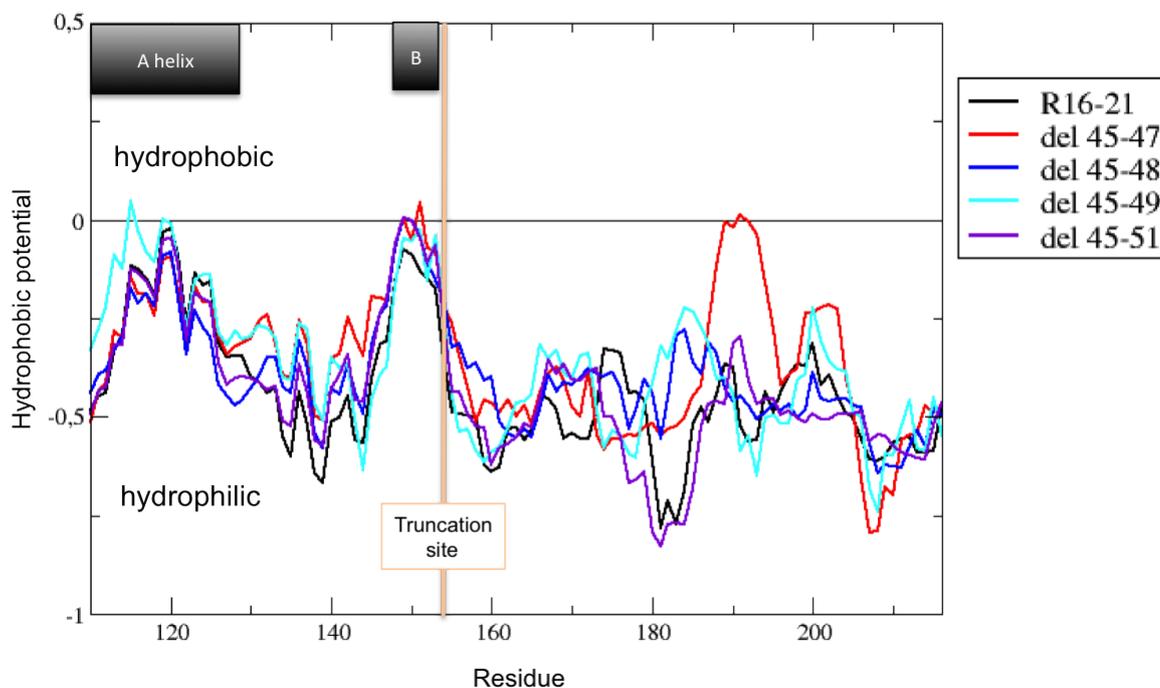


FIGURE 2.22 – Potentiels hydrophobes de la répétition 17 de R16-21 et des répétitions tronquées formées à partir de la partie N-terminale de R17. Les potentiels hydrophobes ont été calculés par *Platinum* sur les structures représentatives de la trajectoire de DM et décrits dans la figure 2.16 puis moyennés par résidu. Les courbes correspondent à ces potentiels moyens et sont lissées sur une fenêtre de 7 résidus.

Nous remarquons que les répétitions hybrides R17//19 de  $\Delta 45-48$  et R17//20 de  $\Delta 45-51$  ont un potentiel hydrophobe proche de la répétition 17 native. Ces répétitions hybrides respectent la succession des motifs en heptade permettant aux résidus hydrophobes d'être enfouis au cœur des faisceaux d'hélices comme dans les répétitions natives. On remarque cependant que  $\Delta 45-48$  possède un patch plus hydrophobe que  $\Delta 45-51$ , probablement à cause de la grande flexibilité de  $\Delta 45-48$  qui présente plus souvent des résidus hydrophobes en surface.

A l'inverse, les répétitions fracturées R17//18 de  $\Delta 45-47$  et R17//19 de  $\Delta 45-49$  ont, comme attendu, un potentiel hydrophobe plus important, particulièrement la répétition fracturée de  $\Delta 45-47$  qui est déstructurée.

Nous ne notons pas de différence majeure d'hydrophobicité entre la protéine native R16-21 et les protéines possédant une répétition hybride  $\Delta 45-48$  et  $\Delta 45-51$ . Ces protéines sont toutes amphiphiles.

### 2.3.2.3 L'analyse des contacts moléculaires

Les linkers entre deux répétitions sont situés au milieu des hélices communes à deux répétitions successives. Comme nous l'avons précédemment vu dans l'étude de la fluctuation de R16-21, ils peuvent être une région hautement flexible. La fluctuation des linkers est notamment déterminée par les interactions des acides aminés qui les compose avec les acides aminés des hélices et boucles proches. Ces interactions peuvent être des interactions hydrophobes caractéristiques des faisceaux d'hélices, des liaisons hydrogènes et des interactions de van der Waals. Les cinq protéines étudiées ont deux linkers en communs : celui qui lie R16 et R17 et celui qui lie R20 et R21. Cependant, selon les délétions, ces linkers peuvent se situer en face de résidus différents comparés à la protéine native et leur flexibilité peut être modifiée.

#### Analyse des interactions du linker R16-17 avec les répétitions environnantes

Afin d'analyser en détail les contacts moléculaires nous avons, dans un premier temps, cartographié les acides aminés en contact avec le linker R16-17 (figure 2.23). D'un côté, nous remarquons que les acides aminés de R16 en contact avec le linker sont tous situés dans la région entourant la boucle AB de R16. De l'autre côté, nous remarquons que les acides aminés en contact avec le linker sont situés dans la région entourant la boucle BC de R17, R17//19 et R17//20 des protéines R16-21,  $\Delta 45-48$  et  $\Delta 45-51$  respectivement. Le linker présente peu de contacts avec la répétition fracturée de  $\Delta 45-49$  mais est en contact avec la charnière H3. Sans surprise, les résidus de la petite répétition fracturée de  $\Delta 45-47$  présente peu de zones de contact avec le linker et différent de ceux de R16-21.

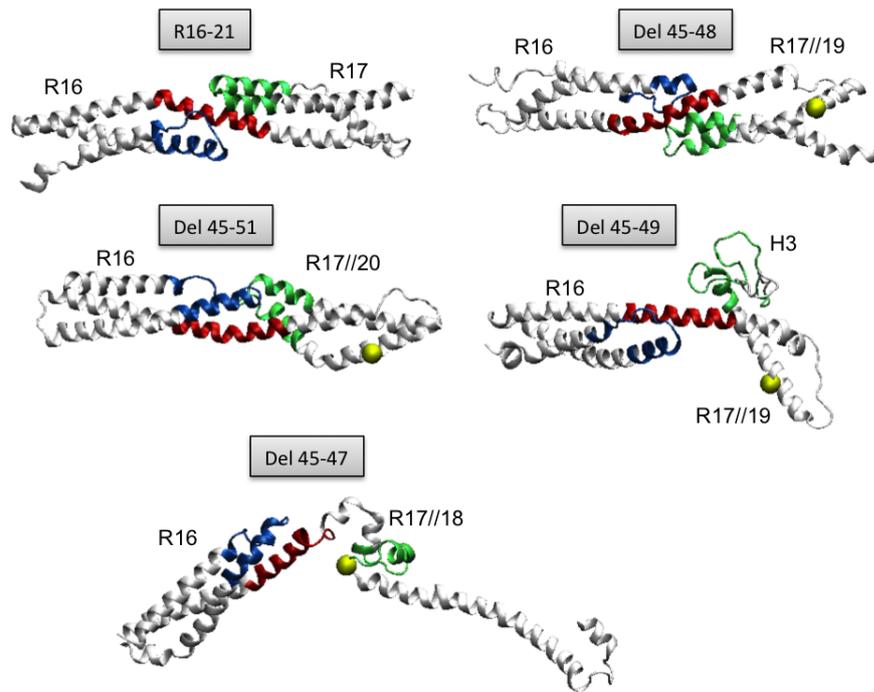


FIGURE 2.23 – Régions en contact avec le linker entre R16 et R17. Seuls les tandems de répétitions entourant le linker sont représentés. Le linker est représenté en rouge, la région autour de la boucle AB de R16 en contact est en bleu et la région de R17, R17//18, R17//19, H3 ou R17//20 en contact est en vert. La sphère jaune indique la jonction de la délétion.

Nous observons également que les boucles en contact avec le linker ont tendance à se placer du même côté du linker dans R16-21,  $\Delta 45-48$  et  $\Delta 45-51$ . Ces boucles peuvent alors créer des interactions entre elles et rigidifier encore un peu plus le linker. Ce type de comportement semble donc conservé dans les protéines présentant des répétitions hybrides.

Une analyse plus détaillée de ce linker R16-17, résidu par résidu, nous indique que les contacts les plus nombreux et les plus fréquents sont sur R16 comparé à R17 et ses équivalents en C-terminal (figure 2.24). Nous dénombrons 13 contacts effectifs pendant plus de 80% du temps pour R16-21 et  $\Delta 45-49$ , 12 pour  $\Delta 45-47$  et  $\Delta 45-51$  et seulement 8 pour  $\Delta 45-48$ . A l'exception de  $\Delta 45-48$ , tous les résidus hydrophobes des heptades de l'hélice B sont fortement en contact avec le linker suggérant des interactions hydrophobes constituant un faisceau à deux hélices.

Les contacts de R17 avec le linker sont moins fréquents et moins nombreux que ceux de R16. On dénombre 8 contacts effectifs pendant plus de 80% du temps pour R16-21, principalement au niveau des hélices. La comparaison des contacts du linker R16-17 avec la répétition R17 native et les répétitions hybrides de  $\Delta 45-49$  et  $\Delta 45-51$  est délicate. En effet, elles conservent des interactions entre les résidus hydrophobes de l'hélice B et le linker et présentent plus de contacts au niveau de la boucle BC. D'une manière générale, on peut dire qu'elles ont des profils de contacts différents de ceux de la répétition 17 native.

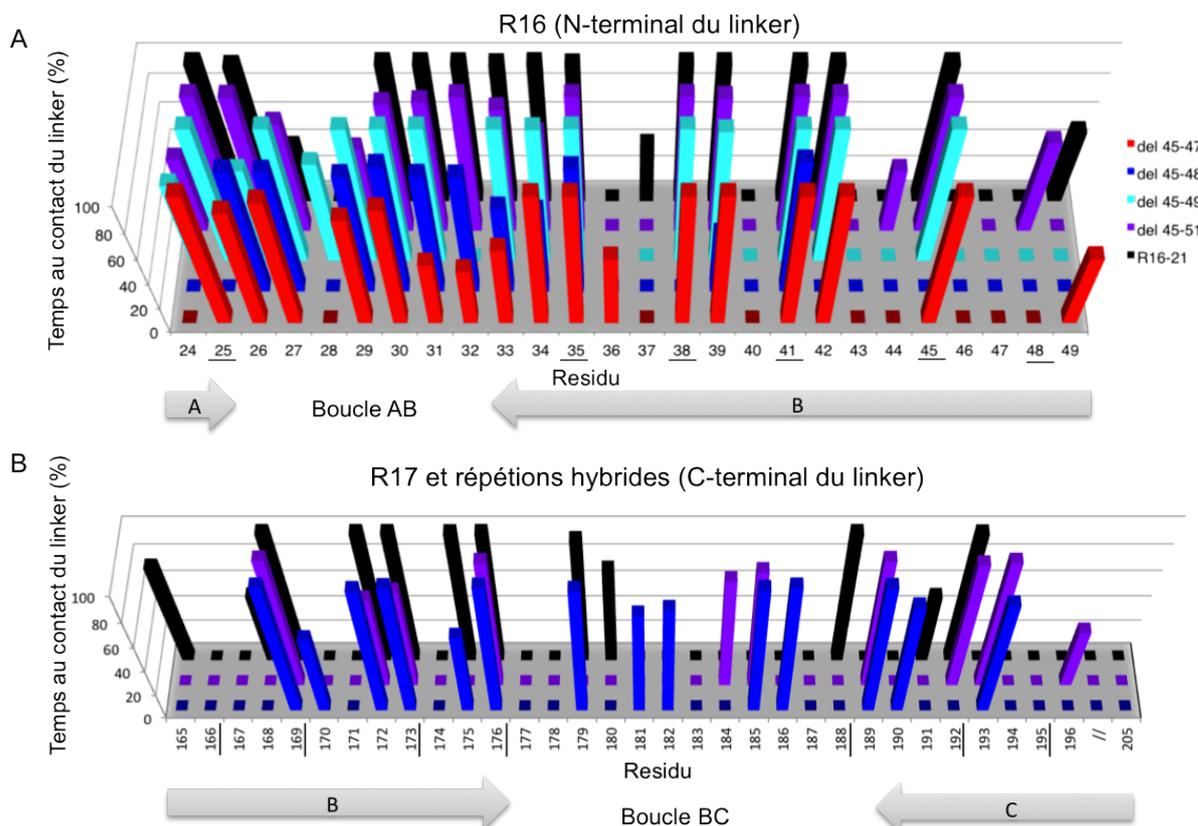


FIGURE 2.24 – Temps de contact des acides aminés de R16, R17, R17//19 et R17//20 avec le linker entre R16 et R17 (défini entre les positions 100 et 119). Deux résidus sont considérés en contact s'ils sont espacés de moins de 5,5 Å. Seuls les contacts effectifs pendant plus de 40 % de la DM sont représentés ici. Les numéros de résidus soulignés indiquent les acides aminés hydrophobes en position a et d de l'heptade. Les flèches grises indiquent les hélices. A. Contacts entre le linker et les résidus de R16. B. Contacts entre le linker et les acides aminés de R17 de R16-21, R17//19 de  $\Delta 45-48$  et R17//20 de  $\Delta 45-51$ .

### Analyse des interactions du linker R20-21 avec les répétitions environnantes

Le linker 20-21 est entouré par des répétitions entières à l'exception de R17//20 de  $\Delta 45-51$  qui est une répétition hybride. Sans surprise, nous observons que les régions en contact avec le linker sont très similaires entre elles : la région entourant la boucle AB de R20 (ou R17//20) d'un côté et la région entourant la boucle BC de R21 de l'autre côté (figure 2.25).

Contrairement au linker R16-17, nous observons des placements différents des deux boucles par rapport au linker. En effet, seules les boucles de R16-21 et  $\Delta 45-47$  semblent se placer du même côté du linker. Bien que le linker R20-21 soit entouré de faisceaux d'hélices dans les cinq protéines, les boucles au contact du linker n'adoptent pas le même positionnement. On remarque que la répétition 21 de R16-21 étant partiellement destructurée, son comportement pourrait être artefactuel.

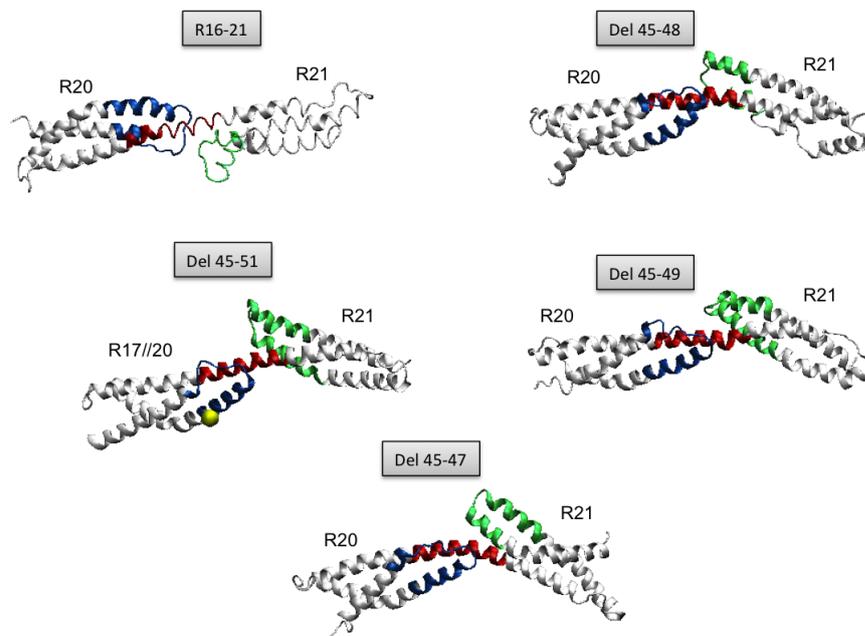


FIGURE 2.25 – Régions en contact avec le linker entre R20 et R21. Seuls les tandems de répétitions entourant le linker sont représentés. Le linker est représenté en rouge, la région autour de la boucle AB de R20 (ou R17//20) en contact est en bleu et la région autour de la boucle BC de R21 en contact est en vert. La sphère jaune indique la jonction de la délétion.

Les régions en contact avec le linker 20-21 sont globalement similaires entre les cinq protéines. Cependant, les profils de contact précis diffèrent (figure 2.26). Les contacts de R20 de la protéine R16-21 avec le linker sont plus étendus et plus dispersés que pour les protéines tronquées. Globalement, les boucles AB des protéines tronquées présentent plus de contacts avec le linker que celle de R16-21. Les interactions hydrophobes entre le linker et les résidus a et d des heptades de l'hélice B sont globalement maintenues.

Comme pour le linker R16-17, la répétition en C-terminal du linker présente moins de contacts avec le linker que la répétition en N-terminal. On dénombre moins de contacts pour la répétition native, ce qui peut s'expliquer par la destructuration de la répétition. Les contacts de l'hélice B et de la boucle BC sont dispersés et aucun contact impliquant l'hélice B ou la boucle BC n'est commun aux cinq protéines. Par contre, on remarque que trois résidus de l'hélice B sont fréquemment impliqués dans les contacts au linker 20-21 dont deux résidus hydrophobes.

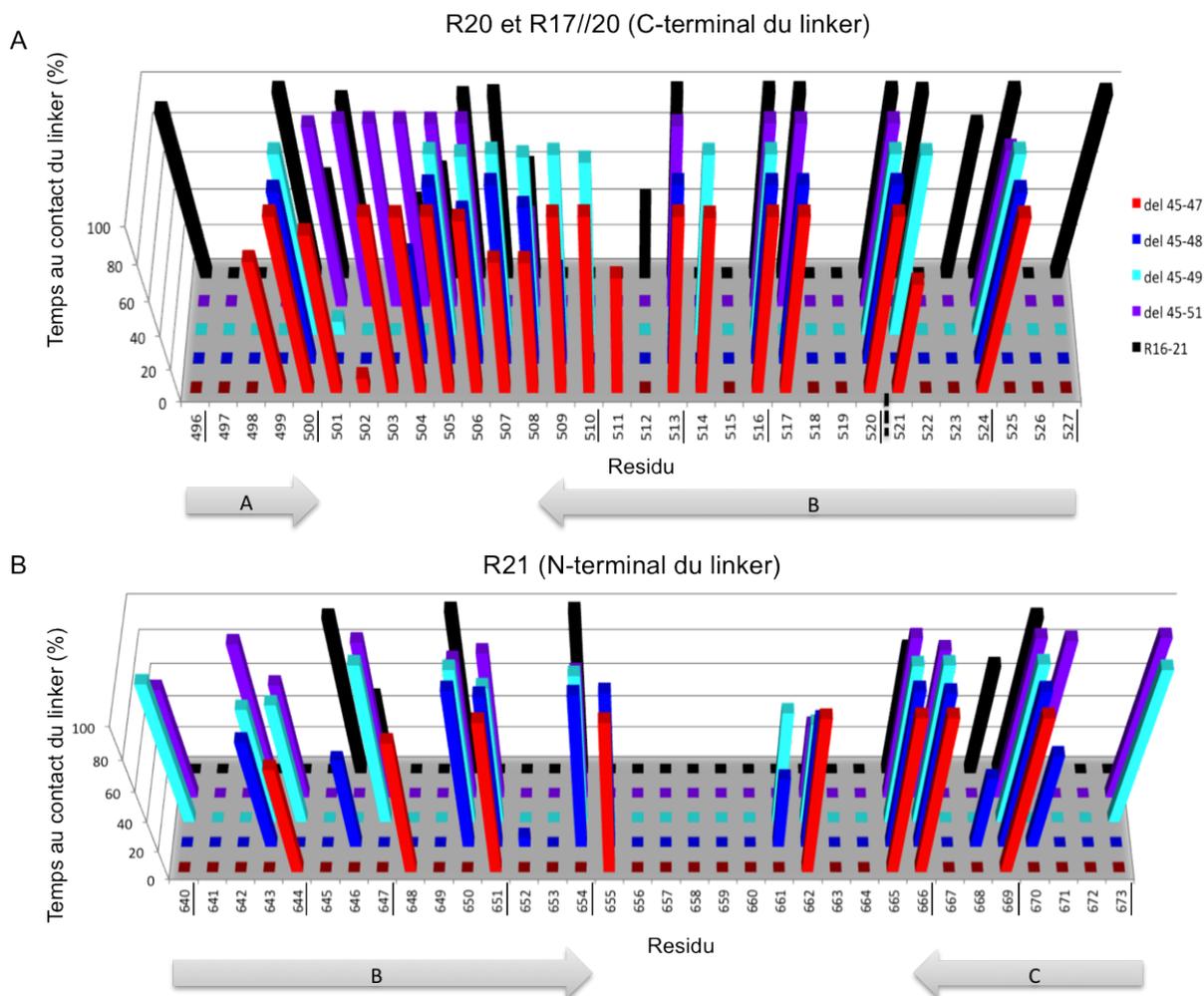


FIGURE 2.26 – Temps de contact des acides aminés avec le linker entre R20 et R21 (défini entre les positions 575 et 594 pour R16-21). Deux résidus sont considérés en contact s'ils sont espacés de moins de 5,5 Å. Seuls les contacts effectifs pendant plus de 40 % de la DM sont représentés ici. Les résidus sont numérotés en fonction de la dystrophine native R16-21 et les numéros de résidus soulignés indiquent les acides aminés hydrophobes en position a et d de l'heptade. Les protéines sont alignées sur R20-R21 (cf. annexe B). La ligne pointillée indique le site de jonction de la délétion de R17//20 de  $\Delta 45-51$ . Les flèches grises indiquent les hélices de la protéine native. A. Contacts entre le linker et les résidus de R20 ou R17//20 pour  $\Delta 45-51$ . B. Contacts entre le linker et les acides aminés de R21.

Les linkers situés à proximité d'une répétition fracturée ont des profils d'interaction influencés par les structures tertiaires des régions les entourant. Ainsi la répétition fracturée R17//18 de  $\Delta 45-47$  offre peu de surface de contacts avec le linker R16-17 alors que la répétition fracturée de  $\Delta 45-49$  permet au linker d'interagir avec la charnière H3, augmentant ainsi les contacts du linker avec les acides aminés l'entourant.

D'un autre côté, les linkers situés à proximité d'une répétition hybride conservent les interactions hydrophobes permettant de maintenir les faisceaux de trois hélices. Cependant les autres types d'interactions, dépendantes des chaînes latérales des acides aminés, peuvent varier en fonction de la séquence de la répétition hybride comme c'est le cas pour les contacts entre le linker R16-17 et les répétitions R17, R17//19 et R17//20.

La répartition des contacts du linker R20-21 avec la répétition R20 ou R17//20 est un cas particulier. A l'exception de la protéine  $\Delta 45-51$ , les protéines ont la même répétition R20 mais les contacts de la répétition R20 de R16-21 avec le linker R20-21 sont moins nombreux que les contacts entre le linker R20-21 et les répétitions des protéines tronquées entraînant une rigidification du linker. Or, l'augmentation et la diminution de la flexibilité des protéines peuvent être délétères. Il semble donc que les interactions des linkers avec les répétitions natives dans des protéines tronquées peuvent être modifiées par les délétions même si elles sont situées dans une région de la protéine à distance de la délétion. Cependant ce type de phénomène est à nuancer car il n'est observé que pour R20 et R21 mais pas pour R16. Or, la fin de la répétition 20 et la répétition 21 se déstructurent partiellement selon les protéines. Il est donc difficile de dire si ce phénomène est significatif ou non.

### 2.3.2.4 Le cas particulier de la charnière H3

Les charnières (H1 à H4) sont des régions composées de tours et de boucles sans structure tertiaire particulière et leurs rôles restent inconnus. Elles ne sont pas présentes dans les protéines homologues à la dystrophine dont les structures 3D sont résolues. La prédiction de la structure par homologie de H3 est donc difficile, faute de patrons disponibles dans la PDB. De plus, les boucles constituant la charnière H3 sont flexibles et il est donc difficile de converger vers une structure tertiaire définitive. Le modèle proposée par I-TASSER utilise des structures-patrons de différentes protéines : ATPase bactérienne (1CZ4), endonucléase I de champignon (2OVW), Dihydrolipoyl déshydrogénase (1zy8). L'utilisation de la DM permet donc d'affiner le modèle de H3 en s'éloignant des structures-patrons et en réalisant les simulations en présence des répétitions l'entourant.

En se référant à l'analyse de la conservation des hélices  $\alpha$  réalisée précédemment et tout particulièrement au tableau 2.7, nous remarquons que H3 s'agrandit dans trois protéines (+8 et +1 de part et d'autre pour R16-21, +4 et +6 pour  $\Delta 45-47$ , +4 et +5 pour  $\Delta 45-49$ ) par rapport à la définition de départ issue de l'alignement proposé par Winder [Winder 1995a]. De plus, la charnière de  $\Delta 45-48$  se décale d'un résidu vers R20. Cependant, pour faciliter la comparaison entre les quatre charnières H3 des quatre protéines tronquées, nous avons choisi de conserver la définition originale de H3.

Les écarts quadratiques moyens de H3 diffèrent entre les quatre protéines (figure 2.27A). Les charnières H3 entourées des répétitions natives de R16-21 et de  $\Delta 45-47$  atteignent plus rapidement un état d'équilibre et s'éloignent moins du modèle d'origine que H3 issue de  $\Delta 45-48$  et  $\Delta 45-49$ . La charnière de la protéine  $\Delta 45-48$  subit un grand réarrangement au début de la trajectoire, probablement à cause de la répétition hybride R17//19 qui la borde.

C'est également la charnière fluctuant le plus au cours de la phase de production (figure

2.27B). Il semblerait donc que la répétition hybride R17//19 de  $\Delta 45-48$  interagisse moins avec H3 que la répétition fracturée de  $\Delta 45-49$  et qu'elle semble donc plus mobile. Étrangement, les plus hauts points de fluctuation diffèrent entre les protéines. Il semblerait donc que l'environnement structural joue un plus grand rôle dans la fluctuation de H3 que sa séquence primaire.

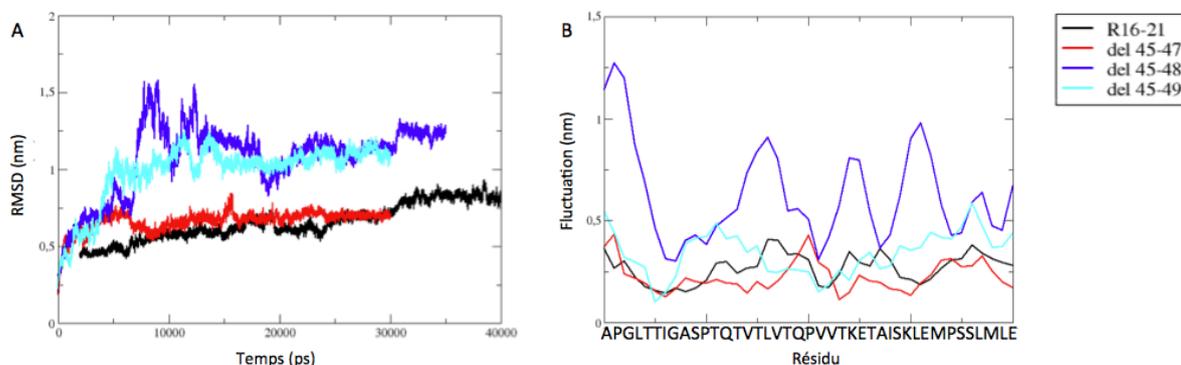


FIGURE 2.27 – A. *Ecart quadratique moyen entre les carbones  $\alpha$  des charnières H3 issues des trajectoires de DM des protéines R16-21,  $\Delta 45-47$ ,  $\Delta 45-48$  et  $\Delta 45-49$ .* B. *Fluctuation des carbones  $\alpha$  de chaque résidu de la charnière H3 au cours des DM.*

Comme attendu dans le cas de régions largement non structurées, la superposition des quatre structures de H3 extraites des groupes les plus représentatifs des trajectoires révèle peu de ressemblance structurale (tableau 2.9 et figure 2.28). En accord avec les tracés de RMSD, les deux charnières les plus proches sont celles issues de R16-21 et  $\Delta 45-47$ .

	R16-21	$\Delta 45-47$	$\Delta 45-48$	$\Delta 45-49$
R16-21	0	7,3	12,3	10,1
$\Delta 45-47$		0	11,1	11,2
$\Delta 45-48$			0	11
$\Delta 45-49$				0

Tableau 2.9 – *Déviat ion quadratique moyenne (en Å) entre les charnières H3 extraites du groupe le plus représentatif de la trajectoire de chaque protéine décrits dans la figure 2.16.*

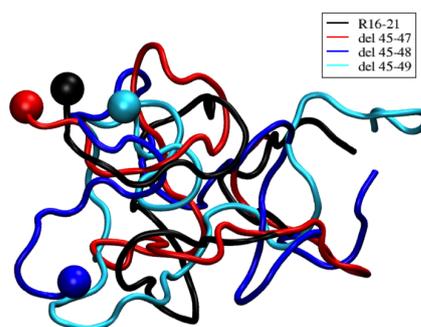


FIGURE 2.28 – *Superposition des charnières H3 extraites du groupe le plus représentatif de la trajectoire de chaque protéine. Les boules indiquent l'extrémité N-terminale de chaque charnière.*

La conformation de la charnière H3 dépend des interactions qu'elle a avec les répétitions environnantes. Ainsi, bien que H3 soit entourée par les mêmes répétitions R19 et R20 dans R16-21 et  $\Delta 45-47$ , les zones d'interactions sont légèrement décalées (figure 2.29). De plus, l'hélice B de la répétition R17//19 de  $\Delta 45-48$  est plus en contact avec la charnière que la boucle AB. Sans surprise, la répétition fracturée de  $\Delta 45//49$  présente une région d'interaction complètement différente avec deux zones d'interactions dans les deux hélices de la répétition fracturée.

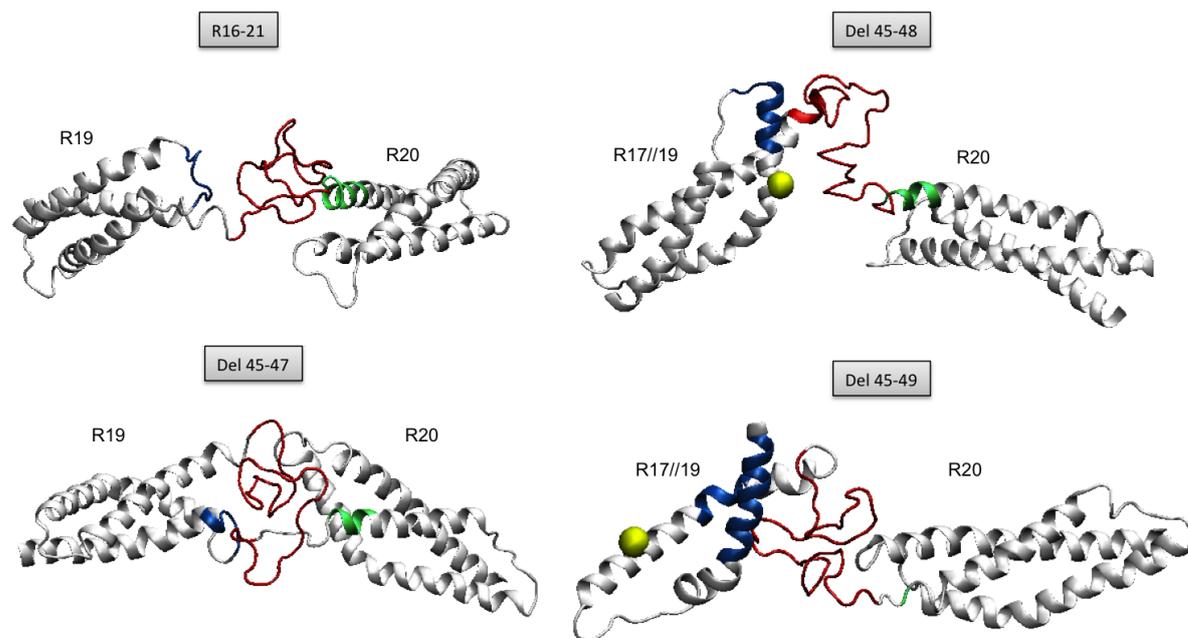


FIGURE 2.29 – Régions en contact avec la charnière H3. Seules les répétitions entourant H3 sont représentées. H3 est en rouge, la région autour de R19 (ou R17//19) en contact est en bleu et le début de l'hélice A de R20 en contact est en vert. La boule jaune indique la jonction de la délétion.

La conformation de la charnière dépend également des interactions entre les acides aminés la constituant. Sur ce premier point, la charnière de  $\Delta 45-48$  se distingue par un grand nombre de liaisons hydrogènes. En effet, on dénombre 13 liaisons hydrogènes sur la structure présentée dans la figure 2.28 contre 6 liaisons H pour H3 de R16-21 et 7 pour les deux autres charnières. La charnière de  $\Delta 45-48$  semble la moins compacte mais le premier acide aminé en position N-terminale se structure en hélice  $\alpha$ , ce qui explique un nombre plus important de liaisons hydrogènes dans la région.

H3 est donc une charnière s'adaptant aux répétitions environnantes en modifiant sa conformation. Etant très mobile, il sera possible d'observer de nouvelles conformations de la charnière dans d'autres protéines tronquées. Par conséquent, les structures de H3 que nous proposons sont assez hypothétiques comparées au reste de nos modèles. Ainsi, il est difficile de caractériser de façon générale H3 mais, comme nous l'avons vu dans l'étude des mouvements (section 2.6), elle apporte de la flexibilité aux protéines tronquées lorsque la mini-charnière entre entre R18 et R19 est absente.

### 2.3.3 La validation par les dynamiques gros-grains

L'utilisation de dynamiques gros-grain permet d'étudier les structures des protéines sur une plus grande échelle de temps ( $5 \mu\text{s}$ ). Les dynamiques gros-grain sont donc utilisées en complément des dynamiques tout-atome afin de proposer un meilleur échantillonnage conformationnel. Cependant, les liaisons hydrogènes ne sont pas prises en compte et les structures secondaires sont, par conséquent, imposées en gros-grain. De plus, le champs de force MARTINI que nous avons choisi décrit moins bien les interactions électrostatiques que les champs de force tout-atome mais décrit précisément les interactions hydrophobes.

Seules les deux protéines  $\Delta 45-49$  et  $\Delta 45-51$  ont subi des simulations gros-grains au cours de ce travail de thèse.

#### 2.3.3.1 La protéine $\Delta 45-49$

La structure servant de point de départ à la dynamique gros-grain est le modèle tout-atome minimisé par NAMD et ainsi les deux trajectoires tout-atome et gros-grain ont le même point de départ.

La trajectoire s'équilibre rapidement (figure 2.30A). Les fluctuations des trajectoires gros-grain et tout-atome sont globalement superposables (figure 2.30B). On observe des différences de fluctuation dans la charnière H3 qui se réarrange différemment dans les deux trajectoires. De plus, on observe une hausse de fluctuation du linker R20-21 dans la trajectoire gros-grain correspondant à la formation d'un coude.

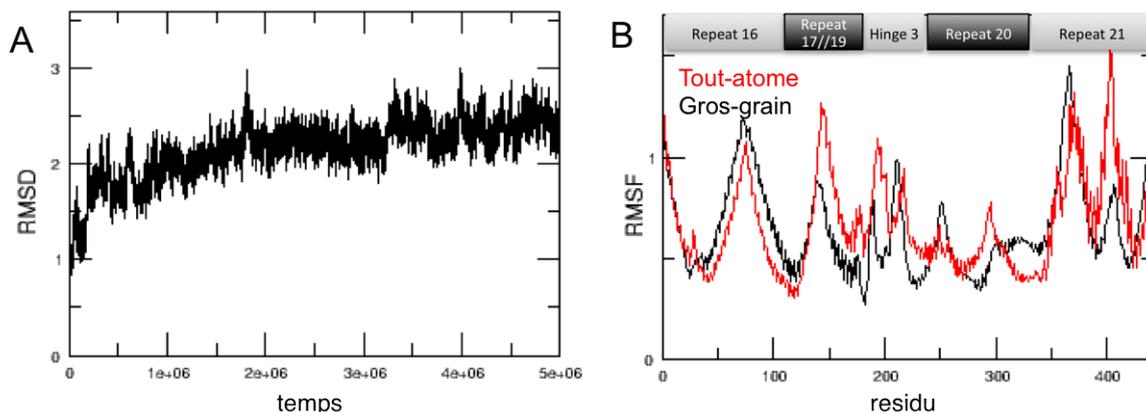


FIGURE 2.30 – A. *Ecart quadratique moyen des  $C\alpha$  de la trajectoire gros-grain de  $\Delta 45-49$  en nm (RMSD).* B. *Fluctuation des structures secondaires au cours de la trajectoire en nm.*

La trajectoire gros-grain a été réduite à sept structures représentatives représentant 81 % de la trajectoire totale (tableau 2.10).

En extrayant les trois structures majoritaires, on remarque qu'elles adoptent le même type de conformation. En effet, le tandem de répétition R16-R17//19 pivote et la charnière devient plus compacte, ce qui a pour conséquence un raccourcissement de la protéine. De plus, on observe la formation du coude correspondant à la hausse de fluctuation observée

	1	2	3	4	5	6	7
1 (14%)	0	8	6,6	9,3	10,4	8,3	10,2
2 (4%)	-	0	8	10,6	12,1	9	13,8
3 (4 %)	-	-	0	5,2	6,7	5,4	8,1
4 (17 %)	-	-	-	0	4,6	4,9	5,8
5 (24 %)	-	-	-	-	0	4,3	5,6
6 (2 %)	-	-	-	-	-	0	7
7 (16 %)	-	-	-	-	-	-	0

Tableau 2.10 – *Déviat*ion quadratique moyenne (en Å) des carbones  $\alpha$  entre les sept structures représentatives majoritaires de la trajectoire gros-grain.

entre R20 et R21 (figure 2.31). Bien que les structures secondaires soient figées, la simulation gros-grain laisse la possibilité aux hélices de couder. Nous avons déjà observé ce type de coude dans une autre simulation gros-grain où le linker R10-11 couvait également.

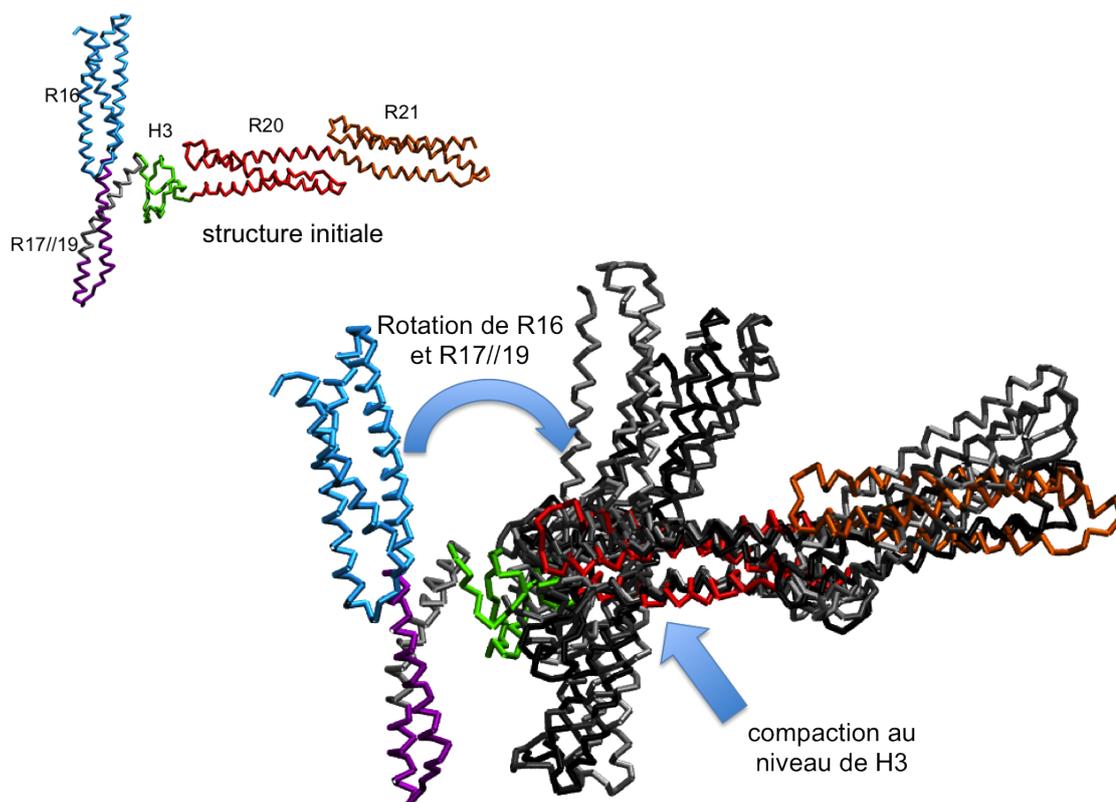


FIGURE 2.31 – *Comparaison de la structure initiale de  $\Delta_{45-49}$  gros-grain avec les trois structures les plus représentatives de la DM. La structure initiale est colorée en fonction des régions la composant. Les 3 structures représentatives sont colorées du gris au noir. Les structures sont superposées sur la répétition 20.*

Ces conformations diffèrent des structures représentatives tout-atome obtenues précédemment, ce qui rend difficile une superposition entre les structures aux deux résolutions (tableau 2.11 et figure 2.32). En effet, bien que les points de flexibilité dans les trajectoires gros-grain et tout-atome soient similaires, les mouvements qu'ils entraînent diffèrent.

GG \ TA	TA		
	0	1	2
1	17,6	15,6	17,7
2	17,8	17,7	18,9
3	19,7	17,7	20,6
4	22,3	20,1	23
5	22,9	20,6	22,9
6	21,3	19,8	20,7
7	23,6	20,7	24

Tableau 2.11 – *Déviation quadratique moyenne (en Å) entre les structures représentatives tout-atome et gros-grain après superposition globale des structures. TA : structures représentatives tout-atome. GG : structures représentatives gros-grain.*

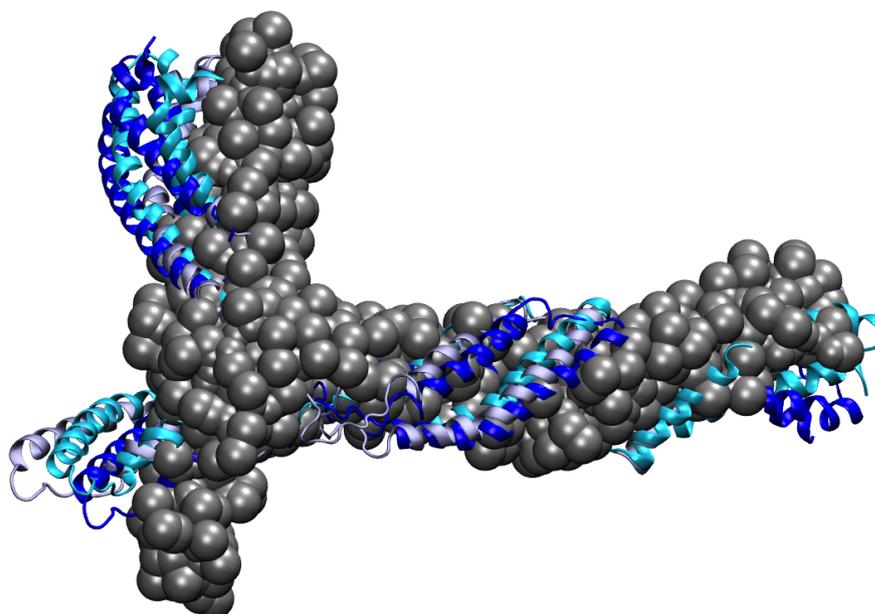


FIGURE 2.32 – *Superposition de la structure représentative 1 de la trajectoire gros-grain (noir) avec les 3 structures représentatives de la trajectoire tout-atome (bleu).*

### 2.3.3.2 La protéine $\Delta 45-51$

Tout comme pour la protéine  $\Delta 45-49$ , le point de départ de la trajectoire gros-grain  $\Delta 45-51$  est le modèle tout-atome minimisé par NAMD.

La trajectoire gros-grain s'équilibre rapidement et on distingue deux conformations majoritaires (figure 2.33A). Les profils de fluctuation sont globalement identiques entre les trajectoires gros-grain et tout-atome (figure 2.33B). Dans les deux trajectoires, la répétition hybride R17//20 présente une fluctuation élevée particulièrement au niveau de l'hélice B indiquant que la répétition hybride est moins stable que les répétitions natives. Les deux répétitions tronquées formant la répétition hybride semblent présenter des propriétés électrostatiques et/ou hydrophobes modifiées engendrant une instabilité de la répétition.

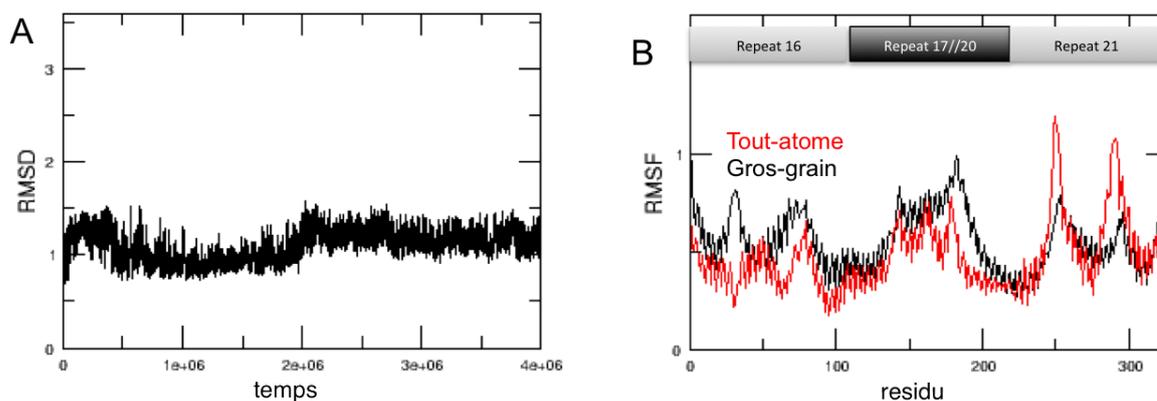


FIGURE 2.33 – A. *Ecart quadratique moyen des  $C\alpha$  de la trajectoire gros-grain de  $\Delta_{45-51}$  en nm (RMSD). B. *Fluctuation des structures secondaires au cours de la trajectoire.**

Trois structures représentatives représentant 79% de la trajectoire totale ont été extraites (tableau 2.12).

	1	2	3
1 (37%)	0	5,4	8,6
2 (35%)	-	0	8,5
3 (7 %)	-	-	0

Tableau 2.12 – *Déviation quadratique moyenne (en Å) des carbones  $\alpha$  entre les trois structures représentatives majoritaires de la trajectoire gros-grain.*

La superposition de ces trois structures met en évidence un mouvement de torsion de R21 entraînant toute la molécule et rapprochant les boucles AB de R16 et BC de R17//20 ainsi que les boucles AB de R17//20 et BC de R21. Ce mouvement de torsion est également observé dans la simulation tout-atome (section 2.6).

La superposition des structures représentatives révèle que la structure représentant le plus grand groupe tout-atome (groupe 0 en rouge dans la figure 2.34) est la plus proche de la structure représentant le plus grand groupe gros-grain (groupe 1 en blanc dans la figure 2.34) (tableau 2.13).

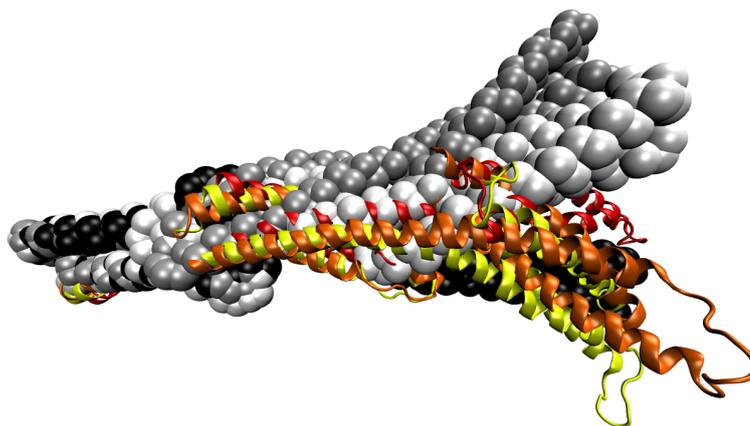


FIGURE 2.34 – *Superposition des structures représentatives gros-grain représentées par des sphères et des structures représentatives tout-atome représentées par des rubans. Les structures sont superposées sur la répétition 16.*

GG \ TA	TA		
	0	1	2
1	14,6	17,6	20
2	18,6	27,7	28,6
3	17,3	23,2	19

Tableau 2.13 – *Déviations quadratiques moyennes (en Å) entre les structures représentatives tout-atome et gros-grain après superposition de la répétition 16. TA : structures représentatives tout-atome. GG : structures représentatives gros-grain.*

L'utilisation de dynamiques gros-grain a permis de mieux échantillonner les conformations possibles des protéines  $\Delta 45-49$  et  $\Delta 45-51$ . Cependant, le champ de force gros-grain MARTINI est moins précis que le champ de force tout-atome pour les interactions électrostatiques et les liaisons hydrogènes. Par conséquent, il est intéressant de coupler les deux approches. Dans l'avenir, nous pouvons envisager d'utiliser les structures représentatives gros-grain comme nouveau point de départ pour des trajectoires tout-atome et réaliser des dynamiques multi-échelle (figure 2.35). Ainsi, les dynamiques tout-atome s'écarteront du biais éventuel des structure-patterns de spectrine utilisées pour construire les modèles par homologie.

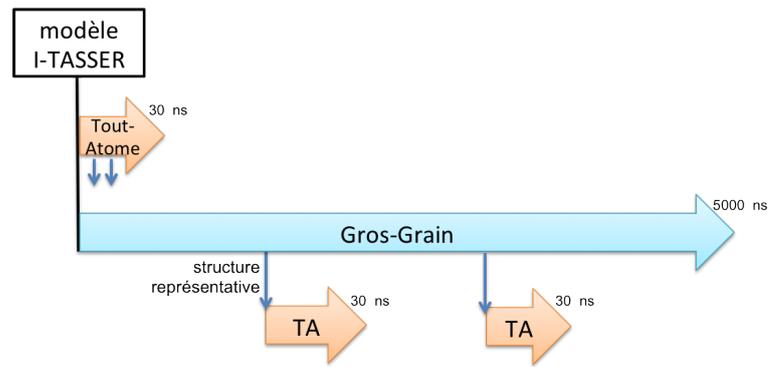


FIGURE 2.35 – Principe de l'approche multi-échelle envisagée.

Nous pourrions également étudier plus précisément les propriétés de surface hydrophobe des structures représentatives gros-grain ou leur éventuelle interaction avec des lipides en les plongeant dans des grilles de potentiels hydrophobes mimant une membrane comme cela à déjà été fait précédemment sur la neurotensine au sein de l'équipe [Da Costa 2012].



# Chapitre 3

## La base de données eDystrophin

eDystrophin est une base de données librement accessible sur internet, développée au sein de l'équipe et complémentaire des bases de données spécifiques du gène *DMD* existantes (cf. section 1.3.2.3). Elle permet une analyse *in silico* des dystrophines mutées, majoritairement tronquées, observées chez les patients présentant une mutation en phase. Elle propose une vue synthétique de l'impact des mutations sur les domaines structuraux et les domaines de liaison des dystrophines tronquées. De plus, lorsqu'il s'agit d'une délétion touchant le domaine central, eDystrophin fournit un modèle structural de la région entourant la délétion afin de mesurer l'impact de la délétion sur la structure de la protéine. Un article correspondant à ce travail a été publié dans Orphanet Journal of Rare Diseases [Nicolas 2012].

### 3.1 Matériel et méthodes

Contrairement aux bases de données gène-spécifiques, eDystrophin n'est pas sous format standard. En effet, il n'existe pas de structure standard de base de données pour les protéines. Par conséquent, nous avons une grande flexibilité pour mettre en place eDystrophin. Je l'ai donc développée pour qu'elle réponde à nos besoins spécifiques.

#### 3.1.1 Matériel

eDystrophin est composée de deux entités : la base de données et le site internet qui est l'interface. Les deux entités sont hébergées sur la plateforme de bioinformatique BioGenouest (<http://www.genouest.org>) et utilisent le serveur web Apache. eDystrophin est librement accessible à l'adresse <http://edydystrophin.genouest.org/>

La base de données utilise le système de gestion de base de données MySQL 5.1.37 (<http://www.mysql.com/>) inclut dans la suite logiciel MAMP (<http://www.mamp.info/en/index.html>). MySQL est un système de gestion de base de données gratuit et très couramment utilisé pour construire des bases de données en bioinformatique.

Le site web a été développé en utilisant les langages XHTML/CSS, PHP 5 (<http://www.php.net/>) et JavaScript. XHTML et CSS sont les langages standards de création de sites internet. PHP 5 est un langage de programmation permettant de créer des sites internet dynamiques et de faire le lien avec la base de données MySQL. PHP permet également, par l'intermédiaire de scripts, d'intégrer, d'analyser et d'extraire les données de la base de données. JavaScript permet d'ajouter des fonctionnalités interactives qui améliore l'attractivité du site internet. J'ai développé la base de données, l'ensemble du site internet et des algorithmes inclus dans eDystrophin.

Les schémas des protéines ont été générés grâce à l'outil MyDomains (<http://prosite.expasy.org/mydomains/>). Pour les schémas des protéines tronquées, j'ai adapté d'un script Python créé par David Goudenège permettant de générer automatiquement les requêtes http.

L'applet Jmol (<http://jmol.sourceforge.net/index.fr.html>) permet de visualiser les structures et les modèles de dystrophine sur le site internet.

### 3.1.2 Construction de la base de données

L'objectif principal d'eDystrophin, défini dès le départ, est d'apporter des informations sur les dystrophines mutées produites suite à des mutations en phase. Plus précisément, il s'agit d'intégrer des informations sur les régions et les liaisons aux partenaires de la dystrophine altérées pour chaque mutation identifiée chez un patient. Nous avons envisagé de réaliser une version privée de la base de données avec des informations sur la structure des protéines mutées et sur les phénotypes des patients mais nous avons finalement décidé de rendre toutes les informations publiques. Le deuxième objectif est de rassembler le plus exhaustivement possible les données disponibles sur la dystrophine sauvage.

#### 3.1.2.1 Les données génétiques

Les séquences des 16 ADN complémentaires (ADNc) issues des sept promoteurs du gène *DMD* ont été obtenues dans la base de données GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) sous forme de fichiers XML contenant toutes les informations sur les séquences ADNc. Ces données ont été extraites puis intégrées dans eDystrophin par l'intermédiaire d'un analyseur syntaxique (parseur) PHP.

#### 3.1.2.2 Les données sur la dystrophine native

Les séquences des 16 isoformes de dystrophine native ont été extraites du fichier XML contenant les séquences des ADNc grâce au même analyseur syntaxique.

Les positions des domaines structuraux et fonctionnels de la dystrophine musculaire Dp427m ont été extraites de 19 articles et 3 outils de recherches de domaines. Au total 35 domaines structuraux et 15 domaines de liaison ont été identifiés sur la dystrophine.

En fonction des études publiées, il peut y avoir un léger décalage dans la définition des domaines, par conséquent, toutes les variations de position des domaines ont été intégrées dans eDystrophin avec la référence originale. Cependant, un domaine par défaut a été choisi à chaque fois pour être utilisé, par la suite, sur les dystrophines mutées.

Les deux domaines structuraux dont la structure 3D a été résolue par cristallographie à rayons X ont été intégrés dans eDystrophin : le domaine de liaison à l'actine ABD1 (PDB: 1DXX), et le domaine de liaison au  $\beta$ -dystroglycane comprenant le domaine WW, et deux mains EF (PDB: 1EG3). Les modèles par homologie des répétitions de la dystrophine que nous avons réalisés sont également disponibles [Legrand 2011].

### 3.1.2.3 Les données cliniques

Afin de collecter le plus exhaustivement possible les mutations en phase du gène *DMD* et d'obtenir des données cliniques fiables, nous avons collaboré avec le Laboratoire de Biochimie et de Génétique de l'Institut Cochin (LGBM) appartenant au réseau de l'UMD-DMD [Tuffery-Giraud 2009]. Nous avons également collecté des données dans des études précédemment publiées. Au total, nous avons collecté des données sur 935 patients : 531 provenant du LGBM et 414 provenant de la littérature. Les données provenant du LGBM étaient sous la forme d'un fichier csv, facilement lisibles par un analyseur syntaxique créé pour intégrer les données dans eDystrophin. Les données provenant de la littérature ont été manuellement intégrées dans un autre fichier csv puis intégrées dans eDystrophin par l'intermédiaire du même analyseur syntaxique.

Les patients sont répartis dans 6 groupes phénotypiques en fonction de la classification internationale en vigueur :

- **DMD** : Patients ayant perdu la marche avant 12 ans.
- **IMD** : Patients ayant perdu la marche entre 12 et 16 ans.
- **BMD** : Patients ayant perdu la marche après 16 ans.
- **Asymptomatic** : Patients ne présentant pas de symptômes musculaires et cardiaques à leur dernière visite.
- **Pending** : Patients avec trop peu d'informations cliniques pour être classifiés (généralement trop jeunes).
- **DCM** : Patients présentant une cardiomyopathie isolée.

Les résultats des examens de biochimie utilisés pour le diagnostic sont disponibles pour 360 patients. Ces résultats comprennent les résultats d'immuno-coloration et/ou de Western blots réalisés avec trois anticorps anti-dystrophine ciblant des régions différentes de la protéine. Les données d'immuno-coloration provenant du LGBM sont présentées en fonction de la régularité de la coloration (normale, régulière, irrégulière, mosaïque, sans signal mais avec des fibres révertantes, sans signal) et de l'intensité (élevée, moyenne, faible). Les Western blots ont été réalisés avec les trois anticorps et décrits en fonction de la quantité de fluo-

rescence (élevée, moyenne, faible) et de la taille (élevée, normale, réduite ou non détectée) de la dystrophine. Toutes ces données ont été synthétisées en deux éléments dans eDystrophin : niveaux d'expression de la dystrophine (élevé, normal, faible) et taille de dystrophine détectée (petite, normale et large).

De plus, lorsque les données sur la présence d'une cardiomyopathie et/ou d'un retard mental sont disponibles, elles sont intégrées dans eDystrophin.

Dans eDystrophin, toutes les données cliniques sont enregistrées par patient. Cependant, le but d'eDystrophin n'est pas de fournir des informations cliniques pour chaque patient mais de présenter une vue globale des atteintes symptomatiques par mutation. Par conséquent, les données cliniques disponibles sur le site internet sont présentées sous forme d'effectifs et de moyennes pour chaque mutation.

### 3.1.2.4 Les données sur les dystrophines mutées

L'innovation d'eDystrophin réside dans les informations que la base de données déduit sur la structure et le maintien des interactions des dystrophines tronquées à partir des mutations en phase collectées. L'approche est décrite dans la figure 3.1.

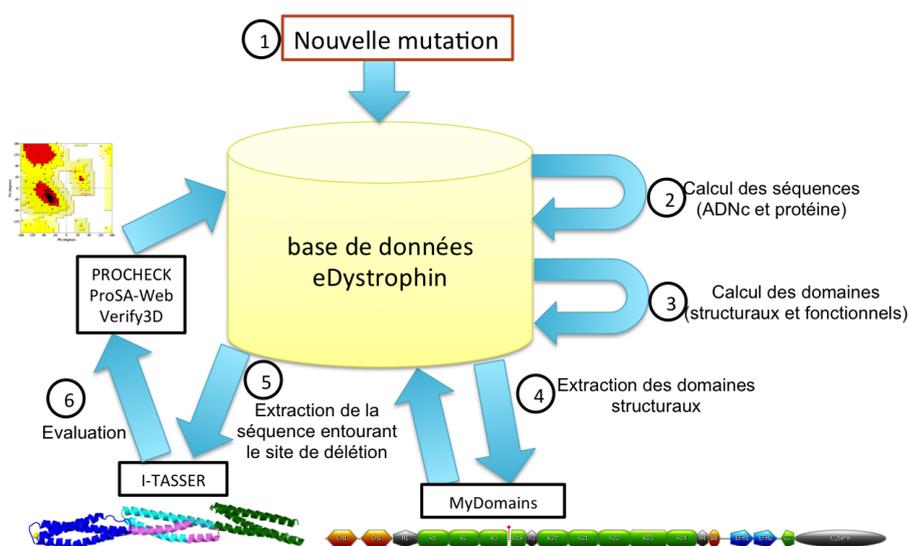


FIGURE 3.1 – Méthode générale d'ajout de nouvelles mutations dans eDystrophin. 1. Une nouvelle mutation est insérée dans eDystrophin. 2. Les séquences d'ADNc et d'acides aminés correspondantes sont calculées. 3. La présence ou l'absence des domaines structuraux et fonctionnels est calculée sur la protéine mutée. 4. Les positions des domaines structuraux sont extraites de la base de données afin d'être soumises à MyDomains qui schématise la protéine. 5. La séquence d'acides aminés entourant le site de délétion est extraite puis soumise au serveur de modélisation par homologie I-TASSER. 6. Les modèles issus de I-TASSER sont évalués par plusieurs logiciels et les résultats intégrés dans la base de données.

### Les mutations

Les mutations observées chez les patients ont été intégrées à la base de données en même temps que les données cliniques. Au total, eDystrophin comprend 209 mutations différentes

provenant de 945 patients. 100 mutations proviennent d'études précédemment publiées, 109 ont été fournies par le LGBM dont 42 sont communes aux 2 sources. Les mutations ont été nommées selon les recommandations de l'Human Genome Variation Society.

### Les séquences mutées

Pour chaque mutation, les séquences d'ADNc du transcrit musculaire (dp427m) et des transcrits plus courts (dp260, dp140, dp116 et dp71) sont déduites grâce à un algorithme qui extrait les positions de la mutation. Puis les séquences sont, soit dupliquées en tandem dans le cas de duplications, soit tronquées dans le cas de délétions, soit simplement mutées dans le cas de mutations ponctuelles. Les séquences d'ADNc sont ensuite traduites. Les 209 séquences d'ADNc et protéiques sont ensuite intégrées à la base de données.

Les tailles des séquences et les masses moléculaires sont également calculées et intégrées à la base de données.

### Les domaines structuraux et fonctionnels

Toutes les séquences mutées de l'isoforme musculaire (dp427m) de la dystrophine sont soumises à un algorithme qui calcule si les domaines caractérisés sur la dystrophine native sont présents ou non sur les dystrophines mutées. Cet algorithme calcule si les domaines sont entièrement présents, totalement absents ou partiellement présents et précise quelle extrémité est présente. Pour cela, l'algorithme compare la position de la mutation avec les positions des différents domaines (figure 3.2). Dans un deuxième temps, la présence de ces domaines est calculée sur les isoformes plus courtes.

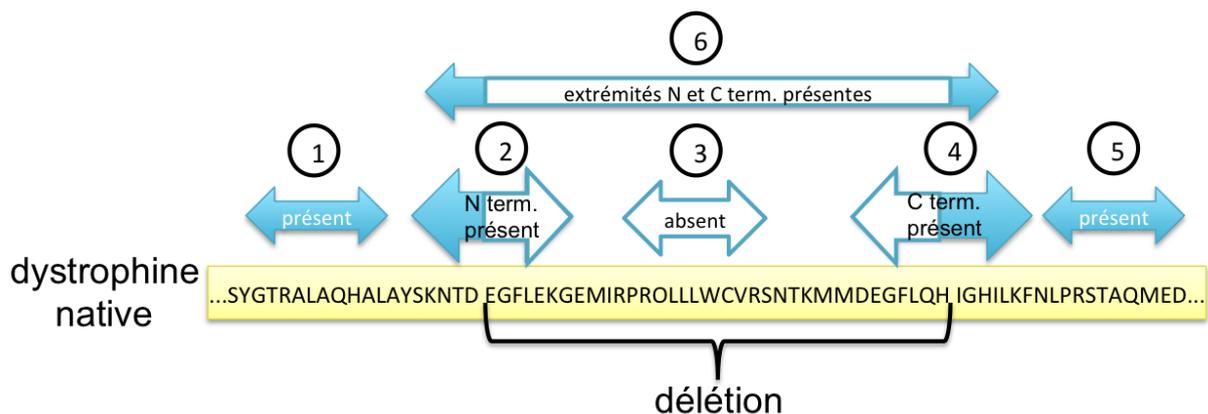


FIGURE 3.2 – Principe de l'algorithme de calcul des domaines sur une protéine tronquée. L'algorithme compare la position de la délétion (représentée par une accolade) avec les positions de chaque région identifiée sur la protéine native (représentées par des doubles flèches). 1. Si la région est en position N-terminal de la mutation, elle est présente. 2. Si la région est à cheval sur l'extrémité N-terminale de la délétion, la partie N-terminale de la région est présente. 3. Si la région est incluse dans la délétion, elle est absente. 4. Si la région est à cheval sur l'extrémité C-terminale de la délétion, la partie C-terminale de la région est présente. 5. Si la région est en position C-terminal de la délétion, elle est présente. 6. Si la région est plus grande et englobe la délétion, les parties N- et C-terminales de la région sont présentes.

Les domaines structuraux (choisis par défaut) sont extraits et écrits dans un fichier texte afin de générer automatiquement le schéma des protéines mutées par l'intermédiaire d'un algorithme soumettant automatiquement toutes les protéines tronquées au logiciel MyDomain. Ces schémas sont ensuite référencés dans la base de données.

### Les modèles par homologie

Pour toutes les protéines présentant une délétion dans le domaine central, la structure 3D de la jonction de la délétion a été modélisée. Les modèles par homologie étant centrés sur la jonction, un algorithme a extrait uniquement les séquences protéiques entourant la jonction, c'est-à-dire le domaine structural où se situe la jonction et les domaines structuraux précédents et suivants. Ces fragments de protéines ont été modélisés selon le protocole décrit dans la section 2.1.1. Les modèles structuraux ont ensuite été soumis à plusieurs programmes classiques de vérification de la qualité des modèles (PROCHECK [Laskowski 1993], ProSA-web et Verify3D [Luthy 1992]). Tous les résultats de ces logiciels sont ensuite intégrés dans la base de données. Cependant, seul le graphe de Ramachandran décrivant la distribution des acides aminés en fonction des structures secondaires est accessible sur le site internet.

#### 3.1.2.5 Mises à jour de la base de données

De nouvelles données peuvent être introduites dans eDystrophin, comme par exemple de nouvelles données cliniques, de nouvelles mutations ou un nouveau domaine de liaison à un partenaire. Ces mises à jour sont principalement des demandes provenant de l'équipe ou de nos collaborateurs mais il est possible de nous contacter par l'intermédiaire du site web d'eDystrophin pour demander l'insertion de nouvelles données. Ces mises à jour se font de façon semi-automatique. En effet, les données intégrées dans eDystrophin sont d'abord manuellement validées afin de n'intégrer que des données fiables. Un des points de vérification important est la technique utilisée pour identifier une mutation chez un patient. Il est important que ces mutations soient délimitées de façon fiable soit par MLPA, soit par mPCR des 79 exons, soit par séquençage complet du gène.

Une fois que les nouvelles données sont validées, plusieurs algorithmes permettent de mettre à jour la base de données. Cette mise à jour se fait en plusieurs étapes notamment à cause de l'utilisation des outils externes MyDomains et I-TASSER. La mise à jour s'étale donc sur quelques jours à cause du délai de réponse d'I-TASSER (72h maximum). Il pourrait être envisagé à l'avenir d'installer I-TASSER localement pour automatiser et accélérer la mise à jour. Comme c'est le cas pour de nombreuses bases de données supervisées par des curateurs, la mise à jour se fait dans un premier temps sur la version locale puis, une fois validée, la mise à jour est effectuée sur le serveur de la version publique.

De plus, il est envisagé de relier eDystrophin à la base de données UMD-DMD en intégrant des liens sur les pages des protéines mutées.

## 3.2 Résultats

### 3.2.1 La base de données relationnelle

La base de données relationnelle d'eDystrophin comprend 18 tables (voir figure 3.3) : 7 tables contiennent les informations sur la dystrophine native et le gène sauvage, 9 tables contiennent les informations liées aux mutations (patients, mutations, dystrophines mutées), une table est commune aux deux parties et une table contient les données des utilisateurs enregistrés. Nous avons choisi de garder les données générées par des logiciels externes sous format de fichiers et de simplement référencer le nom des fichiers dans les tables. C'est par exemple le cas des fichiers de coordonnées (pdb) des modèles structuraux et des fichiers images (ex : MyDomains).

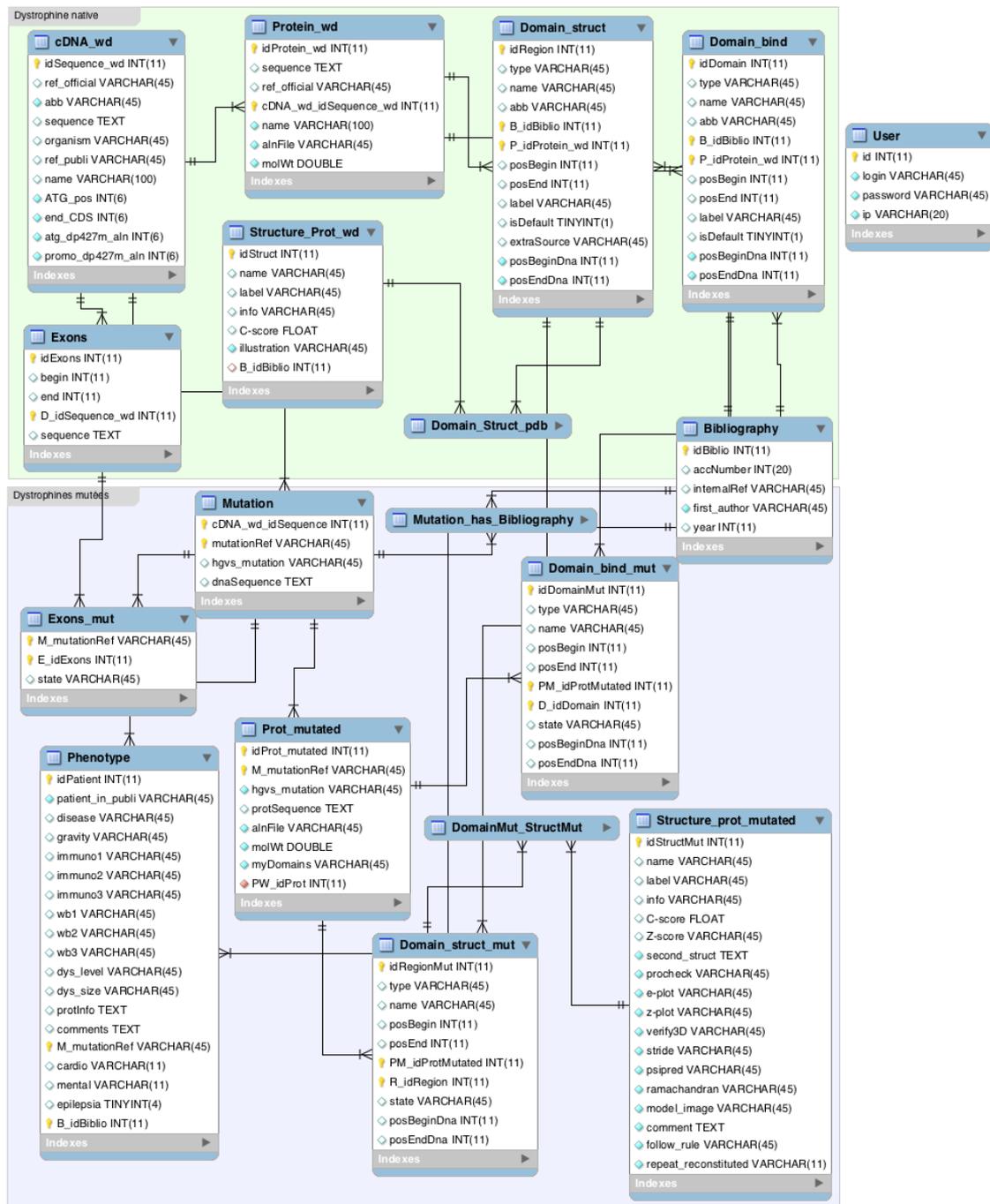


FIGURE 3.3 – Diagramme de la structure de la base de données

Les tables de la section *Wild-type dystrophin* (en vert dans la figure précédente) contiennent principalement des données collectées dans la littérature et dans les bases de données. Les séquences des différents transcrits et leurs références officielles sont dans la table *cDNA\_wd*. Les positions et la séquence de chaque exon du transcrit musculaire (dp427m) sont notés dans la table *Exons*. A chaque transcrit de la table *cDNA\_wd*, correspond une protéine dans la table *Protein\_wd*. Les informations sur les domaines structuraux caractérisés sur la dystrophine native dp427m sont enregistrés dans la table *Domain\_struct*. De la même manière, la table *Domain\_bind* contient les données sur les domaines de liaison. Les fichiers pdb contenant les coordonnées des domaines structuraux

sont référencés dans la table *Structure\_Prot\_wd* reliée à la table *Domain\_struct* par une table intermédiaire *Domain\_Struct\_pdb*. Chaque domaine structural, chaque domaine de liaison et chaque structure 3D sont enregistrés avec leur source. Toutes ces sources sont rassemblées dans la table *Bibliography*.

La section *Mutated dystrophin* (en violet dans la figure précédente) contient les séquences mutées dans la table *Mutation*. Les références de chaque mutation sont inscrites dans la table *Bibliography*. Pour chaque séquence mutée, les exons présents, absents ou modifiés sont enregistrés dans la table *Exons\_mut*. De plus, les informations sur les séquences primaires mutées sont inscrites dans la table *Protein\_mutated*. Les positions des domaines structuraux déduits des séquences protéiques sont enregistrées dans la table *Domain\_struct\_mut* et celles des domaines de liaisons sont enregistrées dans la table *Domain\_bind\_mut*. Les modèles par homologie générés par I-TASSER et les fichiers des logiciels d'évaluation sont référencés dans la table *Structure\_prot\_mutated*.

### 3.2.2 L'interface d'eDystrophin

L'interface d'eDystrophin est un site internet librement accessible. Il comporte quatre grandes parties :

- **La section *Knowledge*** n'est pas liée à la base de donnée mais apporte une vue d'ensemble sur la dystrophine en fournissant des informations sur le gène, la protéine, les pathologies associées et les thérapies envisagées.
- **La section *Explore database*** est l'interface où il possible d'interroger la base de données et de télécharger les informations. Cette section est divisée en deux parties : *Wild-type dystrophin* et *Mutated dystrophin*.
- **La section *Statistics*** fournit les statistiques d'eDystrophin.
- **La section *Links*** rassemble quelques liens utiles et l'ensemble des publications référencées dans eDystrophin.

#### 3.2.2.1 La page *Wild-type dystrophin*

Cette page contient toutes les informations sur les séquences d'ADNc et des protéines natives et les structures 3D des domaines de la dystrophine native. L'ensemble des séquences de cette page sont en format fasta et sont facilement enregistrables par copier-coller dans un fichier texte.

Les séquences d'ADNc et d'acides aminés des 16 isoformes sont accessibles par l'onglet *Isoform full length sequences (cDNA and protein)* (figure 3.4 A).

Les séquences des exons sont accessibles par l'onglet *Exon sequences* (figure 3.4 B). Dans cet onglet, il suffit de cliquer sur un exon du schéma pour voir sa séquence apparaître à

droite et avoir la possibilité de la copier.

Les séquences des domaines structuraux et les domaines de liaison sont accessibles, respectivement, par l'onglet *Structural domain sequences* et par *Binding domain sequences* (figure 3.4 C). De la même façon que dans l'onglet *Exon sequences*, les séquences d'intérêt sont accessibles en les sélectionnant sur le schéma de la protéine. Il peut y avoir plusieurs études publiées positionnant les domaines de façon légèrement différentes. Un tableau permet de choisir la "version" de la séquence du domaine à afficher. La "version" du domaine choisie par défaut dans eDystrophin est indiquée.

Les deux structures obtenues par cristallographie aux rayons X précédemment publiées et les modèles 3D des répétitions que nous avons précédemment publiés [Legrand 2011] sont disponibles dans l'onglet *3D-structure models* (figure 3.4 D). Dans cet onglet, les structures 3D peuvent être directement visualisées par l'intermédiaire de l'applet Jmol. Un schéma annoté, sur la droite des modèles, rappelle l'organisation des structures tridimensionnelles. Les fichiers de coordonnées peuvent également être téléchargés.

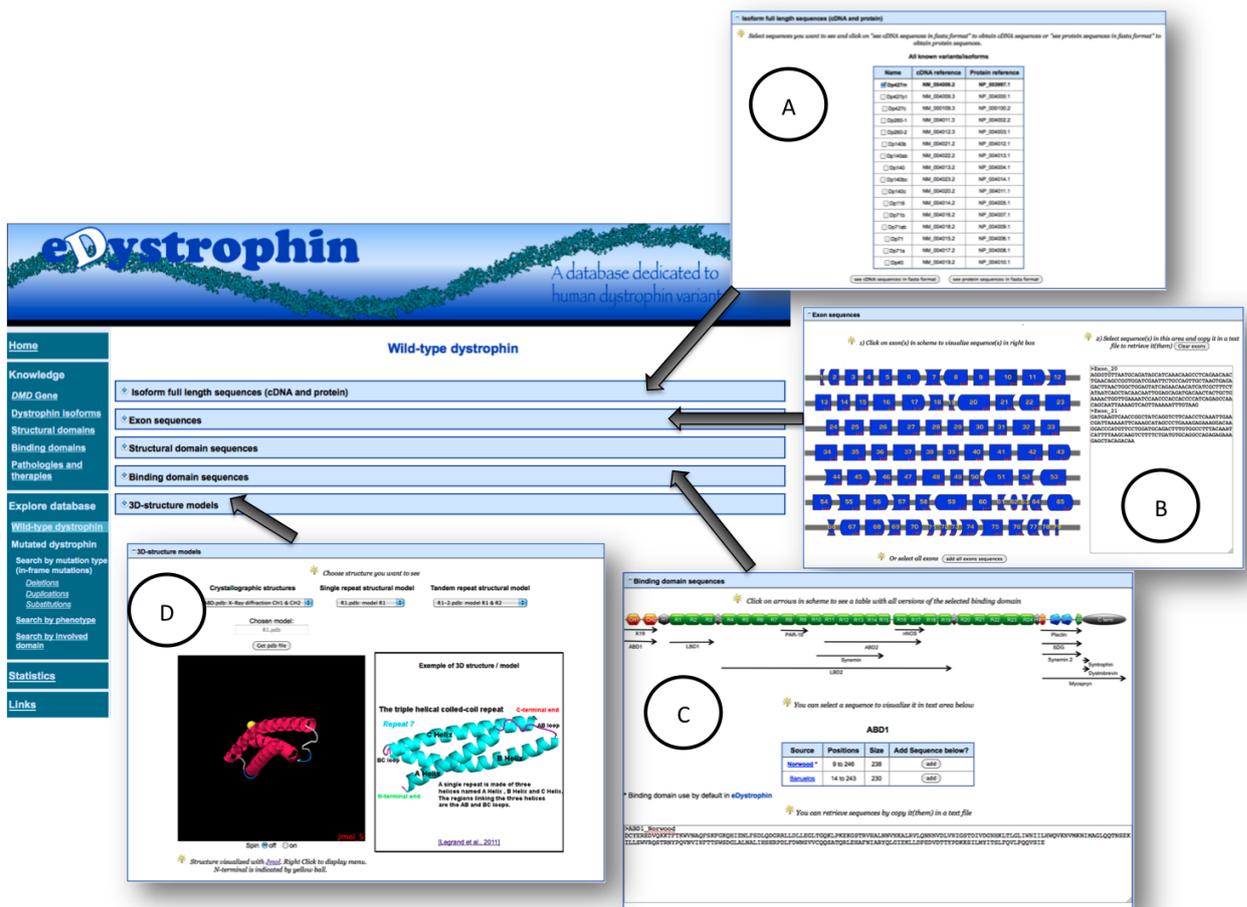


FIGURE 3.4 – Captures d'écran de la page Wild-type dystrophin. A. Onglet Isoform full length sequences (cDNA and protein). B. Onglet Exon sequences. C. Onglet Structural domain sequences. D. Onglet 3D-structure models.

### 3.2.2.2 Les pages *Mutated dystrophin*

Ces pages sont dédiées aux 209 mutations en phase incluses dans eDystrophin. Trois types de recherches sont proposés : par type de mutation, par phénotype associé ou par domaine touché.

La recherche par type de mutation (*Search by mutation type*) permet à l'utilisateur de rechercher des délétions, des duplications ou des substitutions. Lorsque *Deletions* ou *Duplications* est choisi, un schéma représentant les exons apparaît (figure 3.5). La sélection d'un exon sur le schéma affiche une liste de délétions ou duplications affectant cet exon dans laquelle l'utilisateur peut choisir la mutation qui l'intéresse. Il est également possible de sélectionner la mutation d'intérêt directement dans le menu déroulant en bas de la page.

The screenshot displays a web interface for searching mutations. The main area shows a schematic of 79 exons, numbered 1 to 79, arranged in a grid. Exons 1-12 are in the first row, 13-23 in the second, 24-33 in the third, 34-43 in the fourth, 44-53 in the fifth, 54-65 in the sixth, and 66-79 in the seventh. A lightbulb icon indicates '1) Select by click one exon involved in deletion'. To the right, a list of 32 deletion options is shown, each with a radio button. The list is titled 'Exon chosen: 16' and includes options like 'deletion exons 3-16', 'deletion exons 3-18', etc. A 'Get list as csv' button is at the top right. At the bottom, there is a dropdown menu showing 'deletion exons 2-7' and a 'see protein' button.

FIGURE 3.5 – Capture d'écran de la page de la page *Search by mutation type*.

Lorsque *Substitutions* est sélectionné, l'utilisateur choisit la substitution d'intérêt directement dans le tableau présenté.

Il est également possible de faire une recherche par phénotype en sélectionnant *Search by phenotype*. L'utilisateur peut alors rechercher toutes les mutations qui ont été identifiées chez des patients présentant un phénotype spécifique en sélectionnant *affecting at least one patient*, ou au contraire afficher toutes les mutations où un phénotype spécifique n'a pas été identifié en sélectionnant *affecting no patient*. Les phénotypes que l'utilisateur peut choisir sont ceux décrits dans eDystrophin : Asymptomatic, BMD, IMD, DMD, DCM, Pending.

La recherche par domaine (*Search by involved domain*) permet de rechercher toutes les mutations touchant un domaine de liaison ou un domaine structural de la dystrophine. Le schéma de la dystrophine apparaît et permet de choisir de façon intuitive le domaine d'intérêt

puis de sélectionner une mutation dans la liste qui apparaît.

Les listes de mutations sont téléchargeables sous format csv.

Quelque soit la manière d’interroger eDystrophin, après avoir choisi une mutation et cliqué sur le bouton *see protein*, une nouvelle page apparaît (figure 3.6). Cette page rassemble toutes les informations sur la mutation et la protéine mutée présentes dans la base de données. Cette page est constituée d’un tableau résumé et de quatre onglets ouvrables. L’ensemble des informations de cette page est téléchargeable sous format PDF.

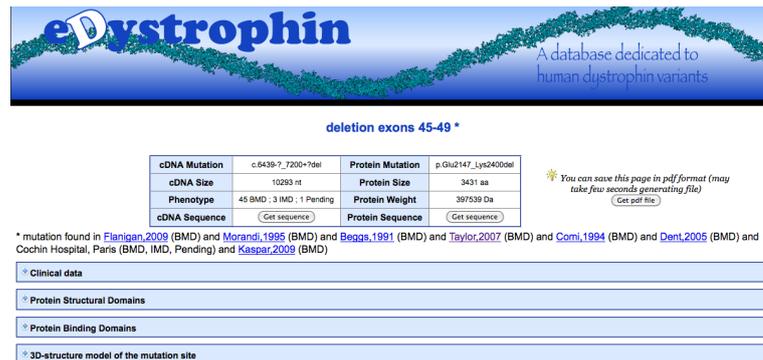


FIGURE 3.6 – Capture d’écran de la page de la délétion des exons 45 à 49.

Le tableau résumé contient la nomenclature de la mutation sur l’ADNc et sur la protéine, la taille des séquences nucléotidiques et protéiques, la masse moléculaire de la protéine et les phénotypes associés à cette mutation.

L’onglet *Clinical data* rassemble les données sur l’ensemble des patients porteurs de cette mutation et présents dans la base de données (figure 3.7). La répartition des phénotypes est représentée. Lorsque les données sont présentes, une représentation en camembert présente le niveau de dystrophine détectée dans les muscles, la taille de la dystrophine détectée, la présence éventuelle de cardiomyopathie et de retard mental à la dernière visite médicale.

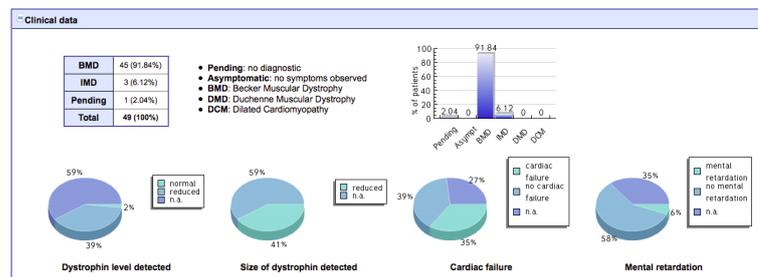


FIGURE 3.7 – Capture d’écran de l’onglet clinique de la délétion des exons 45 à 49.

L’onglet *Protein Structural Domains* présente un schéma de la dystrophine mutée avec le site de mutation indiqué par un spot rouge et les domaines structuraux présents (figure 3.8). Un tableau indique si les isoformes plus courtes sont touchées par la mutation.

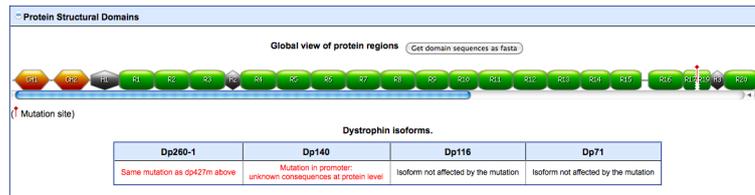


FIGURE 3.8 – Capture d’écran de l’onglet domaines structuraux de la délétion des exons 45 à 49.

L’onglet *Protein binding domains* contient un tableau résumant les conséquences de la mutation sur les domaines de liaison (figure 3.9).

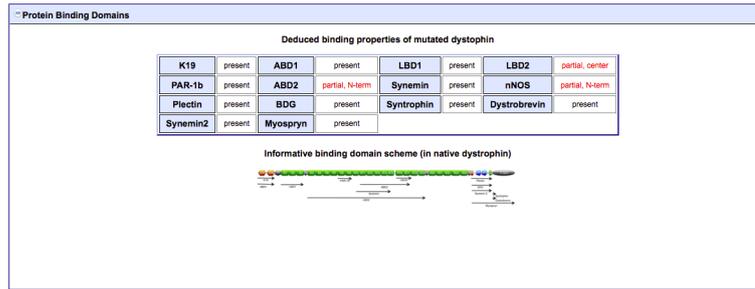


FIGURE 3.9 – Capture d’écran de l’onglet domaines de liaison de la délétion des exons 45 à 49.

L’onglet *3D-structure model of the mutation site* est disponible pour les protéines tronquées dans le domaine central (figure 3.10). Cet onglet présente le modèle structural de la région entourant le point de jonction de la délétion. Ce modèle est visible sur le site par l’intermédiaire de l’applet Jmol et téléchargeable sous format PDB. Le graphe de Ramachandran, indiquant la distribution des acides aminés en fonction de la structure secondaire est également téléchargeable. A droite du modèle, un schéma de trois répétitions de la dystrophine rappelle la structure native du domaine central. De plus, un commentaire décrivant l’impact de la mutation sur la structuration de la dystrophine est affiché.

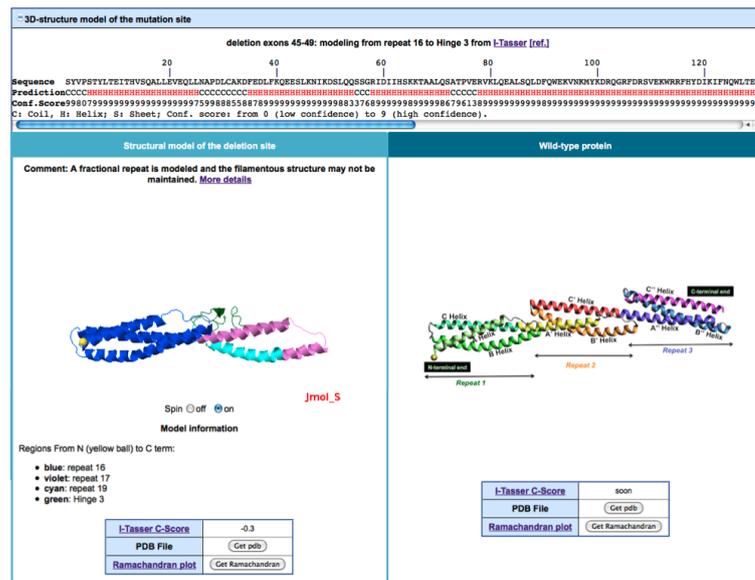


FIGURE 3.10 – Capture d’écran de l’onglet modèle 3D du site de jonction de la délétion des exons 45 à 49.

### 3.2.3 Vue d'ensemble du contenu d'eDystrophin

#### 3.2.3.1 Les statistiques d'eDystrophin

La base de données eDystrophin contient 209 mutations en phase (figure 3.11 A). Ces mutations sont majoritairement des délétions d'un ou plusieurs exons (61%). Les duplications d'un ou plusieurs exons représentent 24% des 209 mutations. Les substitutions et les petites délétions sont minoritaires et représentent respectivement 11% et 4%.

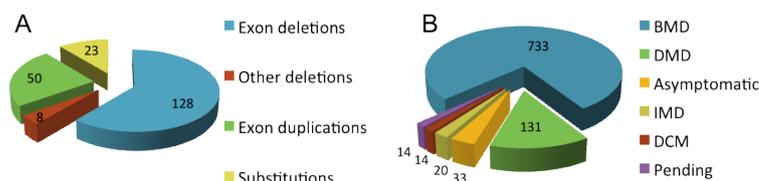


FIGURE 3.11 – *Statistiques d'eDystrophin. A. Répartition des différents types de mutations dans eDystrophin. B. Répartition des différents phénotypes dans eDystrophin.*

Les 945 patients répertoriés dans eDystrophin présentent différents phénotypes (figure 3.11 B). La grande majorité des patients (78%) présente un phénotype BMD. Il est intéressant de noter que 131 patients (14%) sont diagnostiqués DMD et 20 patients (2%) sont de phénotype IMD. Ces patients sont des exceptions à la règle de Monaco (section 1.2.4).

Beaucoup de mutations présentes dans eDystrophin sont associées à plusieurs phénotypes. Ainsi, 111 mutations (53%) ont au moins un patient BMD associé et 104 mutations (50%) ont au moins un patient DMD associé. Le phénotype asymptomatique n'est associé qu'à 10% de mutations.

Les délétions d'exons associées à au moins un patient DMD sont majoritairement situées dans le domaine ABD1 et le domaine riche en cystéine alors que les délétions associées à au moins un patient BMD sont majoritairement situées dans le domaine central (figure 3.12). En effet, 18 délétions commençant à l'exon 3 ont été référencées chez des patients présentant un phénotype DMD alors que seulement 9 délétions débutant à l'exon 3 ont été référencées chez des patients BMD. Ces observations sont en accord avec les précédentes études qui indiquent que les délétions dans les domaines ABD1 et riches en cystéine sont plus délétères [Beroud 2007, Tuffery-Giraud 2009]. Cependant, nous avons également relevé des mutations dans le domaine central associées à des phénotypes DMD.

On remarque également que les duplications d'exons sont le plus souvent associées à un phénotype DMD (31 duplications) qu'à un phénotype BMD (24 délétions). Ceci indiquerait que les duplications sont très délétères.

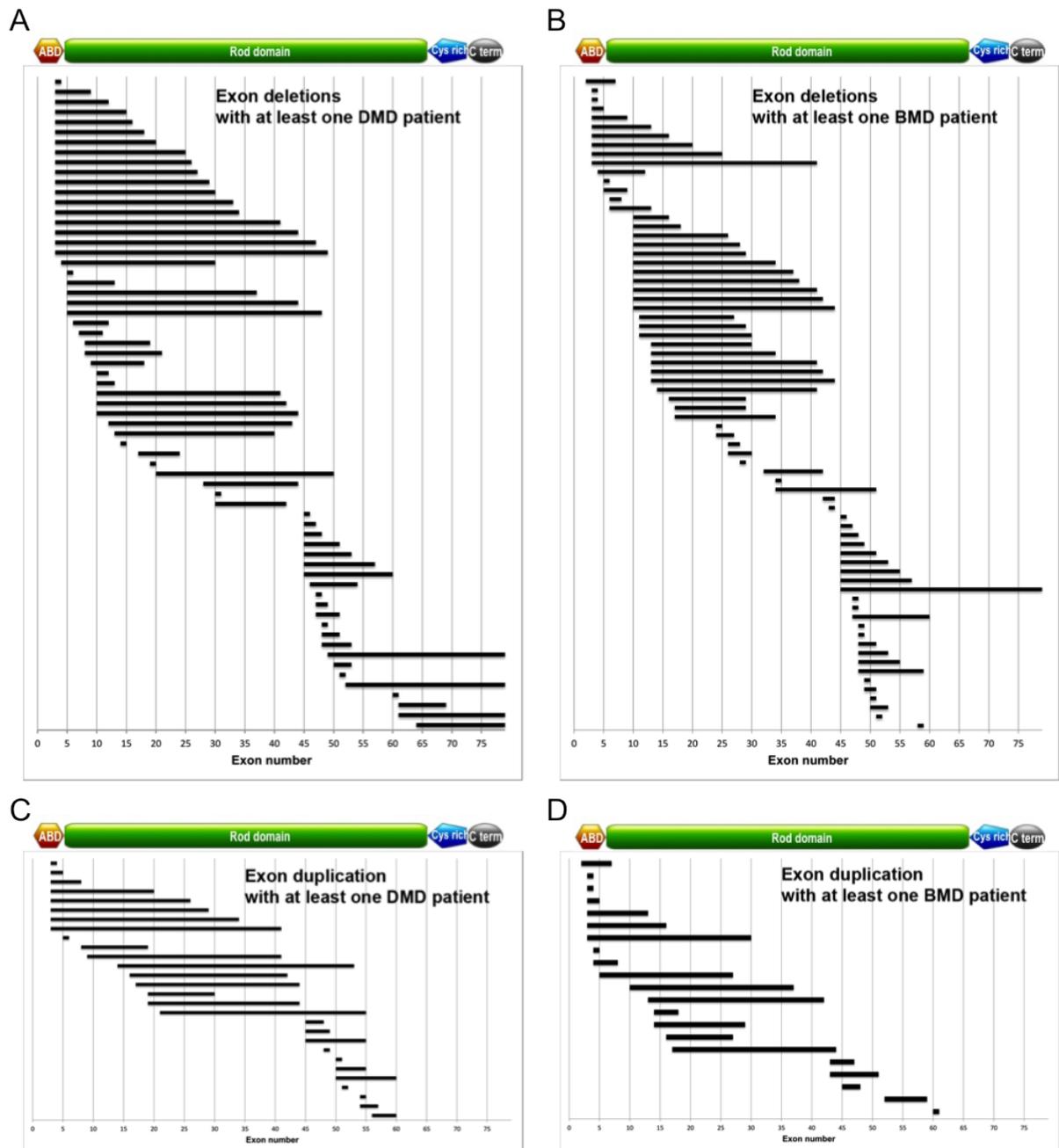


FIGURE 3.12 – Répartition des délétions d'exons (A,B) et duplications (C,D) associées avec au moins un DMD (A,C) ou un BMD (B,D).

### 3.2.3.2 Etude de cas

Pour illustrer l'utilisation d'eDystrophin, j'ai choisi de décrire mutations localisées dans le point chaud de mutation : la délétion des exons 45 à 47 (c.6439-?\_6912+?del/p.Glu2147\_Lys2304del) et la délétion des exons 45 à 55 (c.6439-?\_8217+?del/p.Glu2147\_Gln2739del). J'ai choisi la délétion des exons 45 à 47 car elle est la plus fréquente dans eDystrophin et la délétion des exons 45 à 55 car elle est la cible privilégiée de la thérapie par sauts d'exons multiples.

La manière la plus simple d'accéder aux pages de ces protéines mutées est de choisir

Search by mutation type / Deletions puis de choisir la mutation d'intérêt directement dans la liste apparaissant en bas de la page.

**La délétion des exons 45 à 47**

Le tableau résumé s'affichant sur la page de la délétion d'exons 45-47 indique que la séquence d'ADNc est composée de 10581 nucléotides et que la séquence protéique est composée de 3527 acides aminés. La protéine déduite a un poids moléculaire de 408 kDa. La protéine produite est donc plus petite que la dystrophine native (3685 acides aminés, 427 kDa). Cette mutation a été observée chez des patients présentant différents phénotypes mais est majoritairement associée à un phénotype BMD. Les patients ont été collectés dans huit études précédemment publiées et auprès du LGBM comme indiqué sous le tableau.

En cliquant sur *Get pdf file*, il est possible de télécharger l'ensemble des données de la page dans une fiche au format PDF (figure 3.13 et 3.14).

L'onglet *Clinical data* permet d'avoir accès aux données cliniques avec plus de détails (figure 3.13). On note alors que la délétion a majoritairement été identifiée chez des patients BMD (~95%). On remarque que, lorsque les données sont disponibles, la dystrophine détectée dans les muscles des patients est de taille réduite et présente en quantité moindre par rapport à un muscle sain. En majorité, les patients ne présentent ni cardiomyopathie, ni retard mental lors de leur dernier examen médical.

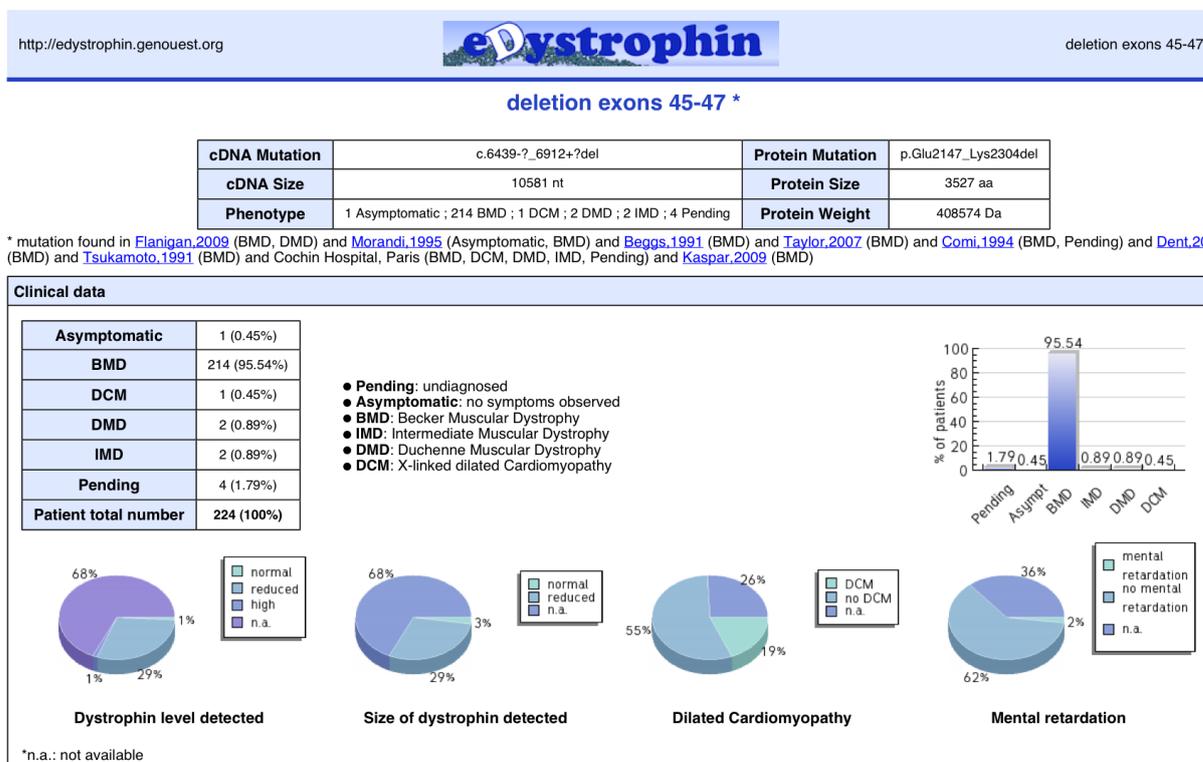


FIGURE 3.13 – Première page de la fiche-résumé de la délétion d'exons 45-47 téléchargeable sur l'interface d'edystrophin.

L'onglet *Protein Structural Domains* présente un schéma de la dystrophine tronquée montrant la répétition 17 tronquée en position C-terminale et la répétition 18 tronquée en position N-terminale (figure 3.14). En cliquant sur le bouton *Get domain sequences as fasta*, les séquences d'acides aminés des domaines structuraux au format fasta sont disponibles en téléchargement.

La mutation touche l'isoforme Dp260 de la même manière que l'isoforme musculaire. Les conséquences de la mutation sur l'isoforme Dp140 sont difficiles à prédire car la mutation se situe dans la région promotrice mais on peut suspecter que l'isoforme sera impactée par la mutation. Les isoformes les plus courtes (Dp116 et Dp71) ne sont pas affectées par la mutation.

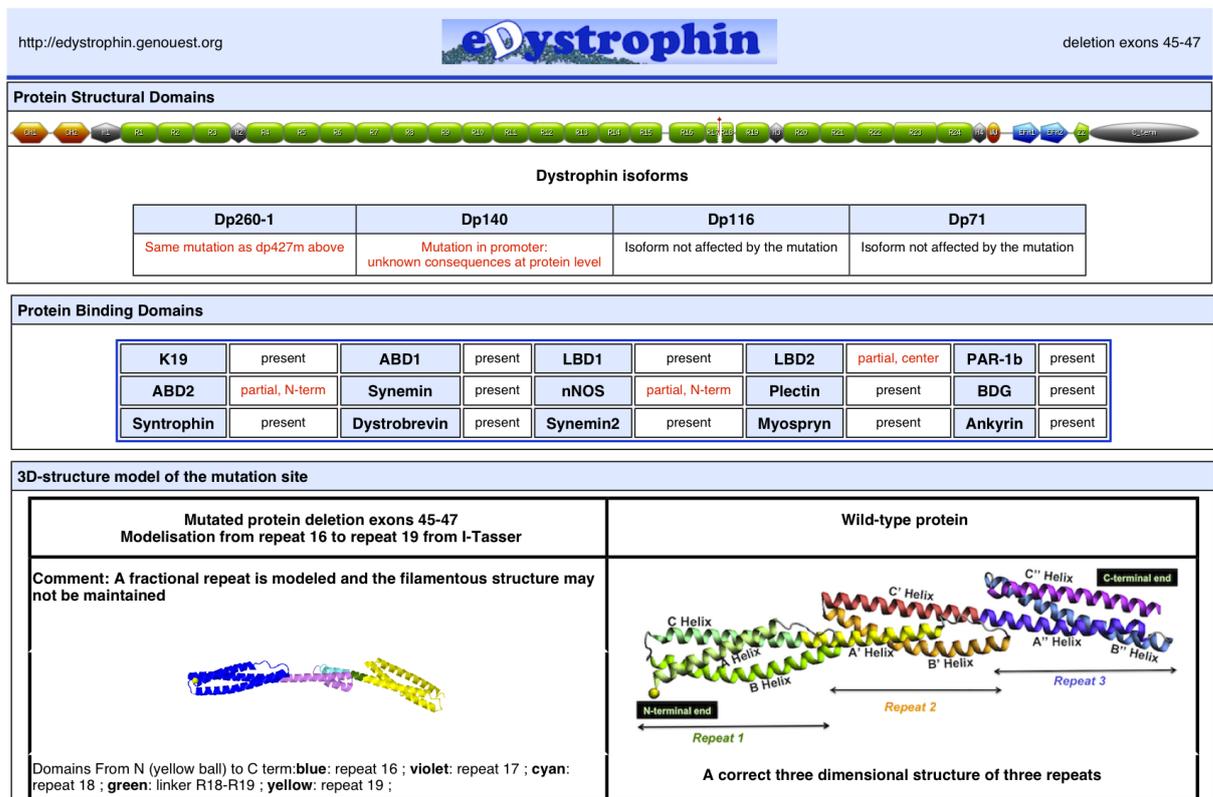


FIGURE 3.14 – Deuxième page de la fiche-résumé de la délétion d'exons 45-47 téléchargeable sur l'interface d'edystrophin.

L'onglet *Protein binding domains* indique les domaines de liaison de la dystrophine modifiés par la mutation (figure 3.14). Les domaines de liaison à l'actine (ABD2), aux lipides (LBD2) et à la nNOS sont partiellement tronqués par la délétion. Les autres domaines de liaison sont entièrement présents.

L'onglet *3D-structure model of the mutation site* propose la structure par homologie de la région entourant la délétion (cf. Chapitre 2). Ce modèle comprend les répétitions 16, la répétition constituée de fragments de 17 et de 18 et la répétition 19. Un commentaire nous indique que la structure 3D de la jonction est composée de petites hélices ne reconstituant

pas une répétition hybride (répétition avec une structure proche de celle des répétitions de la dystrophine native), la répétition est donc fracturée (cf. Chapitre 2).

Le graphe de Ramachandran indique que les résidus du modèle sont à 90,2% dans des régions favorables à 1,6% dans des régions défavorables. Les acides aminés dans des régions défavorables sont principalement localisés dans les boucles reliant les hélices.

### La délétion des exons 45 à 55

Le tableau résumé s'affichant sur la page de la délétion d'exons 45-55 indique que la séquence d'ADNc est composée de 9276 nucléotides et que la séquence protéique est composée de 3092 acides aminés. La protéine a un poids moléculaire de 358 kDa. La protéine produite est donc plus petite que la protéine tronquée précédente. Cette mutation a été observée chez des patients présentant différents phénotypes mais est majoritairement associée à un phénotype BMD. Les patients ont été collectés dans trois études précédemment publiées et auprès du LGBM comme indiqué sous le tableau.

En cliquant sur *Get pdf file*, il est possible de télécharger l'ensemble des données de la page dans une fiche au format pdf (figure 3.15 et 3.16)

L'onglet *Clinical data* nous indique que la délétion a majoritairement été identifiée chez des patients BMD (~89%) mais aussi chez trois patients asymptomatiques et un patient présentant un cardiomyopathie isolée (figure 3.15). On remarque que, lorsque les données sont disponibles, la dystrophine détectée dans les muscles des patients est de taille réduite et présente en moindre quantité par rapport à un muscle sain. En grande majorité, les patients ne présentent ni cardiomyopathie, ni retard mental lors de leur dernière visite.

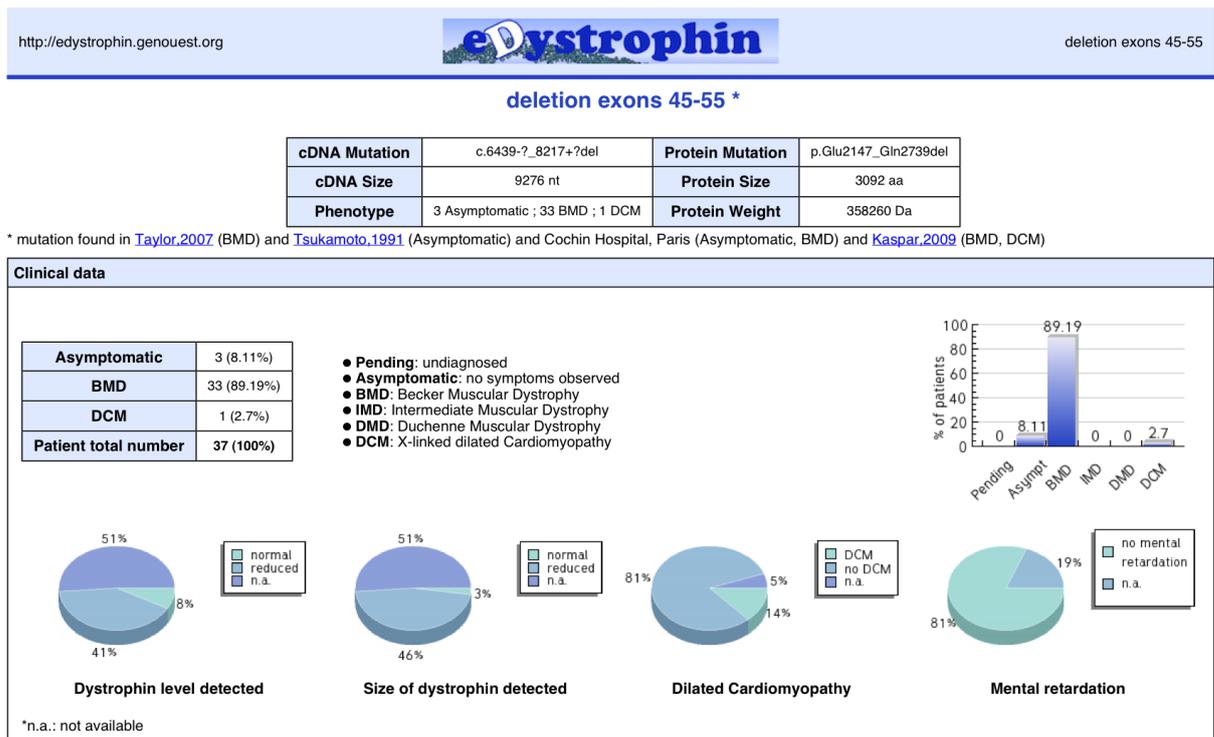


FIGURE 3.15 – Première page de la fiche-résumé de la délétion d'exons 45-55 téléchargeable sur l'interface d'eDystrophin.

L'onglet *Protein Structural Domains* présente le schéma de la dystrophine tronquée (figure 3.16). Les répétitions 18 à 21 sont absentes ainsi que la charnière 3, la répétition 17 est tronquée en position C-terminale et la répétition 22 est tronquée en position N-terminale .

La mutation touche l'isoforme Dp260 de la même manière que l'isoforme musculaire. La mutation touche les régions promotrices des isoformes Dp140 et Dp116. L'isoforme la plus courte (Dp71) n'est pas affectée par la mutation.

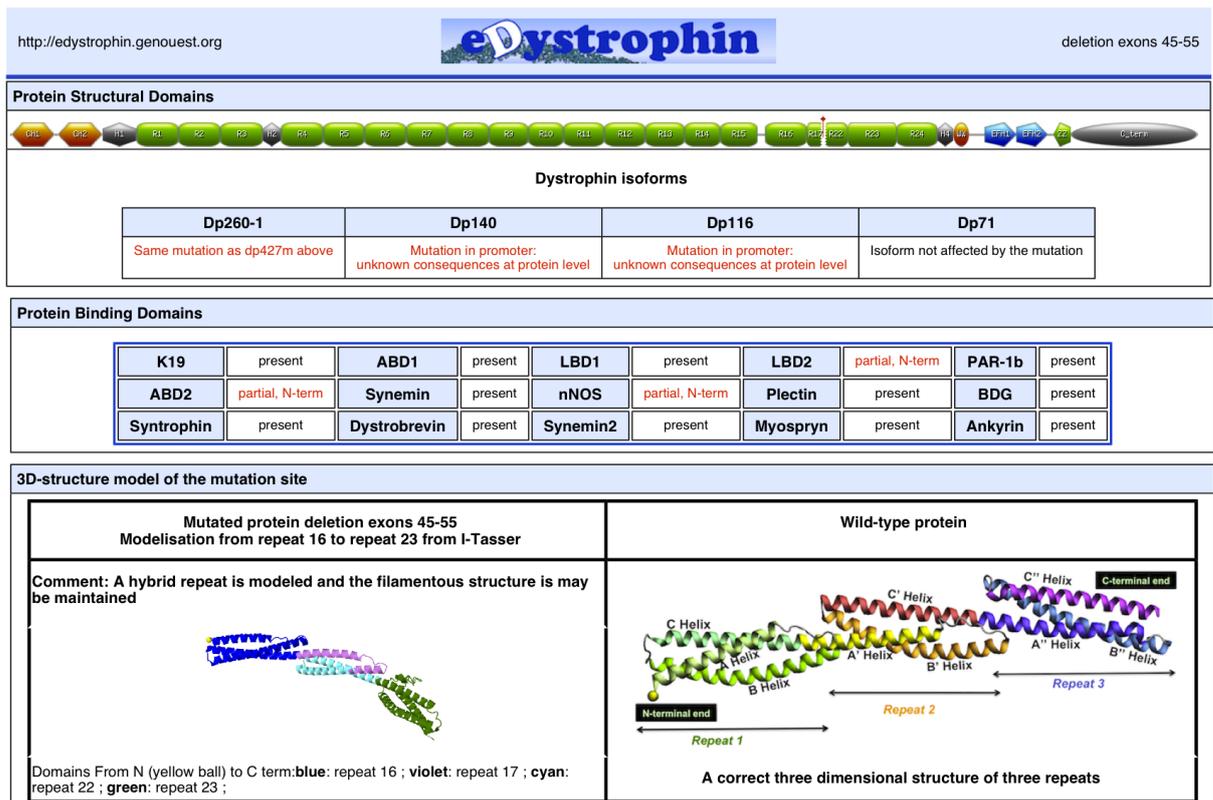


FIGURE 3.16 – Deuxième page de la fiche-résumé de la délétion d'exons 45-55 téléchargeable sur l'interface d'EdyStrophin.

Grâce à l'onglet *Protein binding domains* nous remarquons que les domaines de liaison affectés sont les mêmes que dans la délétion précédente (figure 3.16). En effet, les domaines de liaison à l'actine (ABD2), aux lipides (LBD2) et à la nNOS sont partiellement tronqués par la délétion.

L'onglet *3D-structure model of the mutation site* propose un modèle 3D comprenant les répétitions 16, la répétition constituée de fragments 17 et de 22 et la répétition 23. Un commentaire nous indique que la structure 3D de la jonction est composée d'une répétition hybride avec une structure proche de celle des répétitions de la dystrophine native, et que la structure filamenteuse de la protéine est donc potentiellement maintenue.

Le graphe de Ramachandran indique que les résidus du modèle sont à 90,5% dans des régions favorables à 1,2% dans des régions défavorables. Comme pour la délétion des exons 45-47, les acides aminés dans les régions défavorables sont principalement localisés dans les boucles reliant les hélices.

### 3.2.4 Mise en évidence de l'alignement des répétitions sur l'alignement des exons

Nous avons remarqué que l'alignement des séquences d'acides aminés des répétitions tel que Winder l'a réalisé en 1995 [Winder 1995a] peut être superposé avec les bornes des exons. En effet, chaque hélice B est encodée par deux exons et la limite entre ces deux exons s'aligne parfaitement sur toutes les répétitions à l'exception de la répétition 14 (figure 3.17). De plus, cette limite se superpose parfaitement avec la troisième heptade de l'hélice B et se situe entre les positions b et c de l'heptade.

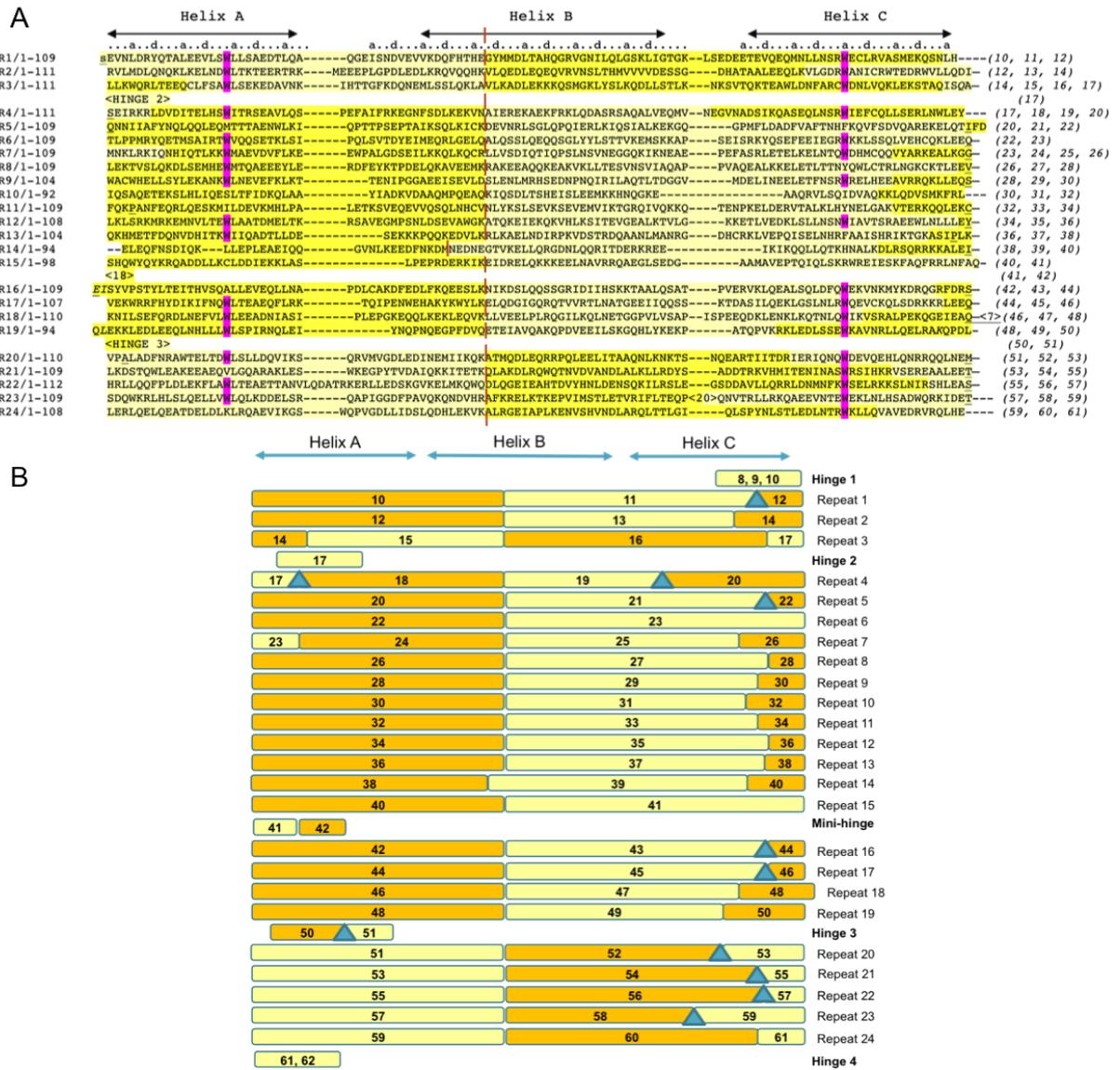


FIGURE 3.17 – A. Alignement des répétitions et des charnières 2 et 3 du domaine central. Chaque ligne représente une répétition dont le nom puis la taille sont indiqués à gauche. Les positions des acides aminés hydrophobes a et d dans les heptades sont indiquées en haut. Chaque exon est colorié alternativement en jaune clair et jaune foncé. Le numéro des exons encodant chaque répétition est indiqué à droite. Le trait pointillé rouge indique la limite entre les deux exons qui encodent chaque hélice B. B. Représentation schématique de l'alignement précédent.

Cet alignement peut être utilisé pour prédire si, à la suite d'une délétion d'exon, il y aura potentiellement la reconstitution d'une répétition hybride ou d'une répétition fracturée. En effet, si une délétion ôte un multiple de deux exons débutant au centre de l'hélice B, le motif en heptade caractéristique des faisceaux d'hélices est conservé et une répétition hybride est reconstituée. Cette répétition hybride est alors constituée du début de l'hélice B encodée par l'exon précédent la délétion et par la fin de l'hélice B encodée par l'exon suivant la délétion. C'est par exemple, le cas de la délétion des exons 45-55 (figure 3.18 A). Dans tous les cas contraires, la délétion laisse des hélices trop courtes ou trop longues qui ne peuvent pas reconstituer une répétition et la répétition formée est alors trop courte pour être une répétition hybride ; on parle alors de répétition fracturée, comme dans le cas de la délétion des exons 45-47 (figure 3.18 B).

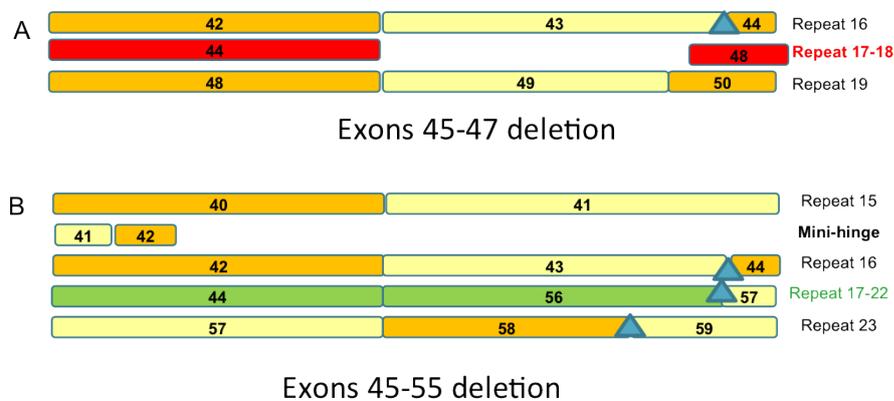


FIGURE 3.18 – A. Alignement des exons entourant la délétion 45-47. La répétition 17//18 créée par la délétion est trop courte pour former une répétition hybride. B. Alignement des exons entourant la délétion 45-55. La répétition 17//22 créée par la délétion est une répétition hybride.

# Chapitre 4

## Discussion générale

Le rôle précis de la dystrophine est encore méconnu même s'il est évident qu'elle joue un rôle de protection des cellules musculaires lors des cycles de contraction-relaxation et que sa structure filamenteuse et ses nombreux domaines de liaisons à des partenaires lui permettent d'assurer le lien entre la membrane plasmique et les protéines du cytosquelette [Deconinck 2007, Le Rumeur 2010]. L'étude expérimentale de la dystrophine est très difficile, notamment parce qu'elle est très grande, filamenteuse et présente une structure répétitive. Cependant, au sein de l'équipe, différentes études *in vitro* de biochimie et biophysique ont été réalisées afin de caractériser la structure et les propriétés d'interactions de la dystrophine avec certains de ses partenaires notamment les lipides et l'actine. En complément de ces approches expérimentales, nous avons choisi de développer des approches *in silico* pour avoir une vue d'ensemble de la dystrophine et l'étudier à un niveau moléculaire.

La modélisation par homologie nous a permis d'étudier la structure des répétitions natives de la dystrophine pour pallier l'absence de structures expérimentales. Les modèles de répétitions de dystrophine montrent une structure en faisceaux de trois hélices et sont donc en accord avec les études de dichroïsme circulaire publiées sur la dystrophine [Le Rumeur 2003, Legardinier 2009a, Mirza 2010] qui mettent en évidence une structuration majoritairement en hélices et en faisceau d'hélices du domaine central de la dystrophine. De plus, la très récente publication de la structure de la répétition 1 de la dystrophine confirme la pertinence de nos modèles par homologie [Muthu 2012].

Le domaine central a longtemps été décrit comme un "bâton rigide" composé de répétitions identiques et interchangeable assurant le lien entre le domaine ABD en position N-terminal et le domaine  $\beta$ DG en position C-terminal [Davies 2006]. L'étude des modèles des répétitions contredit cette hypothèse et met en évidence une grande hétérogénéité entre les répétitions avec des potentiels de surface électrostatiques et hydrophobes différents pouvant expliquer les différentes interactions connues dans la dystrophine. Ces répétitions successives sont reliées entre elles par des linkers présentant également des caractéristiques structurales hétérogènes. En effet, contrairement à ceux de la spectrine, les linkers de la

dystrophine peuvent être non hélicaux. De plus, certains linkers sont flexibles et d'autres forment des coudes rigides, comme par exemple le linker R10-11. Ce type d'étude a déjà été réalisé sur la spectrine [Mirijanian 2007] et l' $\alpha$ -actinine [Zaman 2004] et montre une flexibilité accrue des linkers mais ne met pas en évidence la formation de coudes tels que nous l'observons pour certains linkers de la dystrophine. Ainsi, le domaine central de la dystrophine est flexible et constitué de répétitions hétérogènes.

Des centaines de mutations sont répertoriées sur le gène *DMD* codant pour la dystrophine. Ces mutations sont majoritairement des délétions (ou duplications) de plusieurs exons entiers. Il est admis, à quelques exceptions près, que les mutations hors phase conduisent à l'absence de dystrophine dans les muscles et à un phénotype DMD. Cependant, les mutations en phase produisent différentes dystrophines mutées conduisant à des phénotypes variables classés en phénotype BMD. Actuellement, aucune corrélation génotype-phénotype claire n'existe. Par conséquent, nous avons décidé d'étudier des dystrophines produites suite à des mutations en phase afin d'enrichir cette corrélation. Ces délétions ont majoritairement lieu dans le domaine central où elles tronquent entièrement ou partiellement des répétitions. Ainsi, de nouvelles répétitions constituées de deux répétitions tronquées apparaissent à la jonction des délétions avec potentiellement de nouvelles propriétés. Ces délétions d'exons entiers sont une caractéristique quasi exclusive du gène *DMD* et leur étude nécessite des approches méthodologiques spécifiques permettant de comparer des protéines de tailles variables dont il manque des domaines entiers et/ou partiels. Nous avons donc été amenés à définir de nouvelles méthodes pour étudier notamment les nouvelles répétitions formées.

Il existe un grand nombre de délétions en phase répertoriées dans les bases de données existantes mais elles sont diluées parmi les mutations hors phase et sont donc difficilement exploitables. Nous avons donc créé eDystrophin pour étudier ces altérations et aider à comprendre l'impact de chaque mutation en phase sur le maintien de la structure filamenteuse de la dystrophine et des interactions avec ses partenaires.

L'atout majeur d'eDystrophin est de permettre l'étude des dystrophines mutées (généralement tronquées) résultant d'une mutation en phase du gène *DMD*. En effet, la base de données présente une vision d'ensemble des conséquences d'une mutation sur les domaines structuraux et les domaines de liaison aux partenaires des dystrophines mutées. Elle rassemble également toutes les données disponibles sur la dystrophine native et est enrichie de données cliniques. A ce titre, elle est comparable avec les bases de données spécifiques du gène existantes (UMD-DMD [Tuffery-Giraud 2009] et Leiden Muscular Dystrophy [White 2006]).

L'analyse des domaines de liaison altérés nous a permis de confirmer que les mutations affectant les domaines ABD1 et riche en cystéines sont plus souvent associées à des

phénotypes graves (DMD). Ces observations sont en accord avec des études précédemment publiées [Beroud 2007, Tuffery-Giraud 2009]. Cependant, nous avons également mis en évidence plusieurs mutations du domaine central associées à des phénotypes DMD. Plusieurs hypothèses peuvent expliquer la sévérité des phénotypes comme par exemple l'instabilité de l'ARNm mais il n'existe pas de données de ce type. La mutation peut également avoir été mal délimitée et être en réalité une mutation hors phase comme c'est le cas pour plusieurs patients de l'étude [Taylor 2007]. Des phénomènes épigénétiques encore mal connus peuvent également expliquer la sévérité des phénotypes. Le manque de données cliniques fiables et détaillées rend également difficile la corrélation entre le génotype et le phénotype. Cependant, on ne peut pas exclure que la mutation elle-même ait un impact très délétère sur la fonction de la protéine produite et entraîner un phénotype grave. C'est la raison pour laquelle j'ai mené l'étude *in silico* de ces protéines.

Les modèles structuraux des régions entourant les délétions ont été construits en suivant le même protocole que celui que nous avons mis en place pour les répétitions en tandem de la dystrophine native et sont disponibles pour l'ensemble des délétions du domaine central dans eDystrophin.

Ces modèles structuraux ont deux types de structures tridimensionnelles remarquables en accord avec les hypothèses précédemment émises [Menhart 2006, Kaspar 2009, Yokota 2007]. D'une part, on observe des répétitions hybrides qui conservent la succession des motifs en heptade des hélices et peuvent se structurer en faisceau de trois hélices comme les répétitions natives. C'est le cas lorsqu'une délétion débute dans le milieu de l'hélice B et qu'elle ôte un multiple de deux exons ce qui aboutit à la jonction de deux moitiés d'hélices B pour reformer une répétition similaire à une répétition native. Ces répétitions hybrides présentent une structure similaire aux répétitions de la dystrophine native et permettraient de conserver la structure filamenteuse de la dystrophine [Parry 2008, Lupas 1996] indiquant une meilleure fonctionnalité de la protéine.

D'autre part, on observe des répétitions fracturées, qui ne conservent pas les motifs en heptade et ne reconstituent donc pas de faisceau de trois hélices similaire aux répétitions natives. On distingue plusieurs types de répétitions fracturées : les répétitions fracturées impliquant une modification structurale localisée à la jonction de la délétion et conservant, à priori, la forme filamenteuse de la dystrophine comme c'est le cas dans  $\Delta 45-47$  et celles modifiant sa forme filamenteuse comme le faisceau de deux hélices de  $\Delta 45-49$ . Ces répétitions pourraient s'accompagner d'une perte de fonction partielle de la dystrophine et conduire à un phénotype BMD sévère.

Nous avons donc choisi d'étudier plus précisément ces deux types de répétitions nouvellement formées pour répondre aux deux questions suivantes :

*Les répétitions hybrides ont-elles les mêmes propriétés que les répétitions natives ?*

*Les répétitions hybrides permettent-elle une meilleure fonctionnalité de la protéine que les répétitions fracturées ?*

La dynamique moléculaire a été utilisée pour étudier le comportement et les propriétés dynamiques de quatre protéines tronquées afin de les comparer à la protéine native. Ces quatre protéines sont tronquées suite à des mutations situées dans le point chaud de mutation. Deux de ces protéines sont caractérisées par la présence d'une répétition hybride ( $\Delta 45-48$  et  $\Delta 45-51$ ) et les deux autres protéines sont caractérisées par la présence d'une répétition fracturée ( $\Delta 45-47$  et  $\Delta 45-49$ ).

Cependant, la comparaison des trajectoires de DM entre une protéine pleine longueur et des protéines tronquées n'est pas triviale. En effet, aucune étude de ce type n'est publiée dans la littérature, seules des études de DM d'une protéine native comparée à une protéine avec une mutation ponctuelle ont été publiées. Nous avons donc mis en place un protocole d'étude spécifique à partir des outils existants afin de comparer les caractéristiques des régions présentes de chaque protéine tronquée avec la protéine native comme par exemple pour l'étude de la conservation des hélices (section 2.15). De plus, nous avons comparé spécifiquement les répétitions hybrides aux répétitions natives comme par exemple dans l'étude des propriétés de surface électrostatiques et hydrophobes (section 2.3.2.2).

Les deux répétitions hybrides des protéines  $\Delta 45-48$  et  $\Delta 45-51$  présentent des profils de fluctuation différents entre eux et différents des répétitions natives (section 2.13). En effet, la répétition hybride R17//19 de  $\Delta 45-48$  est plus flexible que les répétitions natives correspondantes et présente de nouveaux points de flexibilité alors que la répétition hybride R17//20 de  $\Delta 45-51$  est plus rigide que les répétitions correspondantes de la protéine R16-21. La fluctuation élevée de la répétition R17//19 de  $\Delta 45-48$  est associée à un nouveau mouvement de dilatation du faisceau d'hélice (section 2.6) qui n'apparaît ni dans la protéine native, ni dans la protéine  $\Delta 45-51$ .

Les contacts entre les faisceaux d'hélices et les linkers situés de part et d'autre des répétitions natives stabilisent les faisceaux d'hélices. Or, ces interactions sont dépendantes de la chaîne latérale des acides aminés constituant le linker et des régions en contact avec celui-ci. Par conséquent, en fonction de la séquence en acides aminés du linker et des répétitions situées de part et d'autre, le nombre et le type d'interactions diffèrent et ne sont donc pas systématiquement transposables d'un linker à l'autre. On peut donc supposer que les contacts d'un linker avec une répétition hybride diffèrent de ceux avec une répétition native.

De plus, dans les répétitions natives, les contacts entre le linker et la répétition en N-terminal sont plus nombreux que ceux avec la répétition en C-terminal du linker (section 2.3.2.3). En effet, une partie de la longue hélice B de la répétition en N-terminal est en contact avec le linker formant ainsi un mini faisceau de deux hélices par le biais des interactions hydrophobes des motifs en heptade (figure 4.1A). Les contacts avec la répétition en C-terminal sont principalement localisés dans la boucle BC qui est mobile. Par conséquent,

lorsqu'une répétition est suivie en C-terminal par une charnière comme c'est le cas pour R19, sa flexibilité est accrue comme nous le voyons dans le graphe de RMSF de R16-21 (section 2.13) et dans ses structures représentatives (section 2.16). La répétition R19 est de plus précédée en position N-terminal d'un linker particulier incluant une mini-charnière. Cette mini-charnière très fluctuante participe également à la fluctuation importante de R19.

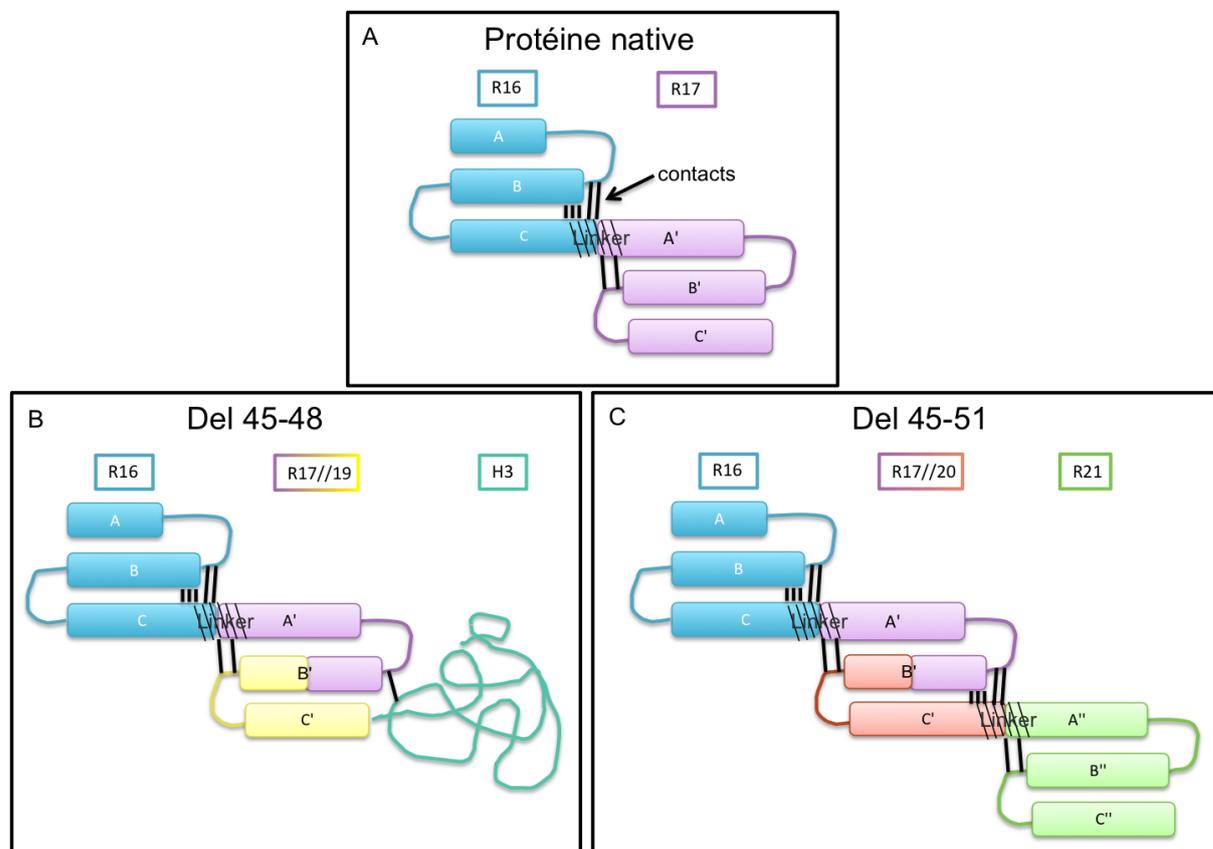


FIGURE 4.1 – A. Représentation schématique des contacts du linker R16-17 avec les répétitions 16 et 17 dans la protéine native R16-21. Contacts de la répétition hybride R17//19 de  $\Delta 45-48$  avec le linker R16-17 et avec la charnière H3. C. Contacts de la répétition hybride R17//20 de  $\Delta 45-51$  avec les linkers R16-17 et R20-21.

La répétition hybride R17//19 de  $\Delta 45-48$  est entourée du linker R16-17 d'un côté et de la charnière H3 de l'autre (figure 4.1B). La charnière offre peu de surface de contacts avec la répétition hybride (section 2.3.2.4) et n'aide donc pas à maintenir le faisceau d'hélices. Seul, le linker R16-17 aide à la cohésion du faisceau d'hélices. Cependant, ce linker situé en N-terminal de la répétition hybride stabilise peu le faisceau d'hélices bien que le nombre de contacts soit du même ordre de grandeur que pour la répétition R17 native et la répétition hybride R17//20. Par conséquent, la répétition hybride R17//19 présente une fluctuation accrue proche de celle de la répétition native R19.

La répétition hybride R17//20 de  $\Delta 45-51$  est entourée de deux linkers favorisant les contacts maintenant le faisceau d'hélice (figure 4.1C). Les boucles AB et BC en contacts avec le linker R16-17 se situent du même côté et sont assez proches pour créer des interactions. De plus, les contacts entre R17//20 et le linker R20-21 sont plus nombreux que ceux de

la répétition native avec ce linker. La délétion hybride R17//20 favorise plus de contacts avec les deux linkers environnants que la répétition native R16-21. Ces contacts accrus et l'absence de la charnière H3 peuvent expliquer la rigidité de la protéine observée dans le graphe de RMSF (section 2.13). Toutefois, à l'heure actuelle, il est impossible de dire si la hausse ou la baisse de flexibilité est délétère ou non pour la fonctionnalité de la dystrophine tronquée.

Les répétitions hybrides étudiées ici présentent donc deux comportements différents en fonction de leur séquence et des linkers ou charnières environnantes. Les répétitions hybrides semblent donc ne pas avoir le même comportement dynamique entre elles ni avoir le même comportement que les répétitions natives.

Les contacts des répétitions fracturées sont difficilement comparables à ceux des répétitions natives car les répétitions fracturées ont des structures tertiaires modifiées. Cependant, on remarque que la répétition fracturée R17//18 de  $\Delta 45-47$  présente peu de contacts avec les linkers environnants et que la répétition fracturée R17//19 de  $\Delta 45-49$  est en contact avec la charnière H3.

La dystrophine native R16-21 est flexible comme nous l'avons vu dans l'étude du RMSF (section 2.13) et des mouvements principaux (section 2.6). La flexibilité de cette protéine est majoritairement due à la mini-charnière rallongeant le linker entre R18 et R19. Or, ce linker est absent des protéines tronquées ou fait partie d'une répétition fracturée et déstructurée dans la protéine  $\Delta 45-47$ . Cependant, la flexibilité de la charnière H3 est accrue dans les trois protéines tronquées et il semblerait qu'elle prenne le relais de la mini-charnière entre R18 et R19.

La charnière H3, majoritairement composée de boucles, est très flexible et peut donc se réarranger en fonction des forces exercées sur elle. Comme nous l'avons vu dans l'étude des mouvements principaux (section 2.6), les mouvements décrits dans les protéines tronquées diffèrent entre eux et diffèrent de ceux observés dans la protéine native R16-21. Or ces mouvements se propagent le long du filament de la dystrophine. Ainsi, on peut supposer qu'un mouvement apparu au niveau d'une répétition tronquée puisse se répercuter dans une région distante comme ici dans la charnière H3 en imposant une conformation différente de la native.

Ce type de phénomène pourrait également expliquer la modification des contacts du linker R20-21 avec les répétitions natives R20 et R21 de R16-21,  $\Delta 45-47$ ,  $\Delta 45-48$  et  $\Delta 45-49$ . Cependant, ce phénomène est à nuancer car il peut être consécutif à la déstructuration plus ou moins importante de R21 dans nos protéines et il n'est pas observé pour le linker R16-17.

La dystrophine est une protéine majoritairement hydrophile comme la plupart des protéines cytoplasmiques. Elle est caractérisée par la succession de motifs en heptades où

des résidus hydrophobes sont présents en position a et d et par de petits sillons hydrophobes globalement observés aussi dans les protéines tronquées. Sans grande surprise, les protéines avec une répétition hybride ( $\Delta 45-48$  et  $\Delta 45-51$ ) conservant les motifs en heptade présentent le même profil hydrophobe que la protéine native et conservent des sillons hydrophobes. Par contre, les répétitions fracturées ( $\Delta 45-47$  et  $\Delta 45-49$ ) présentent un potentiel hydrophobe plus important, principalement la protéine  $\Delta 45-47$  qui apparaît largement déstructurée dans l'étude de la conservation des hélices (section 2.15). De plus, les sillons hydrophobes sont modifiés ou partiellement absents des répétitions fracturées. Cependant, le potentiel de surface hydrophobe global de ces protéines  $\Delta 45-47$  et  $\Delta 45-49$  présente le même profil que la protéine native R16-21. En effet, le potentiel hydrophobe de la dystrophine, déterminé par la succession des motifs en heptade, n'est pas modifié dans les répétitions natives de ces protéines. Il apparaît donc que le potentiel hydrophobe des protéines tronquées et les sillons hydrophobes sont altérés uniquement dans la région localisée aux répétitions fracturées. Cette modification est d'autant plus importante que la répétition fracturée est déstructurée.

Les propriétés électrostatiques de la dystrophine sont hétérogènes comme nous l'avons vu dans l'étude des répétitions natives en tandem. Les protéines avec une répétition hybride possèdent les mêmes propriétés que la protéine native alors que les propriétés électrostatiques des protéines avec une répétition fracturée sont modifiées mais seulement au niveau de la répétition fracturée et de la charnière H3.

La charnière H3 globalement neutre dans la protéine native, se charge positivement notamment dans les protéines  $\Delta 45-49$  possédant une répétition fracturée et  $\Delta 45-48$  possédant une répétition hybride. La charnière étant très mobile et adoptant des conformations différentes entre les protéines, il n'est pas étonnant de la voir présenter des résidus chargés positivement à la surface dans certaines protéines notamment dans la protéine  $\Delta 45-49$  dont la conformation est très éloignée de la structure filamenteuse de la protéine native. Cette charnière n'a pas de propriété de liaison connue mais on peut émettre l'hypothèse qu'elle peut, si nécessaire, facilement se réarranger pour aider la liaison à des partenaires.

Le potentiel électrostatique des répétitions fracturées diffère totalement de la répétition R17 qui présente un alternance de petits patches positifs et négatifs. En effet, la répétition fracturée R17//18 de la protéine  $\Delta 45-47$  est très largement négative après la jonction et la répétition R17//19 de la protéine  $\Delta 45-49$  est positive après la jonction.

Le potentiel électrostatique des répétitions hybrides présente des patches positifs et négatifs après la jonction similaire à R17. Cependant, ces patches sont de plus forte intensité et sont inversés par rapport à la protéine native. En effet, les profils de R18 ( $\Delta 45-48$ ) et de R20 ( $\Delta 45-51$ ) présentent une alternance de patches négatif - positif - négatif alors que la région correspondante dans R17 présente un alternance de patches positif - négatif - positif. Ces modifications peuvent certainement altérer les propriétés d'association de la dystrophine avec ses partenaires.

*Les dystrophines tronquées peuvent-elles toujours lier leurs partenaires ?*

Si la troncature fait totalement disparaître un domaine de liaison, il est raisonnable de supposer que la dystrophine tronquée ne peut plus lier ce partenaire à moins qu'un autre domaine de liaison existe pour ce partenaire dans une autre région de la protéine, comme par exemple les deux domaines de liaison à l'actine. Ce n'est pas le cas des quatre protéines tronquées étudiées ici où les domaines d'interaction sont partiellement tronqués. Les domaines de liaison connus sur le fragment de dystrophine R16-21 étudié ici sont au nombre de trois : la fin du domaine de liaison à l'actine ABD2 dans les répétitions 16 et 17 [Rybakova 2006] ; le domaine de liaison à la nNOS également dans les répétitions 16 et 17 [Lai 2009] et une partie du domaine de liaison aux lipides LBD2 dans les répétitions 16 à 19 [Legardinier 2009b]. La liaison de la dystrophine avec l'actine fait intervenir des interactions électrostatiques tout comme la liaison aux lipides. Le mode de liaison de la dystrophine à la nNOS est pour le moment inconnu mais on peut supposer qu'il fait intervenir les propriétés électrostatiques et/ou hydrophobes de la dystrophine. De plus, une collaboration en cours dans l'équipe avec Lai & Duan (Missouri, Columbia, USA) met en évidence que la répétition 17 est essentielle pour la liaison à la nNOS.

Les protéines avec une répétition hybride sont celles qui peuvent le mieux conserver les propriétés de liaison de la dystrophine mais avec des nuances. Le sillon hydrophobe potentiellement nécessaire aux interactions de la dystrophine avec ses partenaires est conservé dans les protéines avec une répétition hybride.

Par contre, les propriétés électrostatiques sont modifiées. En effet, le tandem de répétitions R18-19 largement électronégatif disparaît, ce qui peut probablement altérer la liaison aux lipides. Par contre, les répétitions hybrides de  $\Delta 45-48$  et  $\Delta 45-51$  présentent un potentiel électrostatique hétérogène proche de celui de la répétition R17 native. On peut donc émettre l'hypothèse que le potentiel électrostatique des répétitions hybrides peut suffire à lier l'actine, les lipides et la nNOS.

Les propriétés de liaison des protéines avec une répétition fracturée semblent plus compromises. Dans un premier temps, la structure tertiaire peut être fortement modifiée, c'est le cas par exemple de la protéine  $\Delta 45-49$  qui adopte une conformation en "L" ou en "T". Ce changement constitue vraisemblablement un obstacle à la liaison de la dystrophine tronquée avec ses partenaires notamment la nNOS, qui plus est, une très grande protéine. De plus, les potentiels hydrophobes et électrostatiques des répétitions tronquées sont modifiés par rapport à la protéine native. Les protéines  $\Delta 45-47$  et  $\Delta 45-49$  doivent avoir une capacité modifiée à lier les lipides, l'actine et la nNOS.

Nos résultats permettent de prédire des modifications entre la protéine native, les pro-

téines caractérisées par une répétition hybride ( $\Delta 45-48$  et  $\Delta 45-51$ ) et celles caractérisées par une répétition fracturée ( $\Delta 45-47$  et  $\Delta 45-49$ ). A partir de nos analyses, il est possible de réaliser un classement des quatre protéines tronquées en les comparant à la dystrophine native (tableau 4.1). Ainsi, la  $\Delta 45-51$ , très courte, est celle qui semble le mieux conserver les caractéristiques de la dystrophine native, puis viennent les  $\Delta 45-48$  et  $\Delta 45-47$  et en enfin la  $\Delta 45-49$  qui présente une structure très modifiée.

	$\Delta 45-51$ (hybride)	$\Delta 45-48$ (hybride)	$\Delta 45-47$ (fracturée)	$\Delta 45-49$ (fracturée)
Structure 3D	filamenteuse sans H3	filamenteuse	filamenteuse	non filamenteuse
Flexibilité	diminuée	accrue	modifiée	modifiée
Hydrophobicité	identique	identique	localement modifiée	localement modifiée
Electrostatique	similaire	similaire	modifiée	modifiée
Hypothèse sur la fonction	conservée	conservée	altérée	fortement altérée

Tableau 4.1 – *Comparaison des propriétés des quatre protéines tronquées avec les propriétés de la dystrophine native.*

Plus généralement, nous pouvons formuler les conclusions suivantes

Pour ce qui concerne les répétitions hybrides :

- Elles conservent la succession des motifs en heptade et présentent une structuration en faisceaux de trois hélices comme les répétitions natives mais ces faisceaux ne sont pas identiques et interchangeables.
- Elles sont globalement stables mais cette stabilité dépend des répétitions environnantes, plus particulièrement de la présence d'un linker en C-terminal de la répétition hybride. En effet, la répétition hybride sera moins stable si une charnière se situe de ce côté.
- Elles présentent un potentiel hydrophobe similaire à celui de la protéine native. Ces protéines tronquées conservent également le sillon hydrophobe.
- Les conséquences de la présence d'une répétition hybride sur le potentiel électrostatique de la protéine sont plus difficiles à généraliser. En effet, elles dépendent en premier lieu des propriétés électrostatiques des deux moitiés de répétitions réunies pour former la répétition hybride. Les deux répétitions hybrides étudiées ici sont tronquées dans la répétition R17 et présentent un profil électrostatique proche de la répétition native R17 potentiellement suffisant pour assurer les propriétés de liaison de R17.
- L'absence de certains domaines structuraux et/ou la présence d'une répétition hybride ont une incidence sur la flexibilité globale de la protéine par rapport à la

protéine native.

Pour ce qui concerne les répétitions fracturées :

- Elles présentent des structures tridimensionnelles plus ou moins modifiées en fonction de la délétion ce qui peut conduire à des conformations très modifiées. Les propriétés dynamiques des protéines avec ce type de répétition dépendent notamment de la structuration globale de la répétition fracturée. Par conséquent, il est difficile de proposer des caractéristiques générales pour ces protéines.
- Plus la structure tertiaire de la protéine s'éloigne de la protéine native, plus la protéine avec une répétition fracturée présente des caractéristiques différentes de la protéine native.
- La flexibilité de la protéine peut être modifiée par une structuration non filamenteuse de la protéine ou par une déstructuration localisée de la répétition fracturée.
- Les propriétés hydrophobes et électrostatiques de ces protéines peuvent être complètement modifiées.
- Les propriétés de liaison de la dystrophine avec des partenaires peuvent être compromises par la nouvelle conformation de la protéine.

Par conséquent, on peut raisonnablement émettre l'hypothèse qu'une répétition hybride est préférable à une répétition fracturée pour assurer une meilleure fonctionnalité de la dystrophine. Cependant, la présence d'une répétition hybride ne garantit pas le maintien des interactions de la dystrophine avec ses partenaires. En effet, tout comme les répétitions natives ont des propriétés hétérogènes, les répétitions hybrides ont aussi des propriétés hétérogènes et ne présentent donc pas les mêmes propriétés que les répétitions natives. Il est donc important de connaître précisément les modes d'interactions de la dystrophine avec ses partenaires afin d'en déduire les propriétés de liaison des dystrophines tronquées.

L'hypothèse émise précédemment supposant que les répétitions hybrides pourraient lier la nNOS est corroborée par une récente étude clinique dont trois patients avec la mutation  $\Delta 45-51$  [Anthony 2011]. Un taux d'expression de 40 à 50% (par rapport au contrôle) de la nNOS dans les cellules musculaires de ces patients a été observé. Ces patients sont atteints d'une forme asymptomatique ou légère. Malheureusement, cette étude ne nous informe pas de la localisation de la nNOS chez ces patients. Par contre, l'étude [Gentil 2012] sur la délétion  $\Delta 45-55$  corrèle la localisation de la nNOS à la membrane à une sévérité phénotypique moins importante et la localisation cytosolique de la nNOS à une sévérité phénotypique plus importante chez des patients.

Très peu d'études ont été réalisées sur la relation entre la structure de la dystrophine mutée et le phénotype observé mais deux études montrent que la présence de la charnière H3 est plus délétère que son absence [Carsana 2005, Kaspar 2009]. Cependant, ce type de

résultat est également à mettre en relation avec la présence d'une répétition hybride ou non. L'étude de [Kaspar 2009] révèle que dans le groupe de patients où la charnière H3 est présente, la présence d'une répétition hybride retarde l'apparition de la cardiomyopathie mais les patients ne présentant pas de cardiomyopathie ne sont pas pris en compte. De plus, cette mise en relation n'a pas été réalisée pour le groupe n'ayant pas la charnière H3, ni avec l'ensemble de la cohorte. L'étude de Carsana [Carsana 2005] n'étudie pas la présence d'une répétition hybride. Or, les patients du groupe "sans H3" ont tous des répétitions hybrides et les patients du groupe "avec H3" ont soit une répétition hybride, soit une répétition fracturée. On peut donc penser que la corrélation entre la présence de H3 et une sévérité plus importante est à revoir en prenant en compte la présence ou l'absence d'une répétition hybride.

Nous avons donc repris les données cliniques détaillées de [Kaspar 2009] afin de réaliser une première corrélation entre les quatre protéines tronquées étudiées par DM et le phénotype des patients (tableau 4.2). En utilisant l'âge d'apparition de la DCM comme témoin de sévérité de la BMD pour les quatre mutations correspondant à nos quatre protéines, on remarque que la délétion  $\Delta 45-51$  est la moins délétère, puis viennent la  $\Delta 45-47$ , la  $\Delta 45-48$  et enfin la  $\Delta 45-49$  qui ne conserve pas la forme filamenteuse de la protéine. Cependant, en utilisant le pourcentage de patients présentant une DCM, ce classement diffère et la sévérité des  $\Delta 45-47$  et  $\Delta 45-48$  s'inverse.

	Nb de patients	Patients avec DCM (%)	Age d'apparition
$\Delta 45-51$ (hybride)	4	25	43
$\Delta 45-48$ (hybride)	23	73	$39,86 \pm 15,74$
$\Delta 45-47$ (fracturé)	26	61	$28,93 \pm 9,48$
$\Delta 45-49$ (fracturé)	8	87	$25,71 \pm 7$

Tableau 4.2 – *Présence et âge d'apparition de DCM chez les patients de la cohorte de [Kaspar 2009].*

Cependant, ce type de corrélation est délicate à établir pour toutes les mutations et avec toutes les cohortes de patients. A l'heure actuelle, les données de la cohorte du LGBM de l'institut Cochin dont nous disposons sont insuffisantes pour réaliser ce type de corrélation. Malgré tout, la délétion  $\Delta 45-51$ , quelque soit l'étude clinique, est la mutation associée au phénotype le plus léger.

Pour conclure, nous pouvons reprendre les trois questions posées dans les objectifs :

- *Comment la structure de la dystrophine permet-elle de comprendre sa fonction ?*

Nous avons montré que, bien que la dystrophine soit majoritairement composée de motifs répétés, ces répétitions présentent des caractéristiques hétérogènes. Nous avons également mis en évidence que la dystrophine est une protéine filamenteuse flexible avec, en plus des

charnières précédemment identifiées, des points de flexion situés au niveau des linkers entre certaines répétitions. Cette nouvelle vision permet de mieux comprendre la présence des nombreux domaines d'interactions caractérisés sur la dystrophine.

- *Que peut-on déduire des mutations identifiées chez les patients ?*

Il est possible de mettre en évidence les domaines de liaison touchés par une mutation ainsi que la structure tertiaire de la protéine mutée. Plus précisément, lorsqu'il s'agit d'une délétion dans le domaine central, la présence d'une répétition hybride ou fracturée à la jonction de la délétion est mise en évidence. eDystrophin permet d'avoir une vue d'ensemble des conséquences de chaque mutation répertoriée sur la protéine produite. De plus, la base de données permet d'avoir un aperçu global des phénotypes associés à cette mutation.

- *Comment l'étude de la structure de la dystrophine et de ses mutations permet-elle de corréler le génotype et le phénotype ?*

L'utilisation combinée des informations répertoriées dans la base de données eDystrophin et de la dynamique moléculaire permet d'apporter les premiers éléments d'une corrélation génotype - phénotype. Cependant, cette corrélation est difficile à mettre en évidence par manque de données cliniques et / ou de données fiables. Il est donc nécessaire de compléter les informations cliniques et d'obtenir plus de détails sur les phénotypes des patients. Une telle étude va démarrer par l'intermédiaire d'un questionnaire réalisé en collaboration avec le LGBM et les curateurs de la base de données UMD-DMD afin de réinterroger l'ensemble des praticiens du réseau français.

Malgré tout, l'étude par dynamique moléculaire de deux répétitions hybrides et de deux répétitions fracturées permet d'émettre l'hypothèse que les répétitions fracturées sont globalement plus délétères que les répétitions hybrides. Les deux répétitions hybrides n'ont pas pour autant les mêmes caractéristiques que les répétitions natives ce qui ne garantit pas que la présence d'une répétition hybride soit associée un phénotype léger.

Nos hypothèses sont toutefois corroborées par les phénotypes observés dans la cohorte de l'étude [Kaspar 2009]. Ce type de mise en relation est une première étape dans la corrélation génotype-phénotype et met en évidence l'importance d'avoir des données cliniques détaillées, fiables et complètes pour mieux comprendre l'impact des délétions sur la sévérité des phénotypes observés.

## Perspectives

Une des perspectives à court terme est de réaliser des dynamiques moléculaires gros-grains sur les trois protéines R16-21,  $\Delta 45-47$  et  $\Delta 45-48$ . Les structures représentatives extraites des dynamiques gros-grain pourront ensuite être utilisées comme nouveau point de départ pour réaliser des dynamiques tout-atome comme décrit dans la section 2.3.3.

De plus, un des moyens pour prédire le maintien ou l'absence des interactions de la dystrophine avec ses partenaires sera d'amarrer les partenaires de la dystrophine à ces dystrophines

tronquées par des méthodes de "docking".

Il reste également à corrélérer les résultats expérimentaux obtenus au sein de l'équipe en biochimie et biophysique avec les résultats obtenus *in silico*. Il sera notamment intéressant de corrélérer les résultats expérimentaux de liaison aux lipides des cinq protéines étudiées par dynamique moléculaire. De plus, l'acquisition des structures de ces protéines par la diffusion de rayons X aux petits angles (SAXS) est en projet et permettrait de les comparer à nos modèles relaxés par MD. La comparaison des résultats *in silico* et expérimentaux obtenus dans l'équipe est en cours et donnera lieu à un article.

L'approche de dynamique moléculaire développée dans ce travail de thèse peut être, à l'avenir, appliquée à d'autres protéines tronquées comme par exemple la protéine  $\Delta 45-46$ . Les données sur cette protéine disponibles dans eDystrophin indiquent la présence d'une répétition hybride à la jonction et peu d'incidence sur la liaison aux partenaires. Or, elle est associée à des phénotypes graves de type DMD. Il serait donc intéressant d'étudier cette protéine de manière plus approfondie afin de comprendre la sévérité des phénotypes associés.

La réinterrogation des praticiens du réseau français UMD-DMD permettra de proposer dans eDystrophin un phénotype clinique détaillé pour toutes les mutations en phase répertoriées. Il est également envisagé de lier eDystrophin à l'UMD-DMD en insérant des liens sur les pages des protéines mutées.

A plus long terme, ce type d'étude sera réalisé dans d'autres régions du domaine central. Les détails phénotypiques nouvellement collectés permettront de choisir les nouvelles protéines mutées à étudier, notamment des protéines associées à des phénotypes "extrêmes" (asymptomatiques et DMD). Nous pourrions ainsi vérifier si nos hypothèses sont transposables à d'autres mutations du domaine central ou si d'autres facteurs entrent en jeu. Il est également envisagé d'étudier plus précisément des protéines avec une mutation faux-sens qui pourrait impliquer un problème de repliement de la protéine. L'étude de ces mutations intéresse particulièrement les cliniciens qui ont souvent un doute sur la pathogénicité de ce type de mutation.



# Bibliographie

- [Aartsma-Rus 2006] A. Aartsma-Rus, J. C. Van Deutekom, I. F. Fokkema, G. J. Van Ommen et J. T. Den Dunnen. *Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule*. Muscle Nerve, vol. 34, no. 2, pages 135–44, 2006.
- [Aartsma-Rus 2009] A. Aartsma-Rus, I. Fokkema, J. Verschuuren, I. Ginjaar, J. van Deutekom, G. J. van Ommen et J. T. den Dunnen. *Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations*. Hum Mutat, vol. 30, no. 3, pages 293–9, 2009.
- [AFM 2009] AFM. *Zoom sur la recherche dans la dystrophie musculaire de Duchenne*, 09 2009.
- [Amann 1998] K. J. Amann, B. A. Renley et J. M. Ervasti. *A cluster of basic repeats in the dystrophin rod domain binds F-actin through an electrostatic interaction*. J Biol Chem, vol. 273, no. 43, pages 28419–23, 1998.
- [Anfinsen 1973] C B Anfinsen. *Principles that govern the folding of protein chains*. Science, vol. 181, no. 4096, pages 223–30, Jul 1973.
- [Anthony 2011] Karen Anthony, Sebahattin Cirak, Silvia Torelli, Giorgio Tasca, Lucy Feng, Virginia Arechavala-Gomez, Annarita Armaroli, Michela Guglieri, Chiara S Straathof, Jan J Verschuuren, Annemieke Aartsma-Rus, Paula Helderma-van den Enden, Katherine Bushby, Volker Straub, Caroline Sewry, Alessandra Ferlini, Enzo Ricci, Jennifer E Morgan et Francesco Muntoni. *Dystrophin quantification and clinical correlations in Becker muscular dystrophy: implications for clinical trials*. Brain, vol. 134, no. Pt 12, pages 3547–59, Dec 2011.
- [Attwood 2011] T.K. Attwood, A. Gisel, N-E. Eriksson et E. Bongcam-Rudloff. *Concepts, historical milestones and the central place of bioinformatics in modern biology: a european perspective*. Intech Online Publishers, 2011.
- [Baines 2003] Anthony J Baines. *Comprehensive analysis of all triple helical repeats in beta-spectrins reveals patterns of selective evolutionary conservation*. Cell Mol Biol Lett, vol. 8, no. 1, pages 195–214, 2003.
- [Baker 2001] N A Baker, D Sept, S Joseph, M J Holst et J A McCammon. *Electrostatics of nanosystems: application to microtubules and the ribosome*. Proc Natl Acad Sci U S A, vol. 98, no. 18, pages 10037–41, Aug 2001.
- [Banuelos 1998] S. Banuelos, M. Saraste et K. Djinic Carugo. *Structural comparisons of calponin homology domains: implications for actin binding*. Structure, vol. 6, no. 11, pages 1419–31, 1998.
- [Beggs 1991] A. H. Beggs, E. P. Hoffman, J. R. Snyder, K. Arahata, L. Specht, F. Shapiro, C. Angelini, H. Sugita et L. M. Kunkel. *Exploring the molecular basis for variability among patients with Becker muscular dystrophy: dystrophin gene and protein studies*. Am J Hum Genet, vol. 49, no. 1, pages 54–67, 1991.

- [Bérout 2005] Christophe Bérout, Dalil Hamroun, Gwenaëlle Collod-Bérout, Catherine Boileau, Thierry Soussi et Mireille Claustres. *UMD (Universal Mutation Database): 2005 update*. Hum Mutat, vol. 26, no. 3, pages 184–91, Sep 2005.
- [Beroud 2007] C. Beroud, S. Tuffery-Giraud, M. Matsuo, D. Hamroun, V. Humbertclaude, N. Monnier, M. P. Moizard, M. A. Voelckel, L. M. Calemard, P. Boisseau, M. Blayau, C. Philippe, M. Cossee, M. Pages, F. Rivier, O. Danos, L. Garcia et M. Claustres. *Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63 Duchenne muscular dystrophy*. Hum Mutat, vol. 28, no. 2, pages 196–202, 2007.
- [Bhosle 2006] R. C. Bhosle, D. E. Michele, K. P. Campbell, Z. Li et R. M. Robson. *Interactions of intermediate filament protein synemin with dystrophin and utrophin*. Biochem Biophys Res Commun, vol. 346, no. 3, pages 768–77, 2006.
- [Blake 1995] D. J. Blake, J. M. Tinsley, K. E. Davies, A. E. Knight, S. J. Winder et J. Kendrick-Jones. *Coiled-coil regions in the carboxy-terminal domains of dystrophin and related proteins: potentials for protein-protein interactions*. Trends Biochem Sci, vol. 20, no. 4, pages 133–5, 1995.
- [Blake 2002] D. J. Blake, A. Weir, S. E. Newey et K. E. Davies. *Function and genetics of dystrophin and dystrophin-related proteins in muscle*. Physiol Rev, vol. 82, no. 2, pages 291–329, 2002.
- [Blau 2008] Helen M Blau. *Cell therapies for muscular dystrophy*. N Engl J Med, vol. 359, no. 13, pages 1403–5, Sep 2008.
- [Bloetzer 2012] Clemens Bloetzer, Joël Fluss et Pierre-Yves Jeannet. *[Therapeutic trials for Duchenne muscular dystrophy: between hopes and disappointments]*. Rev Med Suisse, vol. 8, no. 329, pages 409–12, Feb 2012.
- [Bodensteiner 1978] J B Bodensteiner et A G Engel. *Intracellular calcium accumulation in Duchenne dystrophy and other myopathies: a study of 567,000 muscle fibers in 114 biopsies*. Neurology, vol. 28, no. 5, pages 439–46, May 1978.
- [Boland 1995] B Boland, B Himpens, J F Deneff et J M Gillis. *Site-dependent pathological differences in smooth muscles and skeletal muscles of the adult mdx mouse*. Muscle Nerve, vol. 18, no. 6, pages 649–57, Jun 1995.
- [Bordner 2012] Andrew J Bordner. *Force fields for homology modeling*. Methods Mol Biol, vol. 857, pages 83–106, 2012.
- [Bork 1994] P. Bork et M. Sudol. *The WW domain: a signalling site in dystrophin?* Trends Biochem Sci, vol. 19, no. 12, pages 531–3, 1994.
- [Botstein 1997] D Botstein et J M Cherry. *Molecular linguistics: extracting information from gene and protein sequences*. Proc Natl Acad Sci U S A, vol. 94, no. 11, pages 5506–7, May 1997.
- [Bowles 2012] Dawn E Bowles, Scott W J McPhee, Chengwen Li, Steven J Gray, Jade J Samulski, Angelique S Camp, Juan Li, Bing Wang, Paul E Monahan, Joseph E Rabinowitz, Joshua C Grieger, Lakshmanan Govindasamy, Mavis Agbandje-McKenna, Xiao Xiao et R Jude Samulski. *Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector*. Mol Ther, vol. 20, no. 2, pages 443–55, Feb 2012.
- [Bredt 1999] D S Bredt. *Endogenous nitric oxide synthesis: biological functions and pathophysiology*. Free Radic Res, vol. 31, no. 6, pages 577–96, Dec 1999.
- [Bridwell 1999] K H Bridwell, C Baldus, T M Iffrig, L G Lenke et K Blanke. *Process measures and patient/parent evaluation of surgical management of spinal deformities in patients*

- with progressive flaccid neuromuscular scoliosis (Duchenne's muscular dystrophy and spinal muscular atrophy)*. Spine (Phila Pa 1976), vol. 24, no. 13, pages 1300–9, Jul 1999.
- [Broderick 2005] M. J. Broderick et S. J. Winder. *Spectrin, alpha-actinin, and dystrophin*. Adv Protein Chem, vol. 70, pages 203–46, 2005.
- [Brooke 1989] M H Brooke, G M Fenichel, R C Griggs, J R Mendell, R Moxley, J Florence, W M King, S Pandya, J Robison et J Schierbecker. *Duchenne muscular dystrophy: patterns of clinical progression and effects of supportive therapy*. Neurology, vol. 39, no. 4, pages 475–81, Apr 1989.
- [Bushby 1991] K M Bushby, M Thambyayah et D Gardner-Medwin. *Prevalence and incidence of Becker muscular dystrophy*. Lancet, vol. 337, no. 8748, pages 1022–4, Apr 1991.
- [Bushby 2003] K Bushby, F Muntoni et J P Bourke. *107th ENMC international workshop: the management of cardiac involvement in muscular dystrophy and myotonic dystrophy. 7th-9th June 2002, Naarden, the Netherlands*. Neuromuscul Disord, vol. 13, no. 2, pages 166–72, Feb 2003.
- [Byers 1989] T J Byers, A Husain-Chishti, R R Dubreuil, D Branton et L S Goldstein. *Sequence similarity of the amino-terminal domain of Drosophila beta spectrin to alpha actinin and dystrophin*. J Cell Biol, vol. 109, no. 4 Pt 1, pages 1633–41, Oct 1989.
- [Byers 1993] T. J. Byers, H. G. Lidov et L. M. Kunkel. *An alternative dystrophin transcript specific to peripheral nerve*. Nat Genet, vol. 4, no. 1, pages 77–81, 1993.
- [Carsana 2005] A. Carsana, G. Frisso, M. R. Tremolaterra, R. Lanzillo, D. F. Vitale, L. Santoro et F. Salvatore. *Analysis of dystrophin gene deletions indicates that the hinge III region of the protein correlates with disease severity*. Ann Hum Genet, vol. 69, no. Pt 3, pages 253–9, 2005.
- [Castresana 1995] J Castresana et M Saraste. *Does Vav bind to F-actin through a CH domain?* FEBS Lett, vol. 374, no. 2, pages 149–151, Oct 1995.
- [Cerletti 2003] M Cerletti, T Negri, F Cozzi, R Colpo, F Andreetta, D Croci, K E Davies, F Cornelio, O Pozza, G Karpati, R Gilbert et M Mora. *Dystrophic phenotype of canine X-linked muscular dystrophy is mitigated by adenovirus-mediated utrophin gene transfer*. Gene Ther, vol. 10, no. 9, pages 750–7, May 2003.
- [Chakkalakal 2005] Joe V Chakkalakal, Jennifer Thompson, Robin J Parks et Bernard J Jasmin. *Molecular, cellular, and pharmacological therapies for Duchenne/Becker muscular dystrophies*. FASEB J, vol. 19, no. 8, pages 880–91, Jun 2005.
- [Cheatham 1995] T. E. III Cheatham, J. L. Miller, T. Fox, T. A. Darden et P. A. Kollman. *Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins*. Journal of The American Chemical Society, vol. 117, pages 4193–4194, 1995.
- [Chelly 1990] J. Chelly, G. Hamard, A. Koulakoff, J. C. Kaplan, A. Kahn et Y. Berwald-Netter. *Dystrophin gene transcribed from different promoters in neuronal and glial cells*. Nature, vol. 344, no. 6261, pages 64–5, 1990.
- [Cho 2000] W K Cho, S Ebihara, J Nalbantoglu, R Gilbert, B Massie, P Holland, G Karpati et B J Petrof. *Modulation of Starling forces and muscle fiber maturity permits adenovirus-mediated gene transfer to adult dystrophic (mdx) mice by the intravascular route*. Hum Gene Ther, vol. 11, no. 5, pages 701–14, Mar 2000.
- [Cirak 2011] Sebahattin Cirak, Virginia Arechavala-Gomez, Michela Guglieri, Lucy Feng, Silvia Torelli, Karen Anthony, Stephen Abbs, Maria Elena Garralda, John Bourke,

- Dominic J Wells, George Dickson, Matthew J A Wood, Steve D Wilton, Volker Straub, Ryszard Kole, Stephen B Shrewsbury, Caroline Sewry, Jennifer E Morgan, Kate Bushby et Francesco Muntoni. *Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study*. *Lancet*, vol. 378, no. 9791, pages 595–605, Aug 2011.
- [Claustres 2002] Mireille Claustres, Ourania Horaitis, Marijana Vanevski et Richard G H Cotton. *Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases*. *Genome Res*, vol. 12, no. 5, pages 680–8, May 2002.
- [Clerk 1993] A Clerk, G E Morris, V Dubowitz, K E Davies et C A Sewry. *Dystrophin-related protein, utrophin, in normal and dystrophic human fetal skeletal muscle*. *Histochem J*, vol. 25, no. 8, pages 554–61, Aug 1993.
- [Coffey 1992] A. J. Coffey, R. G. Roberts, E. D. Green, C. G. Cole, R. Butler, R. Anand, F. Giannelli et D. R. Bentley. *Construction of a 2.6-Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach*. *Genomics*, vol. 12, no. 3, pages 474–84, 1992.
- [Connuck 2008] David M Connuck, Lynn A Sleeper, Steven D Colan, Gerald F Cox, Jeffrey A Towbin, April M Lowe, James D Wilkinson, E John Orav, Leigha Cuniberti, Bonnie A Salbert, Steven E Lipshultz et Pediatric Cardiomyopathy Registry Study Group. *Characteristics and outcomes of cardiomyopathy in children with Duchenne or Becker muscular dystrophy: a comparative study from the Pediatric Cardiomyopathy Registry*. *Am Heart J*, vol. 155, no. 6, pages 998–1005, Jun 2008.
- [Cordier 2001] L Cordier, G P Gao, A A Hack, E M McNally, J M Wilson, N Chirmule et H L Sweeney. *Muscle-specific promoters may be necessary for adeno-associated virus-mediated gene transfer in the treatment of muscular dystrophies*. *Hum Gene Ther*, vol. 12, no. 2, pages 205–15, Jan 2001.
- [Cozzetto 2009] Domenico Cozzetto, Andriy Kryshtafovych, Krzysztof Fidelis, John Moulton, Burkhard Rost et Anna Tramontano. *Evaluation of template-based models in CASP8 with standard measures*. *Proteins*, vol. 77 Suppl 9, pages 18–28, 2009.
- [Cramer 2004] C. J. Cramer. *Essentials of computational chemistry: theories and models*. John Wiley and Sons Inc., New York, 2004.
- [Da Costa 2012] Grégory Da Costa, Arnaud Bondon, Olivier Delalande, Liza Mouret et Jean-Pierre Monti. *Elucidation by NMR solution of neurotensin in small unilamellar vesicle environment: molecular surveys for neurotensin receptor recognition*. *J Biomol Struct Dyn*, Aug 2012.
- [Davies 2006] Kay E Davies et Kristen J Nowak. *Molecular mechanisms of muscular dystrophies: old and new players*. *Nat Rev Mol Cell Biol*, vol. 7, no. 10, pages 762–73, Oct 2006.
- [Davison 1988] M D Davison et D R Critchley. *alpha-Actinins and the DMD protein contain spectrin-like repeats*. *Cell*, vol. 52, no. 2, pages 159–60, Jan 1988.
- [De Backer 2002] F De Backer, C Vandebrouck, P Gailly et J M Gillis. *Long-term study of Ca(2+) homeostasis and of survival in collagenase-isolated muscle fibres from normal and mdx mice*. *J Physiol*, vol. 542, no. Pt 3, pages 855–65, Aug 2002.
- [de Brem 2012] Paul de Brem. *Le déluge informationnel*. *Med Sci (Paris)*, vol. 28 Spec No 2, page 6, Jun 2012.

- [Deburgrave 2007] N. Deburgrave, F. Daoud, S. Llense, J. C. Barbot, D. Recan, C. Peccate, A. H. Burghes, C. Beroud, L. Garcia, J. C. Kaplan, J. Chelly et F. Leturcq. *Protein- and mRNA-based phenotype-genotype correlations in DMD/BMD with point mutations and molecular basis for BMD with nonsense and frameshift mutations in the DMD gene*. Hum Mutat, vol. 28, no. 2, pages 183–95, 2007.
- [Deconinck 2007] N. Deconinck et B. Dan. *Pathophysiology of duchenne muscular dystrophy: current hypotheses*. Pediatr Neurol, vol. 36, no. 1, pages 1–7, 2007.
- [Denti 2006] Michela Alessandra Denti, Alessandro Rosa, Giuseppe D’Antona, Olga Sthandier, Fernanda Gabriella De Angelis, Carmine Nicoletti, Mariacarmela Allocca, Orietta Pansarasa, Valeria Parente, Antonio Musarò, Alberto Auricchio, Roberto Bottinelli et Irene Bozzoni. *Body-wide gene therapy of Duchenne muscular dystrophy in the mdx mouse model*. Proc Natl Acad Sci U S A, vol. 103, no. 10, pages 3758–63, Mar 2006.
- [Desguerre 2009] I. Desguerre, C. Christov, M. Mayer, R. Zeller, H. M. Becane, S. Bastuji-Garin, F. Leturcq, C. Chiron, J. Chelly et R. K. Gherardi. *Clinical heterogeneity of duchenne muscular dystrophy (DMD): definition of sub-phenotypes and predictive criteria by long-term follow-up*. PLoS One, vol. 4, no. 2, page e4347, 2009.
- [Doing 2002] Anthony H Doing, Dale G Renlund et Ruth Ann Smith. *Becker muscular dystrophy-related cardiomyopathy: a favorable response to medical therapy*. J Heart Lung Transplant, vol. 21, no. 4, pages 496–8, Apr 2002.
- [Drin 2007] Guillaume Drin, Jean-François Casella, Romain Gautier, Thomas Boehmer, Thomas U Schwartz et Bruno Antonny. *A general amphipathic alpha-helical motif for sensing membrane curvature*. Nat Struct Mol Biol, vol. 14, no. 2, pages 138–46, Feb 2007.
- [D’Souza 1995] V. N. D’Souza, T. M. Nguyen, G. E. Morris, W. Karges, D. A. Pillers et P. N. Ray. *A novel dystrophin isoform is required for normal retinal electrophysiology*. Hum Mol Genet, vol. 4, no. 5, pages 837–42, 1995.
- [Duan 2003] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang et Peter Kollman. *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*. J Comput Chem, vol. 24, no. 16, pages 1999–2012, Dec 2003.
- [Eagle 2007] Michelle Eagle, John Bourke, Robert Bullock, Mike Gibson, Jwalant Mehta, Dave Giddings, Volker Straub et Kate Bushby. *Managing Duchenne muscular dystrophy—the additive effect of spinal surgery and home nocturnal ventilation in improving survival*. Neuromuscul Disord, vol. 17, no. 6, pages 470–5, Jun 2007.
- [Emery 2002] A. E. Emery. *The muscular dystrophies*. Lancet, vol. 359, no. 9307, pages 687–95, 2002.
- [England 1990] S. B. England, L. V. Nicholson, M. A. Johnson, S. M. Forrest, D. R. Love, E. E. Zubrzycka-Gaarn, D. E. Bulman, J. B. Harris et K. E. Davies. *Very mild muscular dystrophy associated with the deletion of 46of dystrophin*. Nature, vol. 343, no. 6254, pages 180–2, 1990.
- [Ervasti 2006] J. M. Ervasti. *Dystrophin, its interactions with other proteins, and implications for muscular dystrophy*. Biochim Biophys Acta, vol. 1772, no. 2, pages 108–17, 2006.
- [Fairclough 2011] Rebecca J Fairclough, Akshay Bareja et Kay E Davies. *Progress in therapy for Duchenne muscular dystrophy*. Exp Physiol, vol. 96, no. 11, pages 1101–13, Nov 2011.

- [Farini 2009] Andrea Farini, Paola Razini, Silvia Erratico, Yvan Torrente et Mirella Meregalli. *Cell based therapy for Duchenne muscular dystrophy*. J Cell Physiol, vol. 221, no. 3, pages 526–34, Dec 2009.
- [Finkel 2010] Richard S Finkel. *Read-through strategies for suppression of nonsense mutations in Duchenne/ Becker muscular dystrophy: aminoglycosides and ataluren (PTC124)*. J Child Neurol, vol. 25, no. 9, pages 1158–64, Sep 2010.
- [Finn 2010] R. D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy et A. Bateman. *The Pfam protein families database*. Nucleic Acids Res, vol. 38, no. Database issue, pages D211–22, 2010.
- [Finsterer 1999] J Finsterer, R E Bittner et M Grimm. *Cardiac involvement in Becker's muscular dystrophy, necessitating heart transplantation, 6 years before apparent skeletal muscle involvement*. Neuromuscul Disord, vol. 9, no. 8, pages 598–600, Dec 1999.
- [Finsterer 2003] Josef Finsterer et Claudia Stöllberger. *The heart in human dystrophinopathies*. Cardiology, vol. 99, no. 1, pages 1–19, 2003.
- [Finsterer 2008] Josef Finsterer et Claudia Stöllberger. *Cardiac involvement in Becker muscular dystrophy*. Can J Cardiol, vol. 24, no. 10, pages 786–92, Oct 2008.
- [Flanigan 2009] K. M. Flanigan, D. M. Dunn, A. von Niederhausern, P. Soltanzadeh, E. Gappmaier, M. T. Howard, J. B. Sampson, J. R. Mendell, C. Wall, W. M. King, A. Pestronk, J. M. Florence, A. M. Connolly, K. D. Mathews, C. M. Stephan, K. S. Laubenthal, B. L. Wong, P. J. Morehart, A. Meyer, R. S. Finkel, C. G. Bonnemann, L. Medne, J. W. Day, J. C. Dalton, M. K. Margolis, V. J. Hinton et R. B. Weiss. *Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort*. Hum Mutat, vol. 30, no. 12, pages 1657–1666, 2009.
- [Fokkema 2005] Ivo F A C Fokkema, Johan T den Dunnen et Peter E M Taschner. *LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach*. Hum Mutat, vol. 26, no. 2, pages 63–8, Aug 2005.
- [Fokkema 2011] Ivo F A C Fokkema, Peter E M Taschner, Gerard C P Schaafsma, J Celli, Jeroen F J Laros et Johan T den Dunnen. *LOVD v.2.0: the next generation in gene variant databases*. Hum Mutat, vol. 32, no. 5, pages 557–63, May 2011.
- [Foster 2008] Helen Foster, Paul S Sharp, Takis Athanasopoulos, Capucine Trollet, Ian R Graham, Keith Foster, Dominic J Wells et George Dickson. *Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer*. Mol Ther, vol. 16, no. 11, pages 1825–32, Nov 2008.
- [Frishman 1995] D Frishman et P Argos. *Knowledge-based protein secondary structure assignment*. Proteins, vol. 23, no. 4, pages 566–79, Dec 1995.
- [Gardner-Medwin D 1974] Walton JN Gardner-Medwin D. In disorders of voluntary muscles. Edinburgh and London., 1974.
- [Gentil 2012] Christel Gentil, France Leturcq, Rabah Ben Yaou, Jean-Claude Kaplan, Pascal Laforet, Isabelle Péniisson-Besnier, Caroline Espil-Taris, Thomas Voit, Luis Garcia et France Piétri-Rouxel. *Variable phenotype of del45-55 Becker patients correlated with nNOS $\mu$  mislocalization and RYR1 hypernitrosylation*. Hum Mol Genet, May 2012.
- [Gilbert 1999] R Gilbert, J Nalbantoglu, B J Petrof, S Ebihara, G H Guibinga, J M Tinsley, A Kamen, B Massie, K E Davies et G Karpati. *Adenovirus-mediated utrophin gene*

- transfer mitigates the dystrophic phenotype of mdx mouse muscles.* Hum Gene Ther, vol. 10, no. 8, pages 1299–310, May 1999.
- [Gimona 1998] M Gimona et R Mital. *The single CH domain of calponin is neither sufficient nor necessary for F-actin binding.* J Cell Sci, vol. 111 ( Pt 13), pages 1813–1821, Jul 1998.
- [Goemans 2011] Nathalie M Goemans, Mar Tulinius, Johanna T van den Akker, Brigitte E Burm, Peter F Ekhardt, Niki Heuvelmans, Tjadine Holling, Anneke A Janson, Gerard J Platenburg, Jessica A Sipkens, J M Ad Sitsen, Annemieke Aartsma-Rus, Gert-Jan B van Ommen, Gunnar Buyse, Niklas Darin, Jan J Verschuuren, Giles V Campion, Sjef J de Kimpe et Judith C van Deutekom. *Systemic administration of PRO051 in Duchenne’s muscular dystrophy.* N Engl J Med, vol. 364, no. 16, pages 1513–22, Apr 2011.
- [Goyenville 2004] A. Goyenville, A. Vulin, F. Fougerousse, F. Leturcq, J. C. Kaplan, L. Garcia et O. Danos. *Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping.* Science, vol. 306, no. 5702, pages 1796–9, 2004.
- [Goyenville 2011] Aurélie Goyenville, Jane T Seto, Kay E Davies et Jeffrey Chamberlain. *Therapeutic approaches to muscular dystrophy.* Hum Mol Genet, vol. 20, no. R1, pages R69–78, Apr 2011.
- [Gregorevic 2006] P. Gregorevic, J. M. Allen, E. Minami, M. J. Blankinship, M. Haraguchi, L. Meuse, E. Finn, M. E. Adams, S. C. Froehner, C. E. Murry et J. S. Chamberlain. *rAAV6-microdystrophin preserves muscle function and extends lifespan in severely dystrophic mice.* Nat Med, vol. 12, no. 7, pages 787–9, 2006.
- [Gregorevic 2008] P. Gregorevic, M. J. Blankinship, J. M. Allen et J. S. Chamberlain. *Systemic microdystrophin gene delivery improves skeletal muscle structure and function in old dystrophic mdx mice.* Mol Ther, vol. 16, no. 4, pages 657–64, 2008.
- [Gurvich 2008] Olga L Gurvich, Therese M Tuohy, Michael T Howard, Richard S Finkel, Livija Medne, Christine B Anderson, Robert B Weiss, Steve D Wilton et Kevin M Flanigan. *DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy.* Ann Neurol, vol. 63, no. 1, pages 81–9, Jan 2008.
- [Hacein-Bey-Abina 2003] Salima Hacein-Bey-Abina, Christof von Kalle, Manfred Schmidt, Françoise Le Deist, Nicolas Wulffraat, Elisabeth McIntyre, Isabelle Radford, Jean-Luc Villeval, Christopher C Fraser, Marina Cavazzana-Calvo et Alain Fischer. *A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency.* N Engl J Med, vol. 348, no. 3, pages 255–6, Jan 2003.
- [Hartigan-O’Connor 2001] D Hartigan-O’Connor, C J Kirk, R Crawford, J J Mulé et J S Chamberlain. *Immune evasion by muscle-specific gene expression in dystrophic muscle.* Mol Ther, vol. 4, no. 6, pages 525–33, Dec 2001.
- [Head 1994] S Head, D Williams et G Stephenson. *Increased susceptibility of EDL muscles from mdx mice to damage induced by contraction with stretch.* J Muscle Res Cell Motil, vol. 15, no. 4, pages 490–2, Aug 1994.
- [Herzog 1999] R W Herzog, E Y Yang, L B Couto, J N Hagstrom, D Elwell, P A Fields, M Burton, D A Bellinger, M S Read, K M Brinkhous, G M Podsakoff, T C Nichols, G J Kurtzman et K A High. *Long-term correction of canine hemophilia B by gene transfer of blood coagulation factor IX mediated by adeno-associated viral vector.* Nat Med, vol. 5, no. 1, pages 56–63, Jan 1999.
- [Heslop 2000] L Heslop, J E Morgan et T A Partridge. *Evidence for a myogenic stem cell that is exhausted in dystrophic muscle.* J Cell Sci, vol. 113 ( Pt 12), pages 2299–308, Jun 2000.

- [Hess 2008] Berk Hess, Carsten Kutzner, David van der Spoel et Erik Lindahl. *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*. Journal of Chemical Theory and Computation, vol. 4, pages 435–447, 2008.
- [Hnia 2007] K. Hnia, D. Zouiten, S. Cantel, D. Chazalotte, G. Hugon, J. A. Fehrentz, A. Mas-moudi, A. Diment, J. Bramham, D. Mornet et S. J. Winder. *ZZ domain of dystrophin and utrophin: topology and mapping of a beta-dystroglycan interaction site*. Biochem J, vol. 401, no. 3, pages 667–77, 2007.
- [Hogeweg 1978] P Hogeweg et B Hesper. *Interactive instruction on population interactions*. Comput Biol Med, vol. 8, no. 4, pages 319–27, 1978.
- [Huang 2000] X. Huang, F. Poy, R. Zhang, A. Joachimiak, M. Sudol et M. J. Eck. *Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan*. Nat Struct Biol, vol. 7, no. 8, pages 634–8, 2000.
- [Hughes 1996] M I Hughes, E M Hicks, N C Nevin et V H Patterson. *The prevalence of inherited neuromuscular disease in Northern Ireland*. Neuromuscul Disord, vol. 6, no. 1, pages 69–73, Jan 1996.
- [Hugnot 1992] J. P. Hugnot, H. Gilgenkrantz, N. Vincent, P. Chafey, G. E. Morris, A. P. Monaco, Y. Berwald-Netter, A. Koulakoff, J. C. Kaplan, A. Kahn et et al. *Distal transcript of the dystrophin gene initiated from an alternative first exon and encoding a 75-kDa protein widely distributed in nonmuscle tissues*. Proc Natl Acad Sci U S A, vol. 89, no. 16, pages 7506–10, 1992.
- [Humbertclaude 2010] V Humbertclaude, S Tuffery-Giraud, C Bareil, C Thèze, D Paulet, F-O Desmet, D Hamroun, D Baux, A Girardet, G Collod-Bérourd, P Khau Van Kien, A-F Roux, M des Georges, C Bérourd et M Claustres. [*Genetic mutation databases: stakes and perspectives for orphan genetic diseases*]. Pathol Biol (Paris), vol. 58, no. 5, pages 387–95, Oct 2010.
- [Humbertclaude 2012] Véronique Humbertclaude, Dalil Hamroun, Kamel Bezzou, Carole Bérard, Odile Boespflug-Tanguy, Christine Bommelaer, Emmanuelle Campana-Salort, Claude Cancès, Brigitte Chabrol, Marie-Christine Commare, Jean-Marie Cuisset, Capucine de Lattre, Claude Desnuelle, Bernard Echenne, Cécile Halbert, Olivier Jonquet, Annick Labarre-Vila, Marie-Ange N’Guyen-Morel, Michel Pages, Jean-Louis Pepin, Thierry Petitjean, Jean Pouget, Elisabeth Ollagnon-Roman, Christian Richelme, François Rivier, Sabrina Sacconi, Vincent Tiffreau, Carole Vuillerot, Marie-Christine Picot, Mireille Claustres, Christophe Bérourd et Sylvie Tuffery-Giraud. *Motor and respiratory heterogeneity in Duchenne patients: implication for clinical trials*. Eur J Paediatr Neurol, vol. 16, no. 2, pages 149–60, Mar 2012.
- [Humphrey 1996] W Humphrey, A Dalke et K Schulten. *VMD: visual molecular dynamics*. J Mol Graph, vol. 14, no. 1, pages 33–8, 27–8, Feb 1996.
- [Ikura 1996] M Ikura. *Calcium binding and conformational response in EF-hand proteins*. Trends Biochem Sci, vol. 21, no. 1, pages 14–17, Jan 1996.
- [Ishikawa-Sakurai 2004] M. Ishikawa-Sakurai, M. Yoshida, M. Imamura, K. E. Davies et E. Ozawa. *ZZ domain is essentially required for the physiological binding of dystrophin and utrophin to beta-dystroglycan*. Hum Mol Genet, vol. 13, no. 7, pages 693–702, 2004.
- [Jearawiriyapaisarn 2010] Natee Jearawiriyapaisarn, Hong M Moulton, Peter Sazani, Ryszard Kole et Monte S Willis. *Long-term improvement in mdx cardiomyopathy after therapy with peptide-conjugated morpholino oligomers*. Cardiovasc Res, vol. 85, no. 3, pages 444–53, Feb 2010.

- [Jeppesen 2003] J Jeppesen, A Green, B F Steffensen et J Rahbek. *The Duchenne muscular dystrophy population in Denmark, 1977-2001: prevalence, incidence and survival in relation to the introduction of ventilator use*. Neuromuscul Disord, vol. 13, no. 10, pages 804–12, Dec 2003.
- [Jin 2007] H. Jin, S. Tan, J. Hermanowski, S. Bohm, S. Pacheco, J. M. McCauley, M. J. Greener, Y. Hinitz, S. M. Hughes, P. T. Sharpe et R. G. Roberts. *The dystrotelin, dystrophin and dystrobrevin superfamily: new paralogues and old isoforms*. BMC Genomics, vol. 8, page 19, 2007.
- [Kaspar 2009] R. W. Kaspar, H. D. Allen, W. C. Ray, C. E. Alvarez, J. T. Kissel, A. Pestronk, R. B. Weiss, K. M. Flanigan, J. R. Mendell et F. Montanaro. *Analysis of dystrophin deletion mutations predicts age of cardiomyopathy onset in becker muscular dystrophy*. Circ Cardiovasc Genet, vol. 2, no. 6, pages 544–51, 2009.
- [Kesari 2008] Akanchha Kesari, Laura N Pirra, Lakshmi Bremadesam, Orinthal McIntyre, Erynn Gordon, Alberto L Dubrovsky, V Viswanathan et Eric P Hoffman. *Integrated DNA, cDNA, and protein studies in Becker muscular dystrophy show high exception to the reading frame rule*. Hum Mutat, vol. 29, no. 5, pages 728–37, May 2008.
- [Khurana 1991] T S Khurana, S C Watkins, P Chafey, J Chelly, F M Tomé, M Fardeau, J C Kaplan et L M Kunkel. *Immunolocalization and developmental expression of dystrophin related protein in skeletal muscle*. Neuromuscul Disord, vol. 1, no. 3, pages 185–94, 1991.
- [Koenig 1987] M Koenig, E P Hoffman, C J Bertelson, A P Monaco, C Feener et L M Kunkel. *Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals*. Cell, vol. 50, no. 3, pages 509–17, Jul 1987.
- [Koenig 1988] M. Koenig, A. P. Monaco et L. M. Kunkel. *The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein*. Cell, vol. 53, no. 2, pages 219–28, 1988.
- [Koenig 1989] M. Koenig, A. H. Beggs, M. Moyer, S. Scherpf, K. Heindrich, T. Bettecken, G. Meng, C. R. Muller, M. Lindlof, H. Kaariainen et et al. *The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion*. Am J Hum Genet, vol. 45, no. 4, pages 498–506, 1989.
- [Kryshtafovych 2011a] Andriy Kryshtafovych, Krzysztof Fidelis et John Moulton. *CASP9 results compared to those of previous CASP experiments*. Proteins, vol. 79 Suppl 10, pages 196–207, 2011.
- [Kryshtafovych 2011b] Andriy Kryshtafovych, John Moulton, Sergio G Bartual, J Fernando Bazan, Helen Berman, Darren E Casteel, Evangelos Christodoulou, John K Everett, Jens Hausmann, Tatjana Heidebrecht, Tanya Hills, Raymond Hui, John F Hunt, Jayaraman Seetharaman, Andrzej Joachimiak, Michael A Kennedy, Choel Kim, Andreas Lingel, Karolina Michalska, Gaetano T Montelione, José M Otero, Anastassis Perrakis, Juan C Pizarro, Mark J van Raaij, Theresa A Ramelot, Francois Rousseau, Liang Tong, Amy K Wernimont, Jasmine Young et Torsten Schwede. *Target highlights in CASP9: Experimental target structures for the critical assessment of techniques for protein structure prediction*. Proteins, vol. 79 Suppl 10, pages 6–20, 2011.
- [Lai 2009] Yi Lai, Gail D Thomas, Yongping Yue, Hsiao T Yang, Dejia Li, Chun Long, Luke Judge, Brian Bostick, Jeffrey S Chamberlain, Ronald L Terjung et Dongsheng Duan. *Dystrophins carrying spectrin-like repeats 16 and 17 anchor nNOS to the sarcolemma and enhance exercise performance in a mouse model of muscular dystrophy*. J Clin Invest, vol. 119, no. 3, pages 624–35, Mar 2009.

- [Laskowski 1993] R. A. Laskowski, M. W. MacArthur, D. S. Moss et J. M. Thornton. *PRO-CHECK: a program to check the stereochemical quality of protein structures*. Journal of Applied Crystallography, vol. 26, no. 2, pages 283–291, Apr 1993.
- [Le Rumeur 2003] Elisabeth Le Rumeur, Yann Fichou, Sandrine Pottier, François Gaboriau, Corinne Rondeau-Mouro, Michel Vincent, Jacques Gallay et Arnaud Bondon. *Interaction of dystrophin rod domain with membrane phospholipids. Evidence of a close proximity between tryptophan residues and lipids*. J Biol Chem, vol. 278, no. 8, pages 5993–6001, Feb 2003.
- [Le Rumeur 2010] E. Le Rumeur, S. J. Winder et J. F. Hubert. *Dystrophin: More than just the sum of its parts*. Biochim Biophys Acta, vol. 1804, no. 9, pages 1713–22, 2010.
- [Leach 2001] A R Leach. Molecular modeling principles and applications. Pearson Education Limited, 2001.
- [Legardinier 2008a] S. Legardinier, J. F. Hubert, O. Le Bihan, C. Tascon, C. Rocher, C. Raguènes-Nicol, A. Bondon, S. Hardy et E. Le Rumeur. *Sub-domains of the dystrophin rod domain display contrasting lipid-binding and stability properties*. Biochim Biophys Acta, vol. 1784, no. 4, pages 672–82, 2008.
- [Legardinier 2008b] Sébastien Legardinier, Jean-François Hubert, Olivier Le Bihan, Christophe Tascon, Chantal Rocher, Céline Raguènes-Nicol, Arnaud Bondon, Serge Hardy et Elisabeth Le Rumeur. *Sub-domains of the dystrophin rod domain display contrasting lipid-binding and stability properties*. Biochim Biophys Acta, vol. 1784, no. 4, pages 672–82, Apr 2008.
- [Legardinier 2009a] S. Legardinier, B. Legrand, C. Raguènes-Nicol, A. Bondon, S. Hardy, C. Tascon, E. Le Rumeur et J. F. Hubert. *A Two-amino Acid Mutation Encountered in Duchenne Muscular Dystrophy Decreases Stability of the Rod Domain 23 (R23) Spectrin-like Repeat of Dystrophin*. J Biol Chem, vol. 284, no. 13, pages 8822–32, 2009.
- [Legardinier 2009b] S. Legardinier, C. Raguènes-Nicol, C. Tascon, C. Rocher, S. Hardy, J. F. Hubert et E. Le Rumeur. *Mapping of the lipid-binding and stability properties of the central rod domain of human dystrophin*. J Mol Biol, vol. 389, no. 3, pages 546–58, 2009.
- [Legrand 2011] B. Legrand, E. Giudice, A. Nicolas, O. Delalande et E. Le Rumeur. *Computational study of the human dystrophin repeats: interaction properties and molecular dynamics*. PLoS One, vol. 6, no. 8, page e23819, 2011.
- [Letunic 2009] I. Letunic, T. Doerks et P. Bork. *SMART 6: recent updates and new developments*. Nucleic Acids Res, vol. 37, no. Database issue, pages D229–32, 2009.
- [Li 2010] Deji Li, Akshay Bareja, Luke Judge, Yongping Yue, Yi Lai, Rebecca Fairclough, Kay E Davies, Jeffrey S Chamberlain et Dongsheng Duan. *Sarcolemmal nNOS anchoring reveals a qualitative difference between dystrophin and utrophin*. J Cell Sci, vol. 123, no. Pt 12, pages 2008–13, Jun 2010.
- [Lidov 1995] H. G. Lidov, S. Selig et L. M. Kunkel. *Dp140: a novel 140 kDa CNS transcript from the dystrophin locus*. Hum Mol Genet, vol. 4, no. 3, pages 329–35, 1995.
- [Lu 2005] Q. L. Lu, A. Rabinowitz, Y. C. Chen, T. Yokota, H. Yin, J. Alter, A. Jadoon, G. Bou-Gharios et T. Partridge. *Systemic delivery of antisense oligoribonucleotide restores dystrophin expression in body-wide skeletal muscles*. Proc Natl Acad Sci U S A, vol. 102, no. 1, pages 198–203, 2005.
- [Lupas 1996] A Lupas. *Prediction and analysis of coiled-coil structures*. Methods Enzymol, vol. 266, pages 513–25, 1996.

- [Luthy 1992] R. Luthy, J. U. Bowie et D. Eisenberg. *Assessment of protein models with three-dimensional profiles*. Nature, vol. 356, no. 6364, pages 83–5, 1992.
- [Malik 2010] Vinod Malik, Louise R Rodino-Klapac, Laurence Viollet, Cheryl Wall, Wendy King, Roula Al-Dahhak, Sarah Lewis, Christopher J Shilling, Janaiah Kota, Carmen Serrano-Munuera, John Hayes, John D Mahan, Katherine J Campbell, Brenda Banwell, Majed Dasouki, Victoria Watts, Kumaraswamy Sivakumar, Ricardo Bien-Willner, Kevin M Flanigan, Zarife Sahenk, Richard J Barohn, Christopher M Walker et Jerry R Mendell. *Gentamicin-induced readthrough of stop codons in Duchenne muscular dystrophy*. Ann Neurol, vol. 67, no. 6, pages 771–80, Jun 2010.
- [Manno 2003] Catherine S Manno, Amy J Chew, Sylvia Hutchison, Peter J Larson, Roland W Herzog, Valder R Arruda, Shing Jen Tai, Margaret V Ragni, Arthur Thompson, Margareth Ozelo, Linda B Couto, Debra G B Leonard, Frederick A Johnson, Alan McClelland, Ciaran Scallan, Erik Skarsgard, Alan W Flake, Mark A Kay, Katherine A High et Bertil Glader. *AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B*. Blood, vol. 101, no. 8, pages 2963–72, Apr 2003.
- [Manzur 2009] A Y Manzur et F Muntoni. *Diagnosis and new treatments in muscular dystrophies*. J Neurol Neurosurg Psychiatry, vol. 80, no. 7, pages 706–14, Jul 2009.
- [Marchler-Bauer 2009] A. Marchler-Bauer, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, S. Lu, G. H. Marchler, M. Mullokandov, J. S. Song, A. Tasneem, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang et S. H. Bryant. *CDD: specific functional annotation with the Conserved Domain Database*. Nucleic Acids Res, vol. 37, no. Database issue, pages D205–10, 2009.
- [Marrink 2007] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman et Alex H. de Vries. *The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations*. Journal of Physical Chemistry B, vol. 111, pages 7812–7824, 2007.
- [Matsumura 1994] K Matsumura, A H Burghes, M Mora, F M Tomé, L Morandi, F Cornello, F Leturcq, M Jeanpierre, J C Kaplan et P Reinert. *Immunohistochemical analysis of dystrophin-associated proteins in Becker/Duchenne muscular dystrophy with huge in-frame deletions in the NH<sub>2</sub>-terminal and rod domains of dystrophin*. J Clin Invest, vol. 93, no. 1, pages 99–105, Jan 1994.
- [Matsuo 1996] M Matsuo. *Duchenne/Becker muscular dystrophy: from molecular diagnosis to gene therapy*. Brain Dev, vol. 18, no. 3, pages 167–72, 1996.
- [McCammon 1977] J A McCammon, B R Gelin et M Karplus. *Dynamics of folded proteins*. Nature, vol. 267, no. 5612, pages 585–90, Jun 1977.
- [McDouall 1990] R M McDouall, M J Dunn et V Dubowitz. *Nature of the mononuclear infiltrate and the mechanism of muscle damage in juvenile dermatomyositis and Duchenne muscular dystrophy*. J Neurol Sci, vol. 99, no. 2-3, pages 199–217, Nov 1990.
- [Mendell 2006] Jerry R Mendell et K Reed Clark. *Challenges for gene therapy for muscular dystrophy*. Curr Neurol Neurosci Rep, vol. 6, no. 1, pages 47–56, Jan 2006.
- [Mendell 2010] Jerry R Mendell, Katherine Campbell, Louise Rodino-Klapac, Zarife Sahenk, Chris Shilling, Sarah Lewis, Dawn Bowles, Steven Gray, Chengwen Li, Gloria Gallo-way, Vinod Malik, Brian Coley, K Reed Clark, Juan Li, Xiao Xiao, Jade Samulski, Scott W McPhee, R Jude Samulski et Christopher M Walker. *Dystrophin immunity in Duchenne’s muscular dystrophy*. N Engl J Med, vol. 363, no. 15, pages 1429–37, Oct 2010.

- [Meng 2009] Xuan-Yu Meng, Qing-Chuan Zheng et Hong-Xing Zhang. *A comparative analysis of binding sites between mouse CYP2C38 and CYP2C39 based on homology modeling, molecular dynamics simulation and docking studies*. *Biochim Biophys Acta*, vol. 1794, no. 7, pages 1066–72, Jul 2009.
- [Menhart 2006] N. Menhart. *Hybrid spectrin type repeats produced by exon-skipping in dystrophin*. *Biochim Biophys Acta*, vol. 1764, no. 6, pages 993–9, 2006.
- [Metzger 2001] T G Metzger, M G Paterlini, D M Ferguson et P S Portoghesi. *Investigation of the selectivity of oxymorphone- and naltrexone-derived ligands via site-directed mutagenesis of opioid receptors: exploring the "address" recognition locus*. *J Med Chem*, vol. 44, no. 6, pages 857–62, Mar 2001.
- [Mirijanian 2007] Dina T Mirijanian, Jih-Wei Chu, Gary S Ayton et Gregory A Voth. *Atomistic and coarse-grained analysis of double spectrin repeat units: the molecular origins of flexibility*. *J Mol Biol*, vol. 365, no. 2, pages 523–34, Jan 2007.
- [Mirza 2010] A. Mirza, M. Sagathevan, N. Sahni, L. Choi et N. Menhart. *A biophysical map of the dystrophin rod*. *Biochim Biophys Acta*, vol. 1804, no. 9, pages 1796–809, 2010.
- [Mok 2006] Elise Mok, Laurent Béghin, Pierre Gachon, Christel Daubrosse, Jean-Eudes Fontan, Jean-Marie Cuisset, Frédéric Gottrand et Régis Hankard. *Estimating body composition in children with Duchenne muscular dystrophy: comparison of bioelectrical impedance analysis and skinfold-thickness measurement*. *Am J Clin Nutr*, vol. 83, no. 1, pages 65–9, Jan 2006.
- [Monaco 1986] A P Monaco, R L Neve, C Colletti-Feener, C J Bertelson, D M Kurnit et L M Kunkel. *Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene*. *Nature*, vol. 323, no. 6089, pages 646–50, 1986.
- [Monaco 1988] A. P. Monaco, C. J. Bertelson, S. Liechti-Gallati, H. Moser et L. M. Kunkel. *An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus*. *Genomics*, vol. 2, no. 1, pages 90–5, 1988.
- [Monaco 1992] A P Monaco, A P Walker, I Millwood, Z Larin et H Lehrach. *A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene*. *Genomics*, vol. 12, no. 3, pages 465–73, Mar 1992.
- [Monticelli 2008] Luca Monticelli, Senthil Kandasamy, Xavier Periolo, Ronald Larson, Peter Tieleman et Siewert-Jan Marrink. *The MARTINI Coarse-Grained Force Field: Extension to Proteins*. *J. Chem. Theory Comput.*, vol. 4, no. 5, pages 819–834, 2008.
- [Mostacciuolo 1987] M L Mostacciuolo, A Lombardi, V Cambissa, G A Danieli et C Angelini. *Population data on benign and severe forms of X-linked muscular dystrophy*. *Hum Genet*, vol. 75, no. 3, pages 217–20, Mar 1987.
- [Moulton 2010] Hong M Moulton et Jon D Moulton. *Morpholinos and their peptide conjugates: therapeutic promise and challenge for Duchenne muscular dystrophy*. *Biochim Biophys Acta*, vol. 1798, no. 12, pages 2296–303, Dec 2010.
- [Muntoni 2003] F. Muntoni, S. Torelli et A. Ferlini. *Dystrophin and mutations: one gene, several proteins, multiple phenotypes*. *Lancet Neurol*, vol. 2, no. 12, pages 731–40, 2003.
- [Muthu 2012] Muralidharan Muthu, Kylie A Richardson et Andrew J Sutherland-Smith. *The crystal structures of dystrophin and utrophin spectrin repeats: implications for domain boundaries*. *PLoS One*, vol. 7, no. 7, page e40066, 2012.
- [Nevo 2003] Y. Nevo, F. Muntoni, C. Sewry, C. Legum, M. Kutai, S. Harel et V. Dubowitz. *Large in-frame deletions of the rod-shaped domain of the dystrophin gene resulting in severe phenotype*. *Isr Med Assoc J*, vol. 5, no. 2, pages 94–7, 2003.

- [Newey 2000] S. E. Newey, M. A. Benson, C. P. Ponting, K. E. Davies et D. J. Blake. *Alternative splicing of dystrobrevin regulates the stoichiometry of syntrophin binding to the dystrophin protein complex*. *Curr Biol*, vol. 10, no. 20, pages 1295–8, 2000.
- [Nicolas 2012] Aurélie Nicolas, Céline Lucchetti-Miganeh, Rabah Ben Yaou, Jean-Claude Kaplan, Jamel Chelly, France Leturcq, Frédérique Barloy-Hubler et Elisabeth Le Rumeur. *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*, vol. 7, no. 1, page 45, Jul 2012.
- [Nishio 1994] H. Nishio, Y. Takeshima, N. Narita, H. Yanagawa, Y. Suzuki, Y. Ishikawa, Y. Ishikawa, R. Minami, H. Nakamura et M. Matsuo. *Identification of a novel first exon in the human dystrophin gene and of a new promoter located more than 500 kb upstream of the nearest known promoter*. *J Clin Invest*, vol. 94, no. 3, pages 1037–42, 1994.
- [Norwood 2000] F. L. Norwood, A. J. Sutherland-Smith, N. H. Keep et J. Kendrick-Jones. *The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy*. *Structure*, vol. 8, no. 5, pages 481–91, 2000.
- [Nudel 1989] U Nudel, D Zuk, P Einat, E Zeelon, Z Levy, S Neuman et D Yaffe. *Duchenne muscular dystrophy gene product is not identical in muscle and brain*. *Nature*, vol. 337, no. 6202, pages 76–8, Jan 1989.
- [Nurisso 2012] Alessandra Nurisso, Antoine Daina et Ross C Walker. *A practical introduction to molecular dynamics simulations: applications to homology modeling*. *Methods Mol Biol*, vol. 857, pages 137–73, 2012.
- [Odom 2008] Guy L Odom, Paul Gregorevic, James M Allen, Eric Finn et Jeffrey S Chamberlain. *Microutrophin delivery through rAAV6 increases lifespan and improves muscle function in dystrophic dystrophin/utrophin-deficient mice*. *Mol Ther*, vol. 16, no. 9, pages 1539–45, Sep 2008.
- [Oostenbrink 2004] Chris Oostenbrink, Alessandra Villa, Alan E Mark et Wilfred F van Gunsteren. *A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6*. *J Comput Chem*, vol. 25, no. 13, pages 1656–76, Oct 2004.
- [Parry 2008] David A D Parry, R D Bruce Fraser et John M Squire. *Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure*. *J Struct Biol*, vol. 163, no. 3, pages 258–69, Sep 2008.
- [Partridge 2004] Terence A Partridge. *Stem cell therapies for neuromuscular diseases*. *Acta Neurol Belg*, vol. 104, no. 4, pages 141–7, Dec 2004.
- [Patrinos 2005] George P Patrinos et Anthony J Brookes. *DNA, diseases and databases: disastrously deficient*. *Trends Genet*, vol. 21, no. 6, pages 333–8, Jun 2005.
- [Pestronk 1982] A Pestronk, I M Parhad, D B Drachman et D L Price. *Membrane myopathy: morphological similarities to Duchenne muscular dystrophy*. *Muscle Nerve*, vol. 5, no. 3, pages 209–14, Mar 1982.
- [Phillips 2005] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé et Klaus Schulten. *Scalable molecular dynamics with NAMD*. *J Comput Chem*, vol. 26, no. 16, pages 1781–802, Dec 2005.
- [Pons 1993] F Pons, L V Nicholson, A Robert, T Voit et J J Leger. *Dystrophin and dystrophin-related protein (utrophin) distribution in normal and dystrophin-deficient skeletal muscles*. *Neuromuscul Disord*, vol. 3, no. 5-6, pages 507–14, 1993.

- [Ponting 1996] C. P. Ponting, D. J. Blake, K. E. Davies, J. Kendrick-Jones et S. J. Winder. *ZZ and TAZ: new putative zinc fingers in dystrophin and other proteins*. Trends Biochem Sci, vol. 21, no. 1, pages 11–13, 1996.
- [Price 2007] F D Price, K Kuroda et M A Rudnicki. *Stem cell based therapies to treat muscular dystrophy*. Biochim Biophys Acta, vol. 1772, no. 2, pages 272–83, Feb 2007.
- [Pyrkov 2009] Timothy V Pyrkov, Anton O Chugunov, Nikolay A Krylov, Dmitry E Nolde et Roman G Efremov. *PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes*. Bioinformatics, vol. 25, no. 9, pages 1201–2, May 2009.
- [Reynolds 2008] J. G. Reynolds, S. A. McCalmon, J. A. Donaghey et F. J. Naya. *Deregulated protein kinase A signaling and myospryn expression in muscular dystrophy*. J Biol Chem, vol. 283, no. 13, pages 8070–4, 2008.
- [Rezniczek 2007] G. A. Rezniczek, P. Konieczny, B. Nikolic, S. Reipert, D. Schneller, C. Abrahamsberg, K. E. Davies, S. J. Winder et G. Wiche. *Plectin 1f scaffolding at the sarcolemma of dystrophic (mdx) muscle fibers through multiple interactions with beta-dystroglycan*. J Cell Biol, vol. 176, no. 7, pages 965–77, 2007.
- [Richter 2008] S. Richter, A. Wenzel, M. Stein, R. R. Gabdouliline et R. C. Wade. *webPIPSA: a web server for the comparison of protein interaction properties*. Nucleic Acids Res, vol. 36, no. Web Server issue, pages W276–80, 2008.
- [Rivière 2006] C Rivière, O Danos et A M Douar. *Long-term expression and repeated administration of AAV type 1, 2 and 5 vectors in skeletal muscle of immunocompetent adult mice*. Gene Ther, vol. 13, no. 17, pages 1300–8, Sep 2006.
- [Roberts 1993] R. G. Roberts, A. J. Coffey, M. Bobrow et D. R. Bentley. *Exon structure of the human dystrophin gene*. Genomics, vol. 16, no. 2, pages 536–8, 1993.
- [Roberts 2000] R G Roberts et M Sheng. *Association of dystrophin-related protein 2 (DRP2) with postsynaptic densities in rat brain*. Mol Cell Neurosci, vol. 16, no. 5, pages 674–85, Nov 2000.
- [Roberts 2001] R G Roberts. *Dystrophins and dystrobrevins*. Genome Biol, vol. 2, no. 4, page 3006, 2001.
- [Rouger 2011] Karl Rouger, Thibaut Larcher, Laurence Dubreil, Jack-Yves Deschamps, Caroline Le Guiner, Gregory Jouvion, Bruno Delorme, Blandine Lieubeau, Marine Carlus, Benoît Fornasari, Marine Theret, Priscilla Orlando, Mireille Ledevin, Céline Zuber, Isabelle Leroux, Stéphane Deleau, Lydie Guigand, Isabelle Testault, Elisabeth Le Rumeur, Marc Fiszman et Yan Chérel. *Systemic delivery of allogenic muscle stem cells induces long-term muscle repair and clinical efficacy in duchenne muscular dystrophy dogs*. Am J Pathol, vol. 179, no. 5, pages 2501–18, Nov 2011.
- [Roy 2010] A. Roy, A. Kucukural et Y. Zhang. *I-TASSER: a unified platform for automated protein structure and function prediction*. Nat Protoc, vol. 5, no. 4, pages 725–38, 2010.
- [Rybakova 2000] I. N. Rybakova, J. R. Patel et J. M. Ervasti. *The dystrophin complex forms a mechanically strong link between the sarcolemma and costameric actin*. J Cell Biol, vol. 150, no. 5, pages 1209–14, 2000.
- [Rybakova 2005] Inna N Rybakova et James M Ervasti. *Identification of spectrin-like repeats required for high affinity utrophin-actin interaction*. J Biol Chem, vol. 280, no. 24, pages 23018–23, Jun 2005.
- [Rybakova 2006] I. N. Rybakova, J. L. Humston, K. J. Sonnemann et J. M. Ervasti. *Dystrophin and utrophin bind actin through distinct modes of contact*. J Biol Chem, vol. 281, no. 15, pages 9996–10001, 2006.

- [Saadat 2006] L. Saadat, L. Pittman et N. Menhart. *Structural cooperativity in spectrin type repeats motifs of dystrophin*. Biochim Biophys Acta, vol. 1764, no. 5, pages 943–54, 2006.
- [Sadoulet-Puccio 1997] H. M. Sadoulet-Puccio, M. Rajala et L. M. Kunkel. *Dystrobrevin and dystrophin: an interaction through coiled-coil motifs*. Proc Natl Acad Sci U S A, vol. 94, no. 23, pages 12413–8, 1997.
- [Sarkis 2011] J. Sarkis, J. F. Hubert, B. Legrand, E. Robert, A. Cheron, J. Jardin, E. Hitti, E. Le Rumeur et V. Vie. *Spectrin-like repeats 11-15 of human dystrophin show adaptations to a lipidic environment*. J Biol Chem, 2011.
- [Sarkis 2012] J. Sarkis, V. Vié, S. J. Winder, A. Renault, E. Le Rumeur et J.-F. Hubert. *Resisting sarcolemmal rupture: Dystrophin repeats increase Membrane-Actin stiffness*. FASEB J, vol. in revision, 2012.
- [Schmalbruch 1984] H. Schmalbruch. *Regenerated muscle fibers in Duchenne muscular dystrophy: a serial section study*. Neurology, vol. 34, no. 1, pages 60–5, Jan 1984.
- [Serrano 2006] María Luisa Serrano, Hilda A Pérez et J D Medina. *Structure of C-terminal fragment of merozoite surface protein-1 from Plasmodium vivax determined by homology modeling and molecular dynamics refinement*. Bioorg Med Chem, vol. 14, no. 24, pages 8359–65, Dec 2006.
- [Snow 1978] M H Snow. *An autoradiographic study of satellite cell differentiation into regenerating myotubes following transplantation of muscles in young rats*. Cell Tissue Res, vol. 186, no. 3, pages 535–40, Jan 1978.
- [Stabach 2009] Paul R Stabach, Ivana Simonović, Miranda A Ranieri, Michael S Aboodi, Thomas A Steitz, Miljan Simonović et Jon S Morrow. *The structure of the ankyrin-binding site of beta-spectrin reveals how tandem spectrin-repeats generate unique ligand-binding properties*. Blood, vol. 113, no. 22, pages 5377–84, May 2009.
- [Stamler 2001] J S Stamler et G Meissner. *Physiology of nitric oxide in skeletal muscle*. Physiol Rev, vol. 81, no. 1, pages 209–237, Jan 2001.
- [Stone 2005] M. R. Stone, A. O'Neill, D. Catino et R. J. Bloch. *Specific interaction of the actin-binding domain of dystrophin with intermediate filaments containing keratin 19*. Mol Biol Cell, vol. 16, no. 9, pages 4280–93, 2005.
- [Straub 1997] V. Straub, J. A. Rafael, J. S. Chamberlain et K. P. Campbell. *Animal models for muscular dystrophy show different patterns of sarcolemmal disruption*. J Cell Biol, vol. 139, no. 2, pages 375–85, 1997.
- [Su 2005] Leonard T Su, Kapil Gopal, Zhonglin Wang, Xiaoqing Yin, Anthony Nelson, Benjamin W Kozyak, James M Burkman, Marilyn A Mitchell, David W Low, Charles R Bridges et Hansell H Stedman. *Uniform scale-independent gene transfer to striated muscle after transvenular extravasation of vector*. Circulation, vol. 112, no. 12, pages 1780–8, Sep 2005.
- [Sudol 1996] M. Sudol. *Structure and function of the WW domain*. Prog Biophys Mol Biol, vol. 65, no. 1-2, pages 113–32, 1996.
- [Taylor 2007] P. J. Taylor, S. Maroulis, G. L. Mullan, R. L. Pedersen, A. Baumli, G. Elakis, S. Piras, C. Walsh, B. Prosper-Gutierrez, F. De La Puente-Alonso, C. G. Bell, D. R. Mowat, H. M. Johnston et M. F. Buckley. *Measurement of the clinical utility of a combined mutation detection protocol in carriers of Duchenne and Becker muscular dystrophy*. J Med Genet, vol. 44, no. 6, pages 368–72, 2007.
- [Taylor 2008] Peter John Taylor. *Molecular genetic analysis of a New South Wales muscular dystrophy cohort*. PhD thesis, The University of New South Wales, 2008.

- [Thampi 2009] S. M. Thampi. *Bioinformatics*, 2009.
- [Tinsley 1992] J M Tinsley, D J Blake, A Roche, U Fairbrother, J Riss, B C Byth, A E Knight, J Kendrick-Jones, G K Suthers et D R Love. *Primary structure of dystrophin-related protein*. *Nature*, vol. 360, no. 6404, pages 591–3, Dec 1992.
- [Tinsley 1996] J M Tinsley, A C Potter, S R Phelps, R Fisher, J I Trickett et K E Davies. *Amelioration of the dystrophic phenotype of mdx mice using a truncated utrophin transgene*. *Nature*, vol. 384, no. 6607, pages 349–53, Nov 1996.
- [Tinsley 1998] J Tinsley, N Deconinck, R Fisher, D Kahn, S Phelps, J M Gillis et K Davies. *Expression of full-length utrophin prevents muscular dystrophy in mdx mice*. *Nat Med*, vol. 4, no. 12, pages 1441–4, Dec 1998.
- [Tinsley 2011] Jonathon M Tinsley, Rebecca J Fairclough, Richard Storer, Fraser J Wilkes, Allyson C Potter, Sarah E Squire, Dave S Powell, Anna Cozzoli, Roberta F Capogrosso, Adam Lambert, Francis X Wilson, Stephen P Wren, Annamaria De Luca et Kay E Davies. *Daily treatment with SMTC1100, a novel small molecule utrophin upregulator, dramatically reduces the dystrophic symptoms in the mdx mouse*. *PLoS One*, vol. 6, no. 5, page e19189, 2011.
- [Tomé 1994] F M Tomé, K Matsumura, M Chevally, K P Campbell et M Fardeau. *Expression of dystrophin-associated glycoproteins during human fetal muscle development: a preliminary immunocytochemical study*. *Neuromuscul Disord*, vol. 4, no. 4, pages 343–8, Jul 1994.
- [Torrente 2007] Y Torrente, M Belicchi, C Marchesi, G Dantona, F Cogiamanian, F Pisati, M Gavina, R Giordano, R Tonlorenzi, G Fagiolari, C Lamperti, L Porretti, R Lopa, M Sampaolesi, L Vicentini, N Grimoldi, F Tiberio, V Songa, P Baratta, A Prella, L Forzenigo, M Guglieri, O Pansarasa, C Rinaldi, V Mouly, G S Butler-Browne, G P Comi, P Biondetti, M Moggio, S M Gaini, N Stocchetti, A Priori, M G D'Angelo, A Turconi, R Bottinelli, G Cossu, P Rebullà et N Bresolin. *Autologous transplantation of muscle-derived CD133+ stem cells in Duchenne muscle patients*. *Cell Transplant*, vol. 16, no. 6, pages 563–77, 2007.
- [Tuffery-Giraud 2009] S. Tuffery-Giraud, 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 et M. Claustres. *Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase*. *Hum Mutat*, vol. 30, no. 6, pages 934–45, 2009.
- [Tyler 2003] K L Tyler. *Origins and early descriptions of "Duchenne muscular dystrophy"*. *Muscle Nerve*, vol. 28, no. 4, pages 402–422, Oct 2003.
- [Venclovas 2012] Ceslovas Venclovas. *Methods for sequence-structure alignment*. *Methods Mol Biol*, vol. 857, pages 55–82, 2012.
- [Venselaar 2010] Hanka Venselaar, Robbie P Joosten, Bas Vrolijk, Coos A B Baakman, Maarten L Hekkelman, Elmar Krieger et Gert Vriend. *Homology modelling and spectroscopy, a never-ending love story*. *Eur Biophys J*, vol. 39, no. 4, pages 551–63, Mar 2010.
- [Vie 2010] V. Vie, S. Legardinier, L. Chieze, O. Le Bihan, Y. Qin, J. Sarkis, J. F. Hubert, A. Renault, B. Desbat et E. Le Rumeur. *Specific anchoring modes of two distinct dystrophin rod sub-domains interacting in phospholipid Langmuir films studied by atomic force microscopy and PM-IRRAS*. *Biochim Biophys Acta*, vol. 1798, no. 8, pages 1503–11, 2010.

- [Wang 2007] Zejing Wang, Christian S Kuhr, James M Allen, Michael Blankinship, Paul Gregorevic, Jeffrey S Chamberlain, Stephen J Tapscott et Rainer Storb. *Sustained AAV-mediated dystrophin expression in a canine model of Duchenne muscular dystrophy with a brief course of immunosuppression*. *Mol Ther*, vol. 15, no. 6, pages 1160–6, Jun 2007.
- [Wang 2010] Zejing Wang, Rainer Storb, Donghoon Lee, Martin J Kushmerick, Baocheng Chu, Carolina Berger, Andrea Arnett, James Allen, Jeffrey S Chamberlain, Stanley R Riddell et Stephen J Tapscott. *Immune responses to AAV in canine muscle monitored by cellular assays and noninvasive imaging*. *Mol Ther*, vol. 18, no. 3, pages 617–24, Mar 2010.
- [Welch 2007] Ellen M Welch, Elisabeth R Barton, Jin Zhuo, Yuki Tomizawa, Westley J Friesen, Panayiota Trifillis, Sergey Paushkin, Meenal Patel, Christopher R Trotta, Seongwoo Hwang, Richard G Wilde, Gary Karp, James Takasugi, Guangming Chen, Stephen Jones, Hongyu Ren, Young-Choon Moon, Donald Corson, Anthony A Turpoff, Jeffrey A Campbell, M Morgan Conn, Atiyya Khan, Neil G Almstead, Jean Hedrick, Anna Mollin, Nicole Risher, Marla Weetall, Shirley Yeh, Arthur A Branstrom, Joseph M Colacino, John Babiak, William D Ju, Samit Hirawat, Valerie J Northcutt, Langdon L Miller, Phyllis Spatrick, Feng He, Masataka Kawana, Hui-sheng Feng, Allan Jacobson, Stuart W Peltz et H Lee Sweeney. *PTC124 targets genetic disorders caused by nonsense mutations*. *Nature*, vol. 447, no. 7140, pages 87–91, May 2007.
- [Whewey 2003] Joanna M Whewey et Roland G Roberts. *The dystrophin lymphocyte promoter revisited: 4.5-megabase intron, or artifact?* *Neuromuscul Disord*, vol. 13, no. 1, pages 17–20, Jan 2003.
- [White 2006] S. J. White et J. T. den Dunnen. *Copy number variation in the genome; the human DMD gene as an example*. *Cytogenet Genome Res*, vol. 115, no. 3-4, pages 240–6, 2006.
- [Wiederstein 2007] M. Wiederstein et M. J. Sippl. *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins*. *Nucleic Acids Res*, vol. 35, no. Web Server issue, pages W407–10, 2007.
- [Willig 1993] T N Willig, L Carlier, M Legrand, H Rivière et J Navarro. *Nutritional assessment in Duchenne muscular dystrophy*. *Dev Med Child Neurol*, vol. 35, no. 12, pages 1074–82, Dec 1993.
- [Winder 1995a] S. J. Winder, T. J. Gibson et J. Kendrick-Jones. *Dystrophin and utrophin: the missing links!* *FEBS Lett*, vol. 369, no. 1, pages 27–33, 1995.
- [Winder 1995b] S J Winder, L Hemmings, S K Maciver, S J Bolton, J M Tinsley, K E Davies, D R Critchley et J Kendrick-Jones. *Utrophin actin binding domain: analysis of actin binding and cellular targeting*. *J Cell Sci*, vol. 108 ( Pt 1), pages 63–71, Jan 1995.
- [Winder 2001] S J Winder. *The complexities of dystroglycan*. *Trends Biochem Sci*, vol. 26, no. 2, pages 118–124, Feb 2001.
- [Wyss 2003] Pierre C Wyss, Paul Gerber, Peter G Hartman, Christian Hubschwerlen, Hans Locher, Hans-Peter Marty et Martin Stahl. *Novel dihydrofolate reductase inhibitors. Structure-based versus diversity-based library design and high-throughput synthesis and screening*. *J Med Chem*, vol. 46, no. 12, pages 2304–12, Jun 2003.
- [Yamashita 2010] K. Yamashita, A. Suzuki, Y. Satoh, M. Ide, Y. Amano, M. Masuda-Hirata, Y. K. Hayashi, K. Hamada, K. Ogata et S. Ohno. *The 8th and 9th tandem spectrin-*

- like repeats of utrophin cooperatively form a functional unit to interact with polarity-regulating kinase PAR-1b.* Biochem Biophys Res Commun, vol. 391, no. 1, pages 812–7, 2010.
- [Yin 2008] HaiFang Yin, Hong M Moulton, Yiqi Seow, Corinne Boyd, Jordan Boutilier, Patrick Iverson et Matthew J A Wood. *Cell-penetrating peptide-conjugated antisense oligonucleotides restore systemic muscle and cardiac dystrophin expression and function.* Hum Mol Genet, vol. 17, no. 24, pages 3909–18, Dec 2008.
- [Yokota 2004] Ryoji Yokota, Manabu Shirotani, Ichiro Kouchi, Taku Hirai, Noritsugu Uemori, Yoshio Ohta, Yoshiyuki Mitsui et Ryuichi Hattori. *Subclinical Becker's muscular dystrophy presenting with severe heart failure.* Intern Med, vol. 43, no. 3, pages 204–8, Mar 2004.
- [Yokota 2007] T Yokota, W Duddy et T Partridge. *Optimizing exon skipping therapies for DMD.* Acta Myol, vol. 26, no. 3, pages 179–84, Dec 2007.
- [Yokota 2009] Toshifumi Yokota, Qi-Long Lu, Terence Partridge, Masanori Kobayashi, Akinori Nakamura, Shiñichi Takeda et Eric Hoffman. *Efficacy of systemic morpholino exon-skipping in Duchenne dystrophy dogs.* Ann Neurol, vol. 65, no. 6, pages 667–76, Jun 2009.
- [Zalaudek 1999] I Zalaudek, R M Bonelli, P Költringer, F Reisecker et K Wagner. *Early diagnosis in Duchenne muscular dystrophy.* Lancet, vol. 353, no. 9168, page 1975, Jun 1999.
- [Zaman 2004] M. H. Zaman et M. R. Kaazempur-Mofrad. *How flexible is alpha-actinin's rod domain?* Mech Chem Biosyst, vol. 1, no. 4, pages 291–302, 2004.
- [Zhang 2005] Yan Zhang, Yuk Y Sham, Ramkumar Rajamani, Jiali Gao et Philip S Portoghese. *Homology modeling and molecular dynamics simulations of the mu opioid receptor in a membrane-aqueous system.* Chembiochem, vol. 6, no. 5, pages 853–9, May 2005.
- [Zhang 2007] Y. Zhang. *Template-based modeling and free modeling by I-TASSER in CASP7.* Proteins, vol. 69 Suppl 8, pages 108–17, 2007.
- [Zhang 2008] Y. Zhang. *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, vol. 9, page 40, 2008.
- [Zhu 2012] Xiao Zhu, Pedro E. M. Lopes et Alexander D. MacKerell. *Recent developments and applications of the CHARMM force fields.* Wiley Interdisciplinary Reviews: Computational Molecular Science, vol. 2, no. 1, pages 167–185, 2012.

## **Annexe A**

### **Exemple de résultats I-TASSER**

Extrait de résultats générés par I-TASSER  
Exemple du tandem de répétitions R21-22

Top 5 Models predicted by I-TASSER



Estimated accuracy of Model1: 0.77±0.10 (TM-score) 4.7±3.1Å (RMSD) ([Read more about C-score of generated models](#))

Top 10 templates used by I-TASSER

Rank	PDB Hit	Iden1	Iden2	Cov.	Norm. Z-score	Download
1	<a href="#">1cunA</a>	0.14	0.16	0.94	2.09	<a href="#">Download</a>
2	<a href="#">1s35A</a>	0.12	0.18	0.93	2.94	<a href="#">Download</a>
3	<a href="#">1cunA</a>	0.13	0.16	0.94	3.51	<a href="#">Download</a>
4	<a href="#">1s35A</a>	0.12	0.18	0.94	2.54	<a href="#">Download</a>
5	<a href="#">1cunA</a>	0.14	0.16	0.94	2.46	<a href="#">Download</a>
6	<a href="#">1cunA</a>	0.16	0.16	0.94	2.38	<a href="#">Download</a>
7	<a href="#">1sjjA</a>	0.15	0.18	0.96	2.02	<a href="#">Download</a>
8	<a href="#">1owaA</a>	0.17	0.15	0.89	2.19	<a href="#">Download</a>
9	<a href="#">1cunA</a>	0.13	0.16	0.94	3.81	<a href="#">Download</a>
10	<a href="#">1u4qA</a>	0.23	0.27	0.93	4.35	<a href="#">Download</a>

(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. ([more about the colors used](#))

(b) Rank of templates represents the top ten threading templates used by I-TASSER.

(c) Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.

(d) Iden2 is the percentage sequence identity of the whole template chains with query sequence.

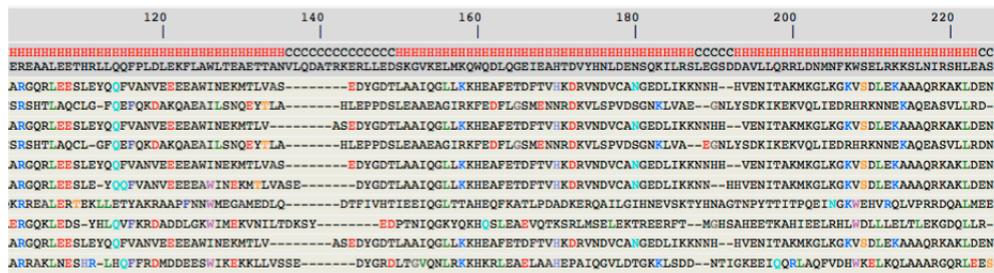
(e) Cov. represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.

(f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

(g) Download Align. provides the 3D structure of the aligned regions of the threading templates.

(h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:

1: MUSTER 2: HHSEARCH 3: SP3 4: PROSPECT2 5: PPA-I 6: HHSEARCH I 7: COMA 8: FUGUE 9: SPARKS 10: SAM T99



- 1cun : Crystal Structure Of Repeats 16 And 17 Of Chicken Brain Alpha Spectrin
- 1s35 : Crystal Structure of Repeats 8 and 9 of Human Erythroid Spectrin
- 1sjj : Cryo-EM Structure of Chicken Gizzard Smooth Muscle alpha-Actinin
- 1owa : Solution Structural Studies on Human Erythrocyte Alpha Spectrin N Terminal Tetramerization Domain
- 1u4q : Crystal Structure of Repeats 15, 16 and 17 of Chicken Brain Alpha Spectrin

## Annexe B

### Alignement des séquences de R16-21 et des quatre protéines tronquées

	10	20	30	40	50	60
del45_47	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSLQQSSGRID					
R16_21	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSLQQSSGRID					
del45_48	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSLQQSSGRID					
del45_49	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSLQQSSGRID					
del45_51	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSLQQSSGRID					
heptad	...a..d...a..d...a..d...a			d...a..d...a..d...a..d...a		
repeat	R16-HA			R16-HB		

	70	80	90	100	110	120
del45_47	IIHSKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDI					
R16_21	IIHSKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDI					
del45_48	IIHSKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDI					
del45_49	IIHSKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDI					
del45_51	IIHSKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDI					
heptad	.d...a..d		d...a..d...a..d...a..d...a..d...a			
repeat	R16-HC & R17-HA					

	130	140	150	160	170	180
del45_47	KIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLK-----					
R16_21	KIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQS					
del45_48	KIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLK-----					
del45_49	KIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLK-----					
del45_51	KIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLK-----					
heptad	..d...a..d...a		d...a..d...a..d...a..d...a..d...a..d			
repeat	R17-HB					

	190	200	210	220	230	240
del45_47	-----					
R16_21	SKTDASILQEKLGSNLRWQEVCKQLSDRKKRLEEQKNILSEFQRDLNEFVLWLEEDNI					
del45_48	-----					
del45_49	-----					
del45_51	-----					
heptad	d...a..d...a..d...a..d...a..d...a..d...a..d...a..d...a..d...					
	R17-HC & R18-HA					

	250	260	270	280	290	300
del45_47	-----					
R16_21	ASIPLEPGKEQQLKEKLEQVKLLVEELPLRQGILKQLNETGGPVLVSAPISPEEQDKLEN					
del45_48	-----					
del45_49	-----					
del45_51	-----					
heptad repeat	a	d...a..d...a..d...a..d...a..d...a..d				d..
		R18-HB				

	310	320	330	340	350	360
del45_47	-----VSRALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEI					
R16_21	KLKQTNLQWIKVSRALPEKQGEIEAQIKDLGQLEKKLEDLEEQLNHLLLWLSPIRNQLEI					
del45_48	-----					
del45_49	-----					
del45_51	-----					
heptad repeat	.a..d...a..d...a..d...a..d...a..d...a..d...a..d...a					
		R18-HC & R19-HA				

	370	380	390	400	410	420
del45_47	YNQPNQEGPFDVQETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAV					
R16_21	YNQPNQEGPFDVQETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAV					
del45_48	-----ETEIAVQAKQPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKAV					
del45_49	-----RKLEDLSSEWKAV					
del45_51	-----					
heptad repeat	a..d...a..d...a..d...a..d...a..d				d...a..d...a..d	
		R19HB			R19-HC	

	430	440	450	460	470	480
del45_47	NRLLQELRAKQPD LAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALA					
R16_21	NRLLQELRAKQPD LAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALA					
del45_48	NRLLQELRAKQPD LAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALA					
del45_49	NRLLQELRAKQPD LAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALA					
del45_51	-----					
heptad repeat	...a..d...a					a.
		H3				

	490	500	510	520	530	540
del45_47	DFNRAWTE	LDWLSLLDQVIKSQRVMVGDLE	EDINEMI	IKQKATM	QDLEQRRP	QLEELITA
R16_21	DFNRAWTE	LDWLSLLDQVIKSQRVMVGDLE	EDINEMI	IKQKATM	QDLEQRRP	QLEELITA
del45_48	DFNRAWTE	LDWLSLLDQVIKSQRVMVGDLE	EDINEMI	IKQKATM	QDLEQRRP	QLEELITA
del45_49	DFNRAWTE	LDWLSLLDQVIKSQRVMVGDLE	EDINEMI	IKQKATM	QDLEQRRP	QLEELITA
del45_51	-----					ATMQDLEQRRPQLEELITA
heptad	.d...a..d...a..d...a	d...a..d...a..d...a..d...a..d...				
repeat	R20-HA		R20-HB			

	550	560	570	580	590	600
del45_47	AQNLKNKTSNQEARTIITDRIERI	QIQNWDEVQEHLQNR	RQQLNEM	LKDSTQW	LEAKEEEAE	
R16_21	AQNLKNKTSNQEARTIITDRIERI	QIQNWDEVQEHLQNR	RQQLNEM	LKDSTQW	LEAKEEEAE	
del45_48	AQNLKNKTSNQEARTIITDRIERI	QIQNWDEVQEHLQNR	RQQLNEM	LKDSTQW	LEAKEEEAE	
del45_49	AQNLKNKTSNQEARTIITDRIERI	QIQNWDEVQEHLQNR	RQQLNEM	LKDSTQW	LEAKEEEAE	
del45_51	AQNLKNKTSNQEARTIITDRIERI	QIQNWDEVQEHLQNR	RQQLNEM	LKDSTQW	LEAKEEEAE	
heptad	a..d	d...a..d...a..d...a..d...a..d...a..d..				
repeat	R20-HC & R21-HA					

	610	620	630	640	650	660
del45_47	QVLGQARAKLESWKEG	PYTVDAIQKKITETK	QLAKDLRQWQTNVDV	ANDLALKLLRD	YSA	
R16_21	QVLGQARAKLESWKEG	PYTVDAIQKKITETK	QLAKDLRQWQTNVDV	ANDLALKLLRD	YSA	
del45_48	QVLGQARAKLESWKEG	PYTVDAIQKKITETK	QLAKDLRQWQTNVDV	ANDLALKLLRD	YSA	
del45_49	QVLGQARAKLESWKEG	PYTVDAIQKKITETK	QLAKDLRQWQTNVDV	ANDLALKLLRD	YSA	
del45_51	QVLGQARAKLESWKEG	PYTVDAIQKKITETK	QLAKDLRQWQTNVDV	ANDLALKLLRD	YSA	
heptad	..a..d...a	d...a..d...a..d...a..d...a..d...a..d..				
	R21-HB					

	670	680	690
del45_47	DDTRKVHMITENINAS	WRSIHKRVSEREAA	EET
R16_21	DDTRKVHMITENINAS	WRSIHKRVSEREAA	EET
del45_48	DDTRKVHMITENINAS	WRSIHKRVSEREAA	EET
del45_49	DDTRKVHMITENINAS	WRSIHKRVSEREAA	EET
del45_51	DDTRKVHMITENINAS	WRSIHKRVSEREAA	EET
heptad	d...a..d...a..d...a..d...a		
	R21-HC		

## Annexe C

Article paru dans Plos One

# Computational Study of the Human Dystrophin Repeats: Interaction Properties and Molecular Dynamics

Baptiste Legrand<sup>1,2,4✉</sup>, Emmanuel Giudice<sup>1,3,4✉</sup>, Aurélie Nicolas<sup>1,2,4</sup>, Olivier Delalande<sup>1,2,4</sup>, Elisabeth Le Rumeur<sup>1,2,4\*</sup>

**1** Université de Rennes 1, Rennes, France, **2** Equipe RMN-ILP, Faculté de médecine, UMR CNRS 6026, Rennes, France, **3** Equipe SDM, UMR CNRS 6026, Rennes, France, **4** Université Européenne de Bretagne, Rennes, France

## Abstract

Dystrophin is a large protein involved in the rare genetic disease Duchenne muscular dystrophy (DMD). It functions as a mechanical linker between the cytoskeleton and the sarcolemma, and is able to resist shear stresses during muscle activity. In all, 75% of the dystrophin molecule consists of a large central rod domain made up of 24 repeat units that share high structural homology with spectrin-like repeats. However, in the absence of any high-resolution structure of these repeats, the molecular basis of dystrophin central domain's functions has not yet been deciphered. In this context, we have performed a computational study of the whole dystrophin central rod domain based on the rational homology modeling of successive and overlapping tandem repeats and the analysis of their surface properties. Each tandem repeat has very specific surface properties that make it unique. However, the repeats share enough electrostatic-surface similarities to be grouped into four separate clusters. Molecular dynamics simulations of four representative tandem repeats reveal specific flexibility or bending properties depending on the repeat sequence. We thus suggest that the dystrophin central rod domain is constituted of seven biologically relevant sub-domains. Our results provide evidence for the role of the dystrophin central rod domain as a scaffold platform with a wide range of surface features and biophysical properties allowing it to interact with its various known partners such as proteins and membrane lipids. This new integrative view is strongly supported by the previous experimental works that investigated the isolated domains and the observed heterogeneity of the severity of dystrophin related pathologies, especially Becker muscular dystrophy.

**Citation:** Legrand B, Giudice E, Nicolas A, Delalande O, Le Rumeur E (2011) Computational Study of the Human Dystrophin Repeats: Interaction Properties and Molecular Dynamics. PLoS ONE 6(8): e23819. doi:10.1371/journal.pone.0023819

**Editor:** Ozlem Keskin, Koç University, Turkey

**Received:** May 6, 2011; **Accepted:** July 26, 2011; **Published:** August 25, 2011

**Copyright:** © 2011 Legrand et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by "Association Française contre les Myopathies", Rennes Métropole, French Ministry of Research and CNRS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: elisabeth.lerumeur@univ-rennes1.fr

✉ These authors contributed equally to this work.

✉ Current address: Laboratoire de chimie physique macromoléculaire, UMR CNRS-INPL 7568, Nancy, France

## Introduction

The stability of muscle cells depends on the ability of cytoskeletal proteins to dynamically resist the mechanical shear stresses which occur during muscle activity. Dystrophin is one of these skeletal muscle cytoskeletal proteins [1–2] and is part of the large dystrophin-glycoprotein sarcolemmal complex [3–4]. Its complete genetic deficit in Duchenne muscular dystrophy (DMD) [5] leads to frequent sarcolemma ruptures followed by cell degeneration. Therefore, the current hypothesis is that dystrophin protects muscle cell membranes from rupture [6].

Dystrophin is a huge scaffolding protein of 427 kDa, made up of four major domains [4,7]. The two N-terminal calponin homology sub-domains constitute an actin-binding domain (#1). After a first hinge is the large central rod domain (#2), composed of 24 spectrin-like repeats interrupted by two more hinges. This domain interacts with membrane phospholipids and with a number of cytosolic proteins such as filamentous actin (F-actin), n-nitric oxide synthase (nNOS) and microtubules. After a fourth hinge, there is the cysteine-rich domain (#3), which anchors dystrophin to the intrinsic membrane protein  $\beta$ -dystroglycan. Finally, the coiled-coil structured

C-terminal domain (#4) interacts with the cytoplasmic proteins syntrophin and dystrobrevin. Through these numerous interactions, dystrophin covers the sub-sarcolemma surface with a dense network and may resist elongation during muscle contraction [4,7].

The dystrophin central rod domain represents about 75% of the entire protein and this conserved structural domain makes it a member of the spectrin-like protein family, which also includes utrophin, spectrin and  $\alpha$ -actinin [8]. The sequence similarity between members of this family is rather low, and their main common feature is the presence of numerous repeated sequences of approximately 100–110 residues called spectrin-like repeats. The structural basis of these repeats is the presence of heptad patterns, i.e., periodic patterns of seven hydrophobic and hydrophilic/charged residues usually denoted by the letters "a" through "g". The residues in positions "a" and "d" are hydrophobic and ensure folding into triple alpha-helical anti-parallel coiled-coils [9–10]. The residues in the other positions are usually hydrophilic and/or charged. In spectrin and  $\alpha$ -actinin, the contiguous repeats are connected by helical linkers that ensure continuity between the last helix of the first repeat and the first helix of the next repeat.

Although we and other groups have tried to solve the three-dimensional (3D) structures of different parts of the dystrophin central domain by X-ray crystallography and NMR, no atomic structures are yet available. At this time, the 3D structures of one isolated spectrin repeat [11], eight multi-repeat spectrin domains [12–19] and the  $\alpha$ -actinin four-repeat domain [20–21] are the only structures that have been solved by X-ray crystallography. Only one 3D spectrin repeat structure has been solved by NMR [22]. The structural study of both spectrin and  $\alpha$ -actinin may have been facilitated because they naturally exist as oligomers [8,23]. In the crystals, spectrin and  $\alpha$ -actinin repeats always appear as dimers, but dystrophin is not expected to. In consequence, the only structural information available for the dystrophin repeats has been obtained by circular dichroism and tryptophan fluorescence [24–29].

Because of the lack of experimental 3D structural data for dystrophin rod domain repeats, it is necessary to use comparative modeling and structural prediction to study their molecular properties. The utility of such approaches in designing experiments and interpreting experimental results is now widely recognized [30–31]. In this context, the goal of the present work is to highlight the structural differences among dystrophin repeats in complementarity of their known biochemical properties. Through the combination of homology modeling and the comparison of surface properties and molecular dynamics simulations, we have proposed a molecular description of the whole dystrophin central rod domain. We have shown that, despite their similar helical coiled-coil structures, dystrophin repeats display a huge diversity of surface electrostatics and hydrophobicity, as well as varying flexibility. The succession of repeats with specific properties and the presence of flexible junctions delineate seven independent structural regions, each of which may play a specific role in dystrophin activity. Our results provide evidence for the role of the dystrophin central rod domain as a scaffold platform interacting with various partners, such as proteins and membrane lipids, through a wide range of surface features and biophysical properties.

## Methods

### Sequence alignment

The sequence of human dystrophin was retrieved from NCBI Dp427m and is consistent with the cDNA sequence of the plasmid pTG11025 harboring the cDNA for the Dp427m muscle isoform of human dystrophin (National Center for Biotechnology Information Nucleotide Data Base NM-004006, provided by S. Braun Transgene, France) used in previous experimental work. Two sequence alignments of the dystrophin repeats are still commonly used, even though they differ somewhat from each other. The first appeared very early after the discovery of dystrophin [32], and the second includes utrophin sequences [33]. It appears that for 14 of the 24 repeats, the repeat starting point shifts by one residue between the two alignments. To optimize the first alignment, Koenig and Kunkel deleted a few residues and introduced some gaps [32]. In the alignment by Winder, there is no deletion. For this reason, we chose to use the alignment by Winder to define the boundaries of the different repeats [33], as shown in Figure S1.

### Secondary and 3D structure prediction

The secondary structure was predicted using PSIPRED [34–35]. For 3D structure prediction, I-TASSER combines various techniques such as threading, *ab initio* modeling and structure refinement approaches [36–37]. The sequences of two successive tandem repeats were submitted with an overlap of one repeat for

the next submission, i.e., first the repeat 1–2, then the repeat 2–3, leading to a total of 21 models. The two tandem repeats that would include known hinges, R3-4 and R19-20, were omitted (Fig. 1A). 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. These representative structures were analyzed on a graphical display with PyMOL [38] and their quality assessed using PROCHECK, ProSA-web [39–40] and Verify3D [41–42].

### Surface-property comparisons: hydrophobicity

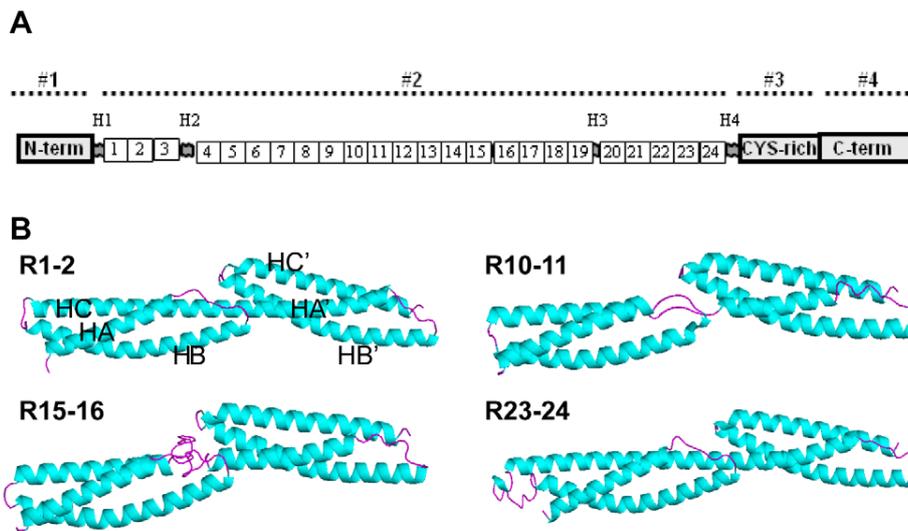
Surface hydrophobicity was obtained using PLATINUM, which is designed to calculate match or mismatch in receptor-ligand complexes [43–44]. This program allows the calculation and visualization of the molecular hydrophobic/hydrophilic surface properties using the concept of molecular hydrophobicity potential (MHP). The PLATINUM algorithm also provides the total hydrophobic and hydrophilic surfaces for a molecule. All the models were submitted, and the PLATINUM web site provided the calculated maps, which were then visualized using PyMOL.

### Surface-property comparisons: electrostatics

The web-based version of PIPSA (Protein Interaction Property Similarity Analysis) [45–48] was used to compare the tandem repeats with respect to their electrostatic potentials. PIPSA quantifies the similarity in the surface properties of homologous proteins and is particularly useful for comparing the surface properties of the dystrophin tandem repeats. The previously fitted models were submitted to the PIPSA server [48], and we chose the Adaptive Poisson-Boltzmann Solver (APBS) software to calculate the electrostatic potentials [49]. A low concentration (<10 mM) of monovalent ion tends to accentuate differences between the electrostatic patches, while at cellular concentration (150 mM), the contrast between the different parts of the electrostatic surface is strongly attenuated. For this reason, we chose to present maps calculated for an intermediate ionic strength of 50 mM. Similarity indexes (SI) for each pair of surface electrostatic potentials were calculated [45] and converted into a distance matrix ranging from 0.5 to 1.5 [46–47]. The distance matrix was subsequently subjected to a clustering procedure, and the corresponding dendrogram was transformed into an electrostatic similarity tree using the neighboring-joining algorithm by PHYLIP [50]. Visualization of the electrostatic potentials on the molecular surfaces of the repeats was produced using PyMOL and the APBS algorithm for consistency.

### Molecular dynamics relaxation

To simulate our systems of tandem repeats, water and ions, we used the program NAMD 2.7b2 [51] and the CHARMM27 force field [52–55]. The initial models of the dystrophin tandem repeats were oriented along the z axis and then solvated in rectangular water boxes generated using the *Solvate* plugin of VMD [56]. We thus ensured that there was a 30 Å thick layer of TIP3P water in the x and y directions and a 15 Å thick layer in the z direction. Subsequently, the VMD plug-in *Autoionize* was used to place ions randomly to neutralize the system while maintaining a 150 mM NaCl concentration. To adjust the position of the solvent (water and ions) around the molecules, each system was energy minimized for 10000 steps using the conjugate gradient method while restraining the solute atoms with a 25 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic restraint. The box size was chosen to be big enough to prevent any bias of the Periodic Boundary Conditions on the



**Figure 1. Dystrophin representation and 3D models of four representative tandem repeats.** (A) Shown are the N-terminal (N-term), 24 spectrin-like repeats, cysteine-rich, C-terminal (C-term) domains and the four hinges (H1 to H4). The succession of the four main domains 1 to 4 is indicated above the diagram by a dotted line. (B) Four representative tandem-repeat models are shown as C $\alpha$  backbone traces. Helical segments are colored in blue and loops are in pink. The N-terminal is on the left, the C-terminal is on the right, and the helix A of the N-terminal repeat of each tandem is at the front of the image. As a reminder, the nomenclature of the helices is given for the R1-2 model. HA, HB and HC are the helices of the first repeat, and HA', HB' and HC' are the helices of the second repeat of the tandem.  
doi:10.1371/journal.pone.0023819.g001

simulations. The simulated system dimensions are given in Table S2.

The entire system (solvent + solute) was then subjected to another 10000 steps of energy minimization to relieve any major stresses, followed by a slow heating to 310 K at constant volume over a period of 50 ps. The production phase was performed for 31 ns under periodic boundary conditions with a 2 fs time step using the SHAKE algorithm. Van der Waals interactions were computed using a cut-off distance of 12 Å with a switching function starting at 10 Å, while long-range electrostatic forces were calculated using the particle-mesh Ewald method with a grid density of 1 Å<sup>-3</sup>. To further reduce the cost of computing full electrostatics, a multiple-time-stepping procedure was employed to calculate long-range electrostatics every 4 fs. Berendsen baths were used to maintain the system temperature and pressure at 310K and 1 atm, respectively.

The post-processing analysis of the MD trajectories was performed with VMD 1.8.7 [56] and Ptraj using the last 20 ns of simulation. To extract representative structures, the coordinate frames from the trajectory were clustered using the K-means algorithm. After testing different values, we chose to split the trajectory into two clusters using the pairwise RMSD between frames as a metric to compare the C $\alpha$  atoms of the protein [57]. The Atomic coordinates of the snapshots closest to the center of the resulting clusters (C1 and C2) are available as pdb files in MD-Clusters S1. The results were verified by repeating the simulation for 20 ns using an identical protocol, the same initial model but a different initial velocity distribution.

## Results

### Sequence analysis and secondary structure prediction

Despite a low overall similarity, the sequence alignment showed that most of the residues in the (a) and (d) positions were well conserved, while most of the residues in the other positions were not (Fig. S1A). In all but seven cases, the heptad pattern was

maintained through the linker (Fig. S1B). In the R4-5 and R10-11 linkers, one residue was missing and in the R1-2 and R13-14 linkers, two residues were missing. In contrast, insertions of 2, 20 and 7 residues were present in R5-6, R15-16 and R18-19, respectively (Fig. S1B). 79 to 91% of the total structure was predicted to form  $\alpha$ -helices by PSIPRED (not shown). This result was in agreement with the assumption that the dystrophin repeats are essentially folded in a triple  $\alpha$ -helical coiled-coil, and also with experimental data obtained on *in vitro* produced repeats of dystrophin [27–29]. It is also worth noting that a decreased tendency to form a helix was predicted for the center of helix B of each repeat.

### Structure models of tandem spectrin-like repeats of dystrophin

Not surprisingly, the I-TASSER threading procedure identified spectrin repeats as the best templates, specifically 1U4Q (chicken-brain  $\alpha$ -spectrin repeat R15-17) [14], 1S35 (erythroid  $\beta$ -spectrin R8-9) [13] and 3EDV ( $\beta$ 2-spectrin repeat R14-16) [15]. As expected for spectrin-like repeats, the identity score was 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 0.76, which indicated that all models had correct folds (Table S1). The models were further assessed with VERIFY3D [42], ProSA [39–40] and PROCHECK [58], with results indicative of high-quality models (Annex S1). All the models are available as pdb files in Models S1.

The models of the tandem repeats all appeared as elongated triple helical coiled-coils of roughly 100×50 Å dimensions (Fig. 1B, Fig. S2), which is in accord with the structural templates. Each repeat consisted of three helices and two loops. The helices were not straight but curved gently to form a left-handed super coil. Compared to the available spectrin-like repeat structures, the canonical kink in the center of helix B of each repeat was observed in all repeats, except for R9 and R14 in the R9-R10 and the R14-15 tandem models, respectively.

A long helical linker between the two consecutive repeats was found in 17 of the 21 tandem models. For R5-6, R10-11, R13-14 and R15-16, the linker was not helical but was a short loop (R5-6, R10-11), a long loop (R15-16) or a break (R13-14). In R5-6, R10-11 and R13-14, the absence of the helical linker may have been due to the presence of one or two proline residues which might impair the helical folding of the linker. In R15-16, a 20-residue insertion between the two repeats has been previously suggested to form a small loop [59], as it appears in our model.

The C $\alpha$ -atoms RMSD of our models ranged from 0.250 to 4.241 Å. These values were comparable to those calculated between the various spectrin repeat structures, which ranged from 0.761 to 4.413 Å for the template structures (1U4Q, 1S35 and 3EDV) used in our modeling (Fig. S3).

### Molecular descriptors

As is often the case with cytoplasmic proteins, the overall tandem repeat surfaces were mostly hydrophilic with an average of  $66.4 \pm 3.5\%$  hydrophilic surfaces (Fig. 2A, B). R13-14 and R14-15 constituted a central region that was highly hydrophilic compared to other regions (by less than one standard deviation from the mean), while R18-19 was highly hydrophobic (more than one standard deviation from the mean) (Fig 2B). We were also able to detect numerous hydrophobic patches dispersed on the surface and small hydrophobic grooves (R2-3, R5-6 and R10-11) that might constitute interaction sites.

The electrostatic surfaces of the tandem repeats were clearly dissimilar, and they showed large positive and negative potential patches (Fig. 3). Such patches often appeared to extend over more than one repeat surface (R5-6 for example), while sometimes the two repeats in a tandem exhibited opposite electrostatic properties (R7-8 for example). To further quantify the surface electrostatic potential similarity of the tandem repeats, we analyzed our models with PIPSA [48]. The resulting dendrogram is divided into four clusters (Fig. 3A). Globally, the ratio of negative/positive potential surfaces decreased from cluster 1 to cluster 4 (Fig. 3B). In cluster 1, the models exhibited large strongly negative patches extending over all tandem repeats. This underlined three strongly negative regions in the dystrophin rod domain, constituted by R1-2, R8-10 and R18-19. The tandem repeat surfaces in cluster 2 had small charged patches with numerous negative and few positive moieties. Cluster 3 was constituted by tandem repeats with large and strongly positive patches in the N-terminal repeat of the tandem and negative patches in the C-terminal end. The electrostatic surfaces in the cluster 4 were comparable to those in cluster 3 but with the large positive moieties rather in the C-terminal repeat of the tandem. The analysis clearly indicated surface-property alternations, particularly in the R10-18 part of the rod domain, i.e., R11-12, R13-14, R15-16, R17-18 are in cluster 3 and R10-11, R12-13, R14-15 are in cluster 4 (Fig. 3C). This region was previously considered as a highly basic region in view of the calculated pI of the single repeats [60], but it appears that the region is in fact made up of alternating repeats with opposite electrostatic properties.

### Molecular dynamics relaxation

To further assess the quality of the proposed models, we studied tandem repeats by molecular dynamics. We applied this approach to four selected models bearing different types of linker and physical properties. Two of the chosen repeats displayed helical linkers (R1-2 and R23-24), and two displayed non-helical linkers (R10-11 and R15-16). R1-2 belonged to the R1-3 domain, which binds to lipids, while R23-24 belonged to the R20-24 domain, which does not [27]. The thermal stability of R1-2 was ten degrees

lower than that of R23-24 [29]. In R10-11, the linker consisted of a short loop, while there was a long unstructured linker of 20 residues in the R15-16 model. The four tandem repeats were also chosen to cover the whole range of electrostatic properties, one from each of the four clusters in the Fig. 3 dendrogram.

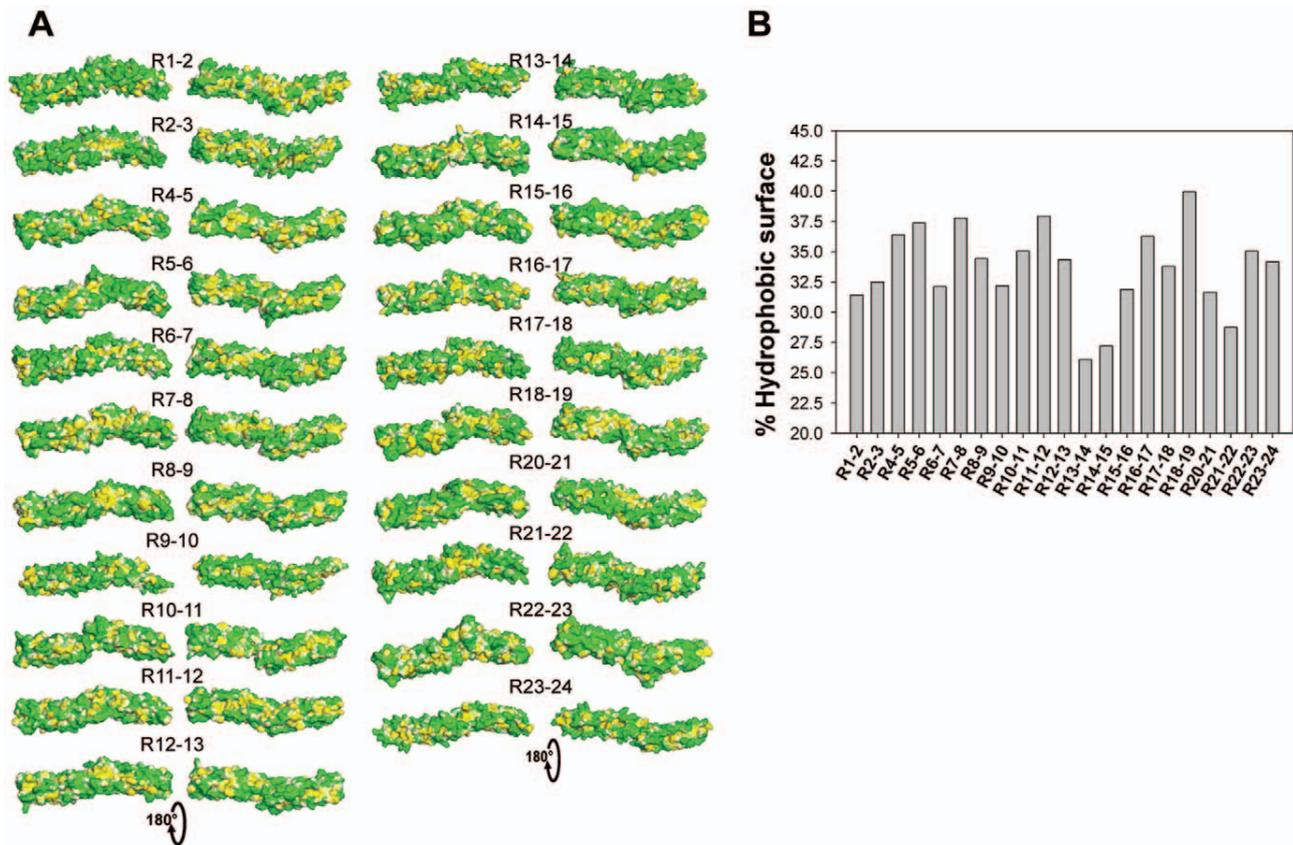
As shown by the RMSD of snapshots measured along the whole trajectory, the three simulations for R1-2, R15-16 and R23-24 converged after 10 ns, while the R10-11 simulation clearly sampled two different conformations (Fig. 4). As expected, the RMSF (Fig. S4) showed that the more flexible regions corresponded to the loops between the helices. All helices remained stable with the exception of R24 HB. It is also worth noting that the heptad pattern (black arrows, Fig. S4) was well maintained, with a lower RMSF for the hydrophobic residues in positions (a) and (d).

The R1-2 tandem was relatively stable throughout the simulation (Fig. 4), with a well-maintained internal structure for both repeats and for the helical linker. The two clusters obtained from the trajectory were very similar. However, the structure diverged rapidly from the initial model. Both R1 HB and R2 HB bent and the two helices rotated around the helical linker to finally interact with each other. At the same time, there was a slight unfolding of R1 HB around residues H55 to G57. The rotation had only a limited effect on the percentage of accessible hydrophobic surface vs. the total molecular surface (Fig. S5 A, B). R1-2 is the most negatively charged tandem repeat of the dystrophin rod domain (Fig. 3), and the conformational rearrangement only slightly modified the electrostatic surface of the molecule (Fig. S5B). Interestingly, the slight unfolding and the bending of R1 HB clearly individualized the putative amphipathic lipid-packing sensor (ALPS) motif of the repeat, i.e., residues Q56 – G73 [61].

Starting from an elongated shape, R10-11 began to bend at the non-helical linker within the first nanosecond. The conformation was then stable for several nanoseconds until the molecule was reorganized into an ever more kinked structure (Fig. 4). Nevertheless, both repeats remained very stable through the trajectory and only their relative positions changed. The same results were observed in the control trajectory. This profound rearrangement was likely driven by the presence of numerous exposed hydrophobic residues at the linker, as pointed out by the significant decrease of the hydrophobic contribution to the molecular surface (Fig. S5A). To minimize the hydrophobic cost, the beginning of R10 HB first interacted with the center of the small R11 HA. The conformation was further stabilized through the formation of multiple contacts between R10 HB, R10 HC and R11 HB, and a tight fit between the two repeats. This structural rearrangement strongly modified the electrostatic properties of the molecule, leading to the appearance of a substantial positively charged pocket with negatively charged surfaces on both sides (Fig. S5B).

The conformation of R15-16 was stable (Fig. 4), although the total RMSD increased slightly along the trajectory due to the flexibility of the long non-helical linker. Interestingly, the cluster analysis identified two preferred conformations that interchanged during the trajectory. The main difference between the two clusters was due to variation in the distance between the non-helical linker and the R16 HB-HC coiled-coil. The interaction between R15 HB and R16 HA was maintained throughout the simulation to minimize the exposure of hydrophobic residues. Nevertheless, the hydrophobic and electrostatic surface properties were mostly the same for the two clusters and differed only slightly from the initial model (Fig. S5A, B).

After a first rearrangement, the R23-24 trajectory converged, although the global conformation remained quite flexible for the



**Figure 2. Molecular hydrophobicity potential of the dystrophin tandem-repeat models.** Molecular hydrophobicity potential surfaces were obtained with PLATINUM and displayed using PyMOL. (A) As in Figure 1B, for the image on the left of each model, the molecules are presented with the helix A in front, the N-terminal end on the left and the C-terminal on the right. The right-hand image of each model is rotated 180° along the molecule's long axis, as indicated on the bottom. The hydrophobicity scale is green-white-yellow, with green representing the most hydrophilic regions and yellow the most hydrophobic. (B) Plot of the calculated % of hydrophobic surface of each tandem repeat by PLATINUM. doi:10.1371/journal.pone.0023819.g002

rest of the simulation (Fig. 4). There was a great flexibility of the internal coil between R23 HB and HC, but the global conformation of this repeat did not change. However, there was a substantial rearrangement of R24 HB, with an unfolding of its C-terminal extremity. In contrast to R1-2, the relative orientation of the two repeats was maintained, and no interaction between the HB helices was observed. Accordingly, the hydrophobic and the electrostatic surface properties were only slightly modified (Fig. S5A, B).

## Discussion

The existing assumption in the field is that dystrophin is a key mechanical linker in the muscle fiber through its association with both the cytoskeletal protein F-actin and the plasma membrane-intrinsic protein  $\beta$ -dystroglycan [4,62-63]. The central rod domain itself has been considered to constitute a passive linker, the role of which is to absorb the mechanical tension created by muscle contraction [6]. However, in contrast to this simple assumption, an increasing number of interacting partners of the central rod domain are being discovered, which suggests a more complex biological role.

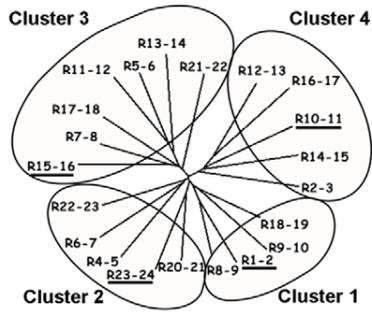
### The models and the nature of the inter-repeat linkers

We have shown that reasonable atomic models can be obtained using I-TASSER, as validated by various assessment procedures.

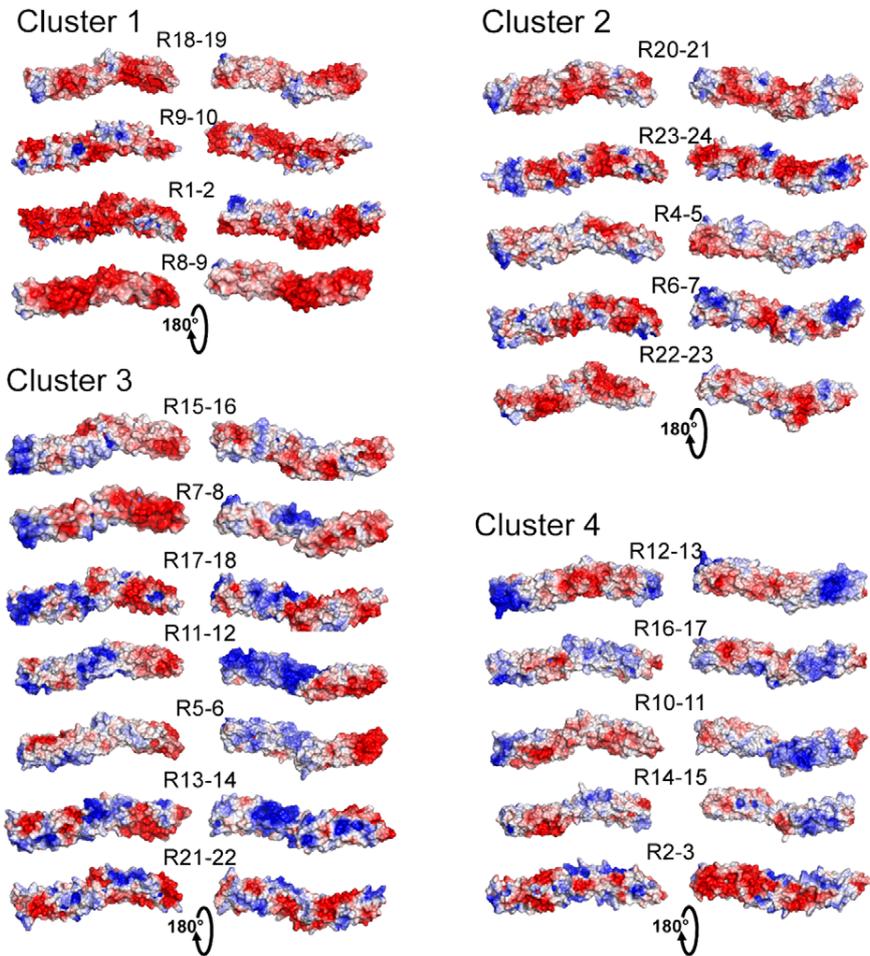
The secondary structures fit well with the spectrin sequences alignment analysis and with the prediction from PSIPRED. In addition, the coiled-coil spectrin-like fold is in agreement with the helicity yields of previously published circular dichroism measurements [24–29]. The low helicity values sometimes observed experimentally were likely due to the absence of neighboring domains. Furthermore, the surface hydrophobicity indicates that the hydrophobic “a” and “d” residue side chains stabilizing the coiled-coil are largely buried inside the models. As a consequence, we observed that the repeat structures remained mostly stable during the molecular dynamics relaxations.

According to the hypothesis of similar folds for dystrophin, spectrin and  $\alpha$ -actinin repeats, all the models fit well with the canonical triple-helical coiled-coil structure obtained by X-ray crystallography and NMR of spectrin repeats [11–21]. However, in contrast to the multi-repeat spectrin structures in which all the linkers were helical, we observed that four tandem-repeat models displayed non-helical linkers (R5-6, R10-11, R13-14 and R15-16). In R15-16, the presence of a 20-residue additional sequence prevented the linker from being helical. The absence of a helical fold for the other three linkers may be due to disruptions of the heptad pattern and/or to the presence of proline residues. However, in two other tandem repeats, R1-2 and R4-5, the heptad pattern rupture did not impair the linker's helical fold. Therefore, we concluded that the presence of proline residues in

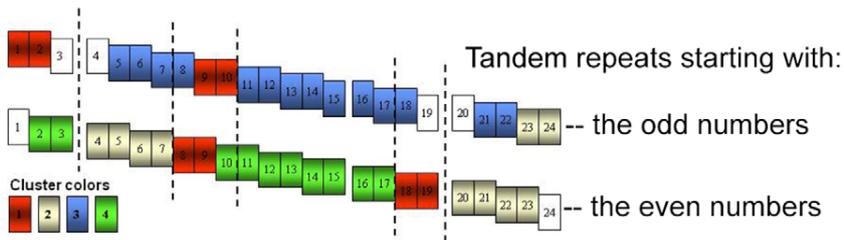
**A**



**B**



**C**



**Figure 3. Molecular electrostatic potential surfaces and PIPSA clustering of the dystrophin tandem-repeat models.** (A) The dendrogram from the PIPSA analysis of the 21 dystrophin tandem repeats showed four clusters. The 4 tandem repeats used in the molecular dynamics simulation are underlined. (B) Representation of the electrostatic potential projected on the solvent accessible surface of the dystrophin repeats, separated into these 4 clusters. Each model was colored using the APBS electrostatic potential calculated for an ionic strength of 50 mM, and the surface colors were clamped at -3 (red) and +3 (blue)  $\text{kTe}^{-1}$ . The molecules are presented with the same orientation as in Fig. 2A. (C) Graphical representation of the repartition of the electrostatic clusters along the rod domain. To take into account the overlapping repeats, the tandems are shown here in two lines, one for tandems starting with an odd number (R1-2, R5-6, etc.), and the other for the evens (R2-3 etc). Each tandem repeat is colored to indicate in which cluster it belongs: red for cluster 1, grey for 2, blue for 3 and green for 4. The uncolored repeats belong to tandem not studied because of the presence of a hinge. Dotted vertical lines are drawn to emphasize the presence of six specific regions. doi:10.1371/journal.pone.0023819.g003

R5-6, R10-11 and R13-14 was likely to be the main cause of the non-helical linkers. Interestingly, no proline residues are located in the inter-repeat linkers of  $\alpha$ - and  $\beta$ -spectrin and  $\alpha$ -actinin. However, these proline residues are conserved in the dystrophin from other vertebrates, such as mice and dogs, and are also present in the linkers of human utrophin repeats 5-6, 10-11, 13-14, 14-15 and 16-17. Thus, the presence of proline residues in the inter-repeats linker constitutes a key feature of dystrophin and utrophin molecules.

### Surface properties of the tandem repeats

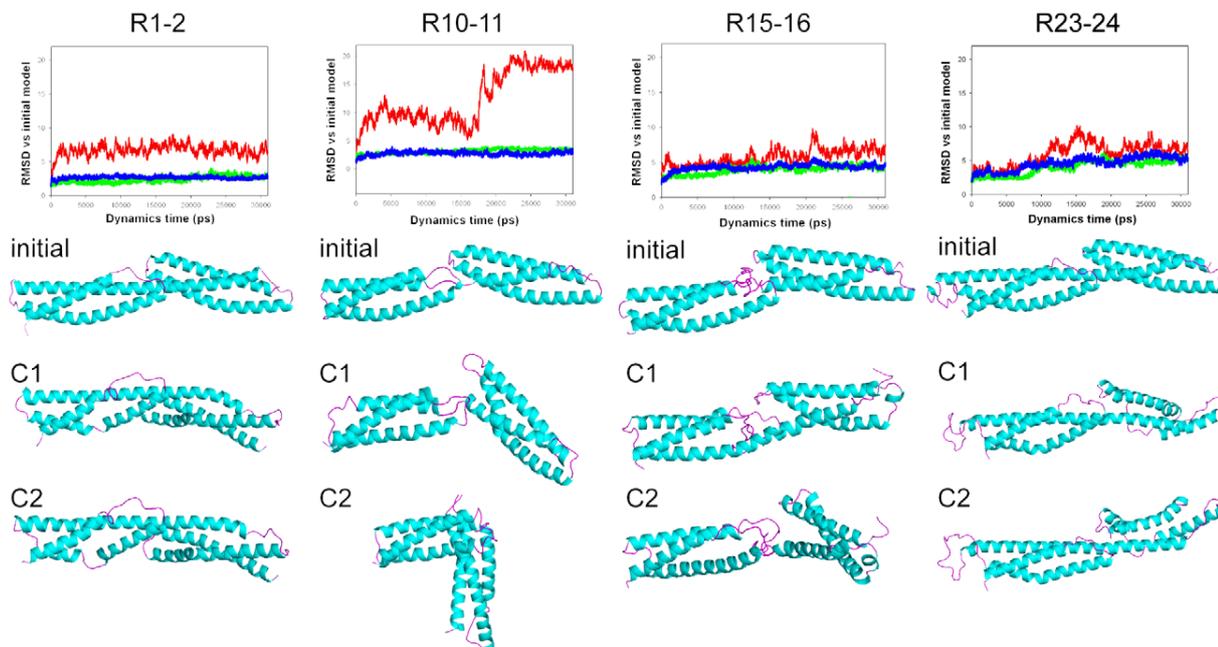
The varied interaction properties of the dystrophin rod domain must be supported by varied surface properties among the repeats. We show here that the repeat surfaces are mostly hydrophilic, in accordance with the heptad pattern's projection of the polar residue side chains outside the coiled-coil. However, the surface hydrophobicity is far from zero, and the repeat surfaces also displayed hydrophobic patches of potential interest.

The use of APBS and PIPSA allowed us to calculate and compare the electrostatic surface potentials of tandem or single repeats of the entire rod domain. Considering all the models, despite substantial heterogeneity of the electrostatic surfaces, the

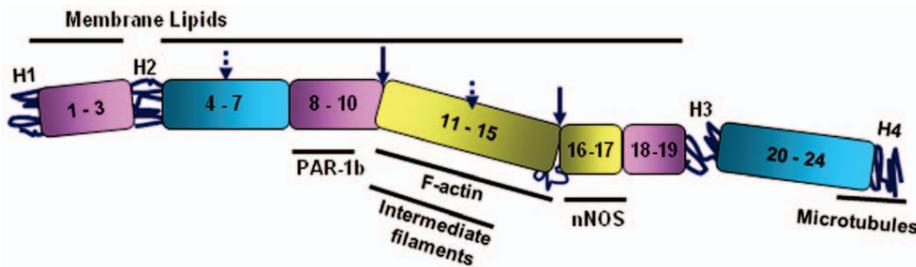
analyses highlighted six distinct regions based on their electrostatic potentials. R1-3, R8-10 and R18-19 are strongly negatively charged regions; R4-7 and R20-24 are less charged with small negative and positive patches; and the R11-17 region is composed of large, alternating strongly negative and positive moieties (Fig. 3B and Fig. 5).

The remaining question is as follows: does this variety of surface properties explain the localization of specific binding sites for the known partners of the dystrophin rod domain? Indeed, the two strongly negatively charged regions R1-3 and R18-19 strangely present strong anionic lipid-binding properties, while R10-17, with its large negatively and positively charged patches, shows a strong affinity to both anionic and zwitterionic lipids [28]. In contrast, the two less charged regions can either bind to lipids, as in the case of R4-7, or not, as for R20-24 [27].

Although the description of lipid binding is quite complex, our models provide some clues as to how the lipids bind to specific repeats. Molecular dynamics experiments with membranes should further define where and how the complementarities lie. On the other hand, surface-property descriptors are clearly insufficient to explain the interaction specificity between the dystrophin repeats and their protein partners. This can be explained by the crucial



**Figure 4. Structural changes in the four tandem repeats observed during the molecular dynamics relaxation.**  $\alpha$  RMSD measured between the initial models and snapshots of the four molecules taken every ps. The red line represents the RMSD of the tandem repeats, the green indicates the RMSD of the N-terminal repeat in the tandem and the blue shows the RMSD of the C-terminal repeat. The simulations were submitted to a clustering procedure in order to identify two clusters per simulation.  $\alpha$  backbone traces of the initial model and the snapshots closest to the center of each cluster (C1 and C2) are shown with the same orientation as in Fig. 1B. doi:10.1371/journal.pone.0023819.g004



**Figure 5. Novel view of the dystrophin central rod domain sub-divided into seven specific structural domains.** H1 to H4 are the four hinges 1 to 4. The new sub-domains are shown here in boxes, with the repeats that they enclose indicated. The colors of the boxes are derived from the electrostatic analysis of Fig. 3B, C. Violet boxes are the most electronegative regions and are mostly constituted by repeats from cluster 1. The blue boxes are intermediate electrostatic regions composed of repeats from clusters 2 and 3. The yellow boxes correspond to a region previously considered as highly basic but which is in fact made up of alternating repeats from clusters 3 and 4 with opposite electrostatic properties. The two solid arrows indicate the newly identified small junctions, while the two dashed arrows indicate the locations of two other putative small junctions between R5 and R6 and R13 and R14. The black straight lines indicate partners and are placed along the dystrophin regions with which they interact. They include nNOS (n-nitric oxide synthase), F-actin (filamentous actin), PAR-1b (polarity-regulating kinase-1b), membrane lipids, intermediate filaments and microtubules.  
doi:10.1371/journal.pone.0023819.g005

role played by both surface and residue complementarities in these processes. To answer this question, we plan to dock our models on atomic structures of well-known dystrophin partners such as nNOS [64] and filamentous actin [65].

### Molecular dynamics relaxation of tandem repeats

Although our models are of high quality, there may be a bias because of the small number of spectrin-repeat templates available. It is therefore of interest to study the molecular dynamics of the dystrophin repeats. Interestingly, three of the four studied models maintained structures very similar to the initial models throughout the entire simulation. The exception was R10-11, where the presence of both a flexible non-helical linker and very strong hydrophobic forces induced a large kink and important changes in the global structure of the tandem, while each of the repeats remained close to the initial model. In contrast, the long linker of R15-16 allows only a restricted bending of the two repeats with respect to each other, and the R1-2 and R23-24 models preserved their helical linkers. These results show the potential limitations of homology modeling and underline the importance of refining structures by MD.

Our observations should also be compared to similar simulation approaches on spectrin repeats. Dynamic flexibility has been previously reported for repeats 8-9 of human erythrocyte  $\beta$ -spectrin and repeats 16-17 of chicken-brain  $\alpha$  spectrin, both with helical linkers [66]. As in our work, the internal structure of each repeat was not modified during the simulations. Atomistic molecular dynamics showed significant bending flexibility governed by the interactions among the AB-loop of the first repeat, the BC-loop of the second repeat and the linker region. At the end of the simulation, the two repeats were in the same face of the linker. This result is in agreement with the observed changes in the respective orientation of the repeats in R1-2, which move onto the same face of the linker early in the simulation. In contrast, R22-23 remained in the same orientation as in the starting models. Therefore, it appears that in dystrophin repeats, the bending directionality is not correlated to the structure of the linker.

Interestingly, these dynamical properties can be directly linked to the presence of specific binding domains. In the case of R1-2, which is known to bind to anionic and curved liposomes, the rapid rotation of repeat 2 around the helical linker placed the potential amphipathic lipid-packing sensor (ALPS) motif [61], situated at the end of HB of R1, on the outside of the molecule. This

mechanism suggests a dynamical control of the interaction with curved membrane surfaces. In the case of R10-11, the large observed kink can be compared with the large differences in the adjacent-repeat orientation observed for  $\beta$ I-spectrin [16-17] and  $\beta$ 2-spectrin [15]. In both studies, repeat 15 exhibits a large tilt angle with respect to repeat 14. Significantly, this tandem repeat constitutes the ankyrin-binding domain, and mutations that interrupt the bonding between two residues essential for maintaining the large tilt angle have been shown to decrease the ankyrin affinity of the tandem repeat [17]. The binding surface of the tandem repeat 14-15 and its electrostatic complementarities [19,67] are disrupted when the angle is no longer present, and this change in turn alters the ankyrin binding. These observations lead to a very interesting hypothesis [17,19] in which modifications of the angle may be dynamically controlled either by the binding of ligands such as lipids [68] or by a mechanical stretch, thus regulating the ankyrin binding. Similarly, the N-terminal end of the  $\alpha$ I and  $\alpha$ II-spectrins forms a large tilt angle with the neighboring repeat 1, consistent with a flexible junction [69]. Again, this flexible junction is thought to play a role in modulating the association affinity of spectrin  $\alpha$  with spectrin  $\beta$  to constitute the spectrin dimer. Our results suggest a similar mechanism for R10-11 that could dynamically control the binding of the neighboring region R11-17 to F-actin.

Altogether, our simulations provided evidence of at least two highly flexible linkers that constitute small “junctions” allowing the individualization of the R11-15 region from its neighbors. Strikingly, this region coincides with the initial description of the central actin-binding domain (ABD2) of dystrophin [60], which was later extended to R11-17 [70]. In the absence of an actual molecular mapping of the actin-interacting domain, it is tempting to consider the possibility that the R11-15 domain is the true ABD2. Furthermore, the remaining R16-17 region has been recently shown to play a specific role through interactions with nNOS. Therefore, the dynamics of the two non-helical linkers R10-11 and R15-16 may regulate the interaction of the two close regions R11-15 and R16-17 with their partners, actin filaments, nNOS and membrane lipids. The presence of non-helical linkers in the tandem repeats R5-6 and R13-14 could constitute two other small “junctions”, but this remains to be determined.

In contrast, the presence of a helical linker in the R20-24 region must rigidify this entire region. Indeed, the region is known to be the most thermally stable portion of the molecule and also shows

specific surface properties. This stability and the presence of the following small hinge are likely crucial to properly individualizing the neighboring cysteine-rich domain, which interacts with the large DGC macromolecular complex through the integral  $\beta$ -dystroglycan protein [4].

## Conclusion

In conclusion, our computational analysis clearly establishes that dystrophin repeats are highly diverse, with electrostatic and hydrophobic surfaces that are far from identical. Our study also identifies new flexible junctions in the rod domain in addition to the already known hinges. Altogether, the dystrophin rod domain is made up of seven different regions distinguished by their surface properties and the presence of key flexible linkers on both sides (Fig. 5). Through the diversity of these properties, the repeats constitute a large scaffold domain for interactions with multiple proteins and with different lipid partners.

This improved description of the dystrophin central rod domain strongly supports the severity grading of Becker muscular dystrophy (BMD) [5,71–73]. Indeed, beside the very severe DMD due to the total deficit of dystrophin, BMD varies from very mild to severe and with or without cardiac involvement in addition to the muscular damage. BMD is mainly caused by in-frame mutations which in the vast majority of cases are deletions of one or several exons. Apart for deletions of the 3' and 5' of the gene, the mutated dystrophin molecules are internally truncated by large parts of the central rod domain. However, the molecular basis of the variability of the BMD severity is difficult to establish because the precise properties of the central rod domain are yet largely unknown. Therefore, our work gives now an integrative view of the central rod domain properties which will help to interpret the BMD variability in view of the missing regions in truncated dystrophin molecules. Similarly, our work may help in the design of truncated dystrophin molecules to be expressed, either by gene replacement or by exon skipping, in Duchenne muscular dystrophy patients who lack dystrophin [63,74].

We emphasized above that crystallization of dystrophin repeats has never been successful. In view of the biophysical properties of the different tandem repeats that we reveal here, key experiments will include crystallization assays on single or tandem repeats and SAXS studies [75] on multi-repeat proteins such as those previously studied by our group [27–28]. Our models also constitute a rational molecular platform for initiating docking studies with atomic structures of known partners such as nNOS, F-actin and lipids, and to guide site-directed mutagenesis to more precisely and experimentally define the surfaces involved in these interactions.

## Supporting Information

**Figure S1 Sequence alignments.** (A) Alignment of the 24 dystrophin repeats and the 8 spectrin repeats used by I-TASSER as templates. Repeats were aligned by ClustalW using default parameters. The alignment was visualized in Jalview and colored using the ClustalX color scheme. Each residue is marked by a specific color only when there is similarity across the repeats. In the bottom line, heptad motifs are indicated, showing the hydrophobic residues in the (a) and (d) positions. The presence of hinges or extra-sequences is mentioned at the end of the corresponding line. (B) The end of a repeat is aligned with the beginning of the following repeat to help visualize the linker within the tandem repeats. The heptad pattern is indicated as in (A). At the right, we indicate whether the heptad pattern is respected (+)

or not (–) in the linker. Insertions are indicated by marking the hinges or the extra-sequences.

(TIF)

### Figure S2 Three dimensional homology models of the 21 dystrophin tandem repeats obtained by I-TASSER.

The models are represented as C $\alpha$  backbone traces. Helical segments are colored in blue and loops are in pink. The N-terminal is on the left, the C-terminal is on the right, and the helix A of the N-terminal repeat of each tandem is at the front of the image. As a reminder, the nomenclature of the helices is given for the R1-2 model. A, B and C are the helices of the first repeat, and A', B' and C' are the helices of the second repeat of the tandem.

(TIF)

### Figure S3 Superposition of the spectrin-repeat crystallographic structures and of the dystrophin tandem-repeat models.

Top: Spectrin repeats shown were those used as templates by I-TASSER: chicken-brain  $\alpha$ -spectrin repeats R15, R16 and R17 (1U4Q) [14], erythroid  $\beta$ -spectrin repeats R8 and R9 (1S35) [13] and  $\beta$ 2-spectrin repeats R14, R15 and R16 (3EDV) [15]. Bottom: superposition of the dystrophin tandem repeats modeled with I-TASSER. The figure was made using PyMOL.

(TIF)

### Figure S4 Quality assessment of the molecular dynamics relaxation of four tandem repeats.

The residue-by-residue backbone fluctuation profile (RMSF) of the eight repeat units R1, R2, R10, R11, R15, R16, R23 and R24 is shown with the primary sequence of each isolated repeat aligned according to the heptad pattern. The (a) and (d) residues are marked with black triangles.

(TIF)

### Figure S5 Influence of the molecular dynamics relaxation on the hydrophobicity and electrostatics of the molecule surfaces.

As in Figure 1B, for the image on the left of each model, the molecules are presented with the helix A in front, the N-terminal end on the left and the C-terminal on the right. The right-hand image of each model is rotated 180° along the molecule's long axis, as indicated on the bottom. The initial model and the snapshots closest to the center of each cluster (C1 and C2) are shown in both cases. (A) Molecular hydrophobicity potential surfaces calculated with PLATINUM are shown using PyMOL. The hydrophobicity scale is green-white-yellow, with green representing the most hydrophilic regions and yellow the most hydrophobic. (B) Representation of the electrostatic potential projected on the solvent accessible surface of the dystrophin tandem repeats. Each model was colored using the APBS electrostatic potential calculated for an ionic strength of 50 mM, and the surface colors were clamped at -3 (red) and +3 (blue) kTe-1.

(TIF)

### Table S1 I-TASSER statistics for the tandem-repeat models of the dystrophin central rod domain.

(TIF)

### Table S2 Simulated system dimensions for the molecular dynamics study of the four tandem-repeats, R1-2, R10-11, R15-16 and R23-24.

(TIF)

### Annex S1 Quality assessment of homology modeling of all tandem repeats of the human dystrophin rod domain.

For each model, the sequence and an image of the model as it appeared in Fig. 1 are shown first and second, respectively. Shown next are the results from PROCHECK, with

the Z score and the graph of the energy for each residue in the sequence. The Verify3D results follow, along with the Ramachandran plot from PROCHECK. This is completed by an image of the model, with the residues in the disallowed regions of the Ramachandran plot colored in red.  
(PDF)

**Models S1** Atomic coordinates (in PDB format) of 3D homology models of the 21 dystrophin tandem repeats obtained with I-TASSER.  
(RAR)

## References

- Koenig M, Monaco AP, Kunkel LM (1988) The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* 53: 219–226.
- Ohlendieck K, Campbell K (1991) Dystrophin constitutes 5% of membrane cytoskeleton in skeletal muscle. *FEBS Letters* 283: 230–234.
- Michele D, Campbell K (2003) Dystrophin-glycoprotein complex: post-translational processing and dystroglycan function. *J Biol Chem* 278: 15457–15460.
- Ervasti JM (2007) Dystrophin, its interactions with other proteins, and implications for muscular dystrophy. *Biochimica Biophysica Acta* 1772: 108–117.
- Koenig M, Beggs A, Moyer M, Scherpf S, Heindrich K, et al. (1989) The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. *Am J Hum Genet* 45: 498–506.
- Davies KE, Nowak KJ (2006) Molecular mechanisms of muscular dystrophies: old and new players. *Nat Rev Mol Cell Biol* 7: 762–773.
- Le Rumeur E, Winder SJ, Hubert JF (2010) Dystrophin: More than just the sum of its parts. *Biochim Biophys Acta* 1804: 1713–1722.
- Broderick MJ, Winder SJ (2005) Spectrin, alpha-actinin, and dystrophin. *Adv Protein Chem* 70: 203–246.
- Lupas A (1996) Coiled coils: new structures and new functions. *Trends Biochem Sci* 21: 375–382.
- Burkhard P, Stetefeld J, Strelkov SV (2001) Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol* 11: 82–88.
- Yan Y, Winograd E, Viel A, Cronin T, Harrison S, et al. (1993) Crystal Structure of the Repetitive Segments of Spectrin. *Science* 262: 2027–2030.
- Grum VL, Li D, MacDonald RI, Mondragon A (1999) Structures of two repeats of spectrin suggest models of flexibility. *Cell* 98: 523–535.
- Kusunoki H, MacDonald R, Mondragon A (2004) Structural insights onto the stability and flexibility of unusual erythroid spectrin repeats. *Structure* 12: 645–656.
- Kusunoki H, Minasov G, MacDonald R, Mondragon A (2004) Independent movement, dimerization and stability of tandem repeats of chicken brain alpha-spectrin. *J Mol Biol* 344: 495–511.
- Davis L, Abdi K, Machius M, Brautigam C, Tomchick DR, et al. (2009) Localization and structure of the ankyrin-binding site on beta2-spectrin. *J Biol Chem* 284: 6982–6987.
- Ipsaro JJ, Huang L, Mondragon A (2009) Structures of the spectrin-ankyrin interaction binding domains. *Blood* 113: 5385–5393.
- Stabach PR, Simonovic I, Ranieri MA, Aboodi MS, Steitz TA, et al. (2009) The structure of the ankyrin-binding site of beta-spectrin reveals how tandem spectrin-repeats generate unique ligand-binding properties. *Blood* 113: 5377–5384.
- Ipsaro JJ, Harper SL, Messick TE, Marmorstein R, Mondragon A, et al. (2010) Crystal structure and functional interpretation of the erythrocyte spectrin tetramerization domain complex. *Blood* 115: 4843–4852.
- Ipsaro JJ, Mondragon A (2010) Structural basis for spectrin recognition by ankyrin. *Blood* 115: 4093–4101.
- Djinovic-Carugo K, Young P, Gaudel M, Saraste M (1999) Structure of the  $\beta$ -actinin rod: molecular basis for cross linking of actin filaments. *Cell* 98: 537–546.
- Ylanne J, Scheffzek K, Young P, Saraste M (2001) Crystal Structure of the alpha-Actinin Rod: Four Spectrin Repeats Forming a Thight Dimer. *Cell Mol Biol Lett* 6: 234.
- Pascual J, Pfuhl M, Walther D, Saraste M (1997) Solution structure of spectrin repeat: a left-handed antiparallel triple-helical coiled-coil. *J Mol Biol* 273: 740–751.
- Baines AJ (2010) Evolution of the spectrin-based membrane skeleton. *Transfus Clin Biol* 17: 95–103.
- Kahana E, Marsh PJ, Henry AJ, Way M, Gratzel WB (1994) Conformation and phasing of dystrophin structural repeats. *J Mol Biol* 235: 1271–1277.
- Kahana E, Gratzel WB (1995) Minimum folding unit of dystrophin rod domain. *Biochemistry* 34: 8110–8114.
- Saadat L, Pittman L, Menhart N (2006) Structural cooperativity in spectrin type repeats motifs of dystrophin. *Biochim Biophys Acta* 1764: 943–954.
- Legardinier S, Hubert J-F, Le Bihan O, Tascon C, Rocher C, et al. (2008) Subdomains of the dystrophin rod domain display contrasting lipid-binding and stability properties. *Biochim Biophys Acta* 1784: 672–682.
- Legardinier S, Raguénès-Nicol C, Tascon C, Rocher C, Hardy S, et al. (2009) Mapping of the lipid-binding and stability properties of the central rod domain of human dystrophin. *J Mol Biol* 389: 546–558.
- Mirza A, Sagathevan M, Sahni N, Choi L, Menhart N (2010) A biophysical map of the dystrophin rod. *Biochim Biophys Acta* 1804: 1796–1809.
- Petrey D, Honig B (2005) Protein structure prediction: inroads to biology. *Mol Cell* 20: 811–819.
- Zhang Y (2009) Protein structure prediction: when is it useful? *Curr Opin Struct Biol* 19: 145–155.
- Koenig M, Kunkel LM (1990) Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. *J Biol Chem* 265: 4560–4566.
- Winder S, Gibson T, Kendrick-Jones J (1995) Dystrophin and utrophin: the missing links! *FEBS Letters* 369: 27–33.
- Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292: 195–202.
- Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, et al. (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res* 33: W36–38.
- 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): 100–113.
- DeLano WL, ed (2002) The PyMOL user's manual. San CarlosCA: DeLano Scientific.
- Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17: 355–362.
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35: W407–410.
- Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253: 164–170.
- Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83–85.
- Efremov RG, Chugunov AO, Pyrkov TV, Priestle JP, Arseniev AS, et al. (2007) Molecular lipophilicity in protein modeling and drug design. *Curr Med Chem* 14: 393–415.
- Pyrkov TV, Chugunov AO, Krylov NA, Nolde DE, Efremov RG (2009) PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics* 25: 1201–1202.
- Blomberg N, Gabdoulina RR, Nilges M, Wade RC (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins* 37: 379–387.
- Wade R, Gabdoulina R, De Rienzo F (2001) Protein Interaction Property Similarity Analysis. *Int J Quant Chem* 83: 122–127.
- Gabdoulina RR, Stein M, Wade RC (2007) qPIPSA: relating enzymatic kinetic parameters and interaction fields. *BMC Bioinformatics* 8: 373.
- Richter S, Wenzel A, Stein M, Gabdoulina RR, Wade RC (2008) webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Res* 36: W276–280.
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98: 10037–10041.
- Felsenstein J (1987) Estimation of hominoid phylogeny from a DNA hybridization data set. *J Mol Evol* 26: 123–131.
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.
- Mackerell AD, Bashford D, Bellott M, Dunbrack RL Jr., Evansck J, et al. (1998) All atom empirical potential for molecular modeling and dynamics studies of protein. *J Phys Chem B* 102: 3586–3615.
- Foloppe N, Mackerell AD, Jr. (2000) Parameter optimization based on small molecule and condensed phase macromolecular target. *J Comput Chem* 21: 86–104.
- MacKerell AD, Jr., Banavali N (2000) Application to molecular dynamics simulations of DNA and RNA in solution. *J Comput Chem* 21: 105–120.

55. Mackerell AD, Jr., Feig M, Brooks CL, 3rd (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25: 1400–1415.
56. Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J Mol Graph* 14: 33–38.
57. Shao JT, Tanner SW, Thompson N, Cheatham TE (2007) III. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J Chem Theoret Comput* 3: 2312–2334.
58. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477–486.
59. Menhart N (2006) Hybrid spectrin type repeats produced by exon-skipping in dystrophin. *Biochim Biophys Acta* 1764: 993–999.
60. Amann KJ, Renley BA, Ervasti JM (1998) A Cluster of Basic Repeats in the Dystrophin Rod Domain Binds F-actin through an Electrostatic Interaction. *J Biol Chem* 273: 28419–28423.
61. Drin G, Casella JF, Gautier R, Boehmer T, Schwartz TU, et al. (2007) A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat Struct Mol Biol* 14: 138–146.
62. Deconinck N, Dan B (2007) Pathophysiology of duchenne muscular dystrophy: current hypotheses. *Pediatr Neurol* 36: 1–7.
63. Banks GB, Chamberlain JS (2008) The value of mammalian models for duchenne muscular dystrophy in developing therapeutic strategies. *Curr Top Dev Biol* 84: 431–453.
64. Li H, Shimizu H, Flinspach M, Jamal J, Yang W, et al. (2002) The novel binding mode of N-alkyl-N'-hydroxyguanidine to neuronal nitric oxide synthase provides mechanistic insights into NO biosynthesis. *Biochemistry* 41: 13868–13875.
65. Fujii T, Iwane AH, Yanagida T, Namba K (2010) Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature* 467: 724–728.
66. Mirjaniyan DT, Chu JW, Ayton GS, Voth GA (2007) Atomistic and coarse-grained analysis of double spectrin repeat units: the molecular origins of flexibility. *J Mol Biol* 365: 523–534.
67. La-Borde PJ, Stabach PR, Simonovic I, Morrow JS, Simonovic M (2010) Ankyrin recognizes both surface character and shape of the 14-15 di-repeat of beta-spectrin. *Biochem Biophys Res Commun* 392: 490–494.
68. Czogalla A, Grzymajlo K, Jezierski A, Sikorski AF (2008) Phospholipid-induced structural changes to an erythroid beta spectrin ankyrin-dependent lipid-binding site. *Biochim Biophys Acta* 1778: 2612–2620.
69. Mehboob S, Jacob J, May M, Kotula L, Thiyagarajan P, et al. (2003) Structural analysis of the alpha N-terminal region of erythroid and nonerythroid spectrins by small-angle X-ray scattering. *Biochemistry* 42: 14702–14710.
70. Rybakova I, Humston J, Sonneman K, Ervasti J (2006) Dystrophin and utrophin bind actin through distinct modes of contact. *J Biol Chem* 281: 9996–10001.
71. Beggs A, Hoffman E, Snyder J, Arahata K, Specht L, et al. (1991) Exploring the molecular basis for variability among patients with Becker muscular dystrophy: dystrophin gene and protein studies. *Am J Hum Genet* 49: 54–67.
72. Kaspar RW, Allen HD, Ray WC, Alvarez CE, Kissel JT, et al. (2009) Analysis of dystrophin deletion mutations predicts age of cardiomyopathy onset in becker muscular dystrophy. *Circ Cardiovasc Genet* 2: 544–551.
73. Tuffery-Giraud S, Beroud C, Leturcq F, Yaou RB, Hamroun D, et al. (2009) Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. *Hum Mutat* 30: 934–945.
74. Nakamura A, Takeda S (2009) Exon-skipping therapy for Duchenne muscular dystrophy. *Neuropathology* 29: 494–501.
75. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40: 191–285.

## Annexe D

Article paru dans **Orphanet Journal  
of Rare Diseases**

RESEARCH

Open Access

# Assessment of the structural and functional impact of in-frame mutations of the *DMD* gene, using the tools included in the eDystrophin online database

Aurélien Nicolas<sup>1,2,3</sup>, Céline Lucchetti-Miganeh<sup>1,3,4</sup>, Rabah Ben Yaou<sup>5,6</sup>, Jean-Claude Kaplan<sup>6,7,8</sup>, Jamel Chelly<sup>6,7,8</sup>, France Leturcq<sup>6,7,8</sup>, Frédérique Barloy-Hubler<sup>1,3,4,9</sup> and Elisabeth Le Rumeur<sup>1,2,4,10\*</sup>

## Abstract

**Background:** Dystrophin is a large essential protein of skeletal and heart muscle. It is a filamentous scaffolding protein with numerous binding domains. Mutations in the *DMD* gene, which encodes dystrophin, mostly result in the deletion of one or several exons and cause Duchenne (DMD) and Becker (BMD) muscular dystrophies. The most common *DMD* mutations are frameshift mutations resulting in an absence of dystrophin from tissues. In-frame *DMD* mutations are less frequent and result in a protein with partial wild-type dystrophin function. The aim of this study was to highlight structural and functional modifications of dystrophin caused by in-frame mutations.

**Methods and results:** We developed a dedicated database for dystrophin, the eDystrophin database. It contains 209 different non frame-shifting mutations found in 945 patients from a French cohort and previous studies. Bioinformatics tools provide models of the three-dimensional structure of the protein at deletion sites, making it possible to determine whether the mutated protein retains the typical filamentous structure of dystrophin. An analysis of the structure of mutated dystrophin molecules showed that hybrid repeats were reconstituted at the deletion site in some cases. These hybrid repeats harbored the typical triple coiled-coil structure of native repeats, which may be correlated with better function in muscle cells.

**Conclusion:** This new database focuses on the dystrophin protein and its modification due to in-frame deletions in BMD patients. The observation of hybrid repeat reconstitution in some cases provides insight into phenotype-genotype correlations in dystrophin diseases and possible strategies for gene therapy. The eDystrophin database is freely available: <http://edydystrophin.genouest.org/>.

**Keywords:** Dystrophin, *DMD* gene mutations, Spectrin-like repeats, Duchenne muscular dystrophy, Becker muscular dystrophy, Phenotype-genotype correlation

## Background

The *Duchenne Muscular Dystrophy (DMD)* gene, located on the short arm of the X chromosome (at Xp21.2), is the largest known gene in humans. It has an open reading frame of ~11.055 kb, containing 79 exons (Mendelian Inheritance in Man [MIM: 300377]) [1], and transcription from seven tissue-specific promoters leads

to the synthesis of 16 isoforms of the dystrophin protein. In humans, dystrophin diseases are caused by mutations in the *DMD* gene and include the allelic phenotypes of Duchenne muscular dystrophy (DMD) [OMIM:310200], Becker muscular dystrophy (BMD) [OMIM:300376] and X-linked dilative cardiomyopathy (XLDCM) [OMIM:302045] [1–3].

Dystrophin is present at the internal face of the plasma membrane in many tissues, including skeletal, cardiac and smooth muscle, and in various central nervous system cells. Dystrophin is highly conserved in vertebrates,

\* Correspondence: [elisabeth.lerumeur@univ-rennes1.fr](mailto:elisabeth.lerumeur@univ-rennes1.fr)

<sup>1</sup>Université de Rennes 1 Rennes France

<sup>2</sup>UMR CNRS 6026, Equipe RMN-ILP, Faculté de Médecine, CS 34317 Rennes Cedex 35043, France

Full list of author information is available at the end of the article

including mouse, chicken and dog, and in invertebrates, such as *Drosophila* [4], and *Caenorhabditis elegans* [5].

The full-length 3685-residue isoform of dystrophin, dp427m, has a molecular weight of 427kDa and is expressed in skeletal and cardiac muscle, where it plays a key role during muscle contraction-relaxation cycles. Dystrophin has four main regions: (i) the N-terminal actin-binding domain (ABD) comprises the first 246 residues; (ii) the central rod domain spans residues 247 to 3045 (accounting for about 76% of the molecule [6,7]), is formed by 24 spectrin-like repeats and four hinges and binds to various partners (filamentous actin, membrane lipids and nitric oxide synthase); (iii) the cysteine-rich domain, from residues 3080 to 3360, binds to the intrinsic membrane protein  $\beta$ -dystroglycan and (iv) the carboxy-terminal domain, comprising the last 325 residues, binds to dystrobrevin and syntrophins (for reviews, see [8,9]). Dystrophin is associated with a large number of proteins, to which it either binds directly or with which it interacts indirectly through intracellular or extracellular proteins. The binding of dystrophin to  $\beta$ -dystroglycan brings it into contact with membrane and extracellular proteins to form the dystrophin-glycoprotein complex (DGC) [10,11]. Dystrophin therefore forms a link between the extracellular matrix and cytoskeletal actin. The function of dystrophin is not completely understood, but its main role is to protect the sarcolemma from rupture during the stresses of muscle contraction [12,13].

Patients with muscular dystrophy have little or no dystrophin, and it has been suggested that this results in the disruption of muscle membranes, which alters calcium-channel activity, thereby strongly increasing intracellular calcium concentration. This ultimately leads to muscle-cell necrosis [13,14], followed by regeneration. The continual cycles of regeneration and necrosis lead to the skeletal muscles being gradually replaced by adipose tissue and unable to sustain any mechanical activity.

The incidence of Duchenne muscular dystrophy (DMD) is about 1 in 3,500 male births. Affected patients have a massively reduced life expectancy and a poor functional prognosis. In most cases, DMD is due to frame-shift mutations in the *DMD* gene, leading to a complete absence or low levels of dystrophin protein (no more than 3% normal levels). This accounts for the severity of the phenotype in all patients, although some variation of disease expression is observed between patients, in terms of motor, respiratory, cardiac and mental functions [15,16]. The expression of the various dystrophin isoforms may depend on the mutation site, potentially accounting, at least in part, for the correlation with the motor and mental status of patients [15]. Becker muscular dystrophy (BMD) is less frequent than DMD, and is usually milder, with slower disease

progression. BMD is caused by in-frame deletions or duplications of one or several exons or by splice-site and missense mutations. These mutations lead to the production of various amounts of internally truncated, lengthened, or slightly modified dystrophin molecules. This results in a broad spectrum of clinical severity, ranging from a complete absence of symptoms, through mild disease, to severe clinical conditions similar to DMD [15–25]. According to the reading-frame rule, frame-shifting mutations lead to the severe DMD phenotype, whereas in-frame mutations lead to the less severe BMD phenotype [16,26]. However, there are exceptions to this rule, with certain in-frame mutations resulting in the severe DMD phenotype [27–31]. These mutations are frequently located at the 5' end of the gene encoding the N-terminus of dystrophin including ABD1, or at the 3' end of the gene encoding the C-terminal domain, usually in the Cys-rich domain, thereby disrupting the DGC.

The structure and function of dystrophin are poorly resolved at the biological and physiological levels, and it is therefore difficult to establish a detailed phenotype-genotype correlation in BMD patients. Phenotypic differences between patients are thought to depend on the site of the deletion or duplication and the conservation of the reading frame, and such differences have recently been shown to be correlated with the residual amount of dystrophin [32]. Such knowledge is essential to anticipate the effects of current exon-skipping treatments on phenotype restoration in treated DMD patients [33–36].

Two databases of *DMD* human mutations are already freely available online: the Leiden Muscular Dystrophy database [37,38] and the UMD-DMD French database [39,40]. The Leiden Muscular Dystrophy database lists the *DMD* mutations in patients reported in publications or submitted by contributors from around the world and includes some biochemical and phenotypic details. The UMD-DMD database provides molecular and clinical data for patients from France carrying a mutation of the *DMD* gene. Both databases include in-frame and frame-shifting mutations and focus on gene-level information. However, as dystrophin acts at the protein level, a more detailed and comprehensive characterization of the protein produced from genes with in-frame gene mutations is required. Such a characterization is particularly important for comparisons of the structural features and molecular interactions of the mutated protein with those of the wild-type protein. For example, the total absence of dystrophin, or its presence in very small amounts in DMD patients, leads to the breakdown of the DGC complex, a histological marker of the disease [41]. However, the site of the mutation determines whether these interactions are abolished in BMD patients, resulting in diverse phenotypes. For both basic research and clinical/

therapeutic purposes, it is therefore of interest to establish a correlation between the genotype and the molecular and structural consequences of in-frame mutations for the encoded protein.

To this end, we have developed a new database called eDystrophin, specifically dedicated to providing information about the in-frame mutations of the *DMD* gene and their consequences for dystrophin molecules. The eDystrophin database includes both in-frame *DMD* mutations identified at a routine diagnostic laboratory for these mutations in France and published mutations. In addition to the genetic and clinical details provided by the other two available databases, the eDystrophin database provides: (1) a synthetic view of the properties of mutated dystrophin, (2) a map of modifications to binding sites for interacting protein partners; for deletions involving the central rod domain, eDystrophin provides (3) a structural model of the mutation site and (4) a specific comment indicating whether a correct filamentous 3D-structure is reconstituted around the mutation site. Finally, this new database focuses on the protein rather than the gene, providing a new vantage point regarding in-frame mutations of the *DMD* gene, with the finding that the gene exons and protein domains are “in phase” for the specific central rod domain of dystrophin. This phasing controls the ability of the internally truncated dystrophin molecules to reconstitute a hybrid repeat unit able to fold into a triple coiled-coil, resembling the native repeats present in full-length dystrophin. This database is freely available from <http://edydystrophin.genouest.org/>, and all the information provided can be downloaded.

## Methods

eDystrophin is a relational database developed in MySQL 5.1.37 within the MAMP package [42]. The website was constructed with XHTML, PHP 5 and JavaScript. The database and the website are hosted by the BioGenouest platform [43] and run with the Apache web server, version 2.2.3. MyDomains [44] was used to represent the mutated proteins. The Jmol applet [45] is embedded for the visualization of three-dimensional structure-homology modeling.

### Database construction

#### Gene data

The *DMD* cDNA sequences for the seven known tissue-specific promoters and the positions of the 79 exons were obtained from GenBank (RefSeq in Table 1).

#### Wild-type dystrophin data

Sequence data for the 16 isoforms of wild-type dystrophin were downloaded from GenBank. Several of these isoforms are generated by alternative splicing in

specific tissues (Table 1). The boundaries of the structural and functional domains were defined according to the findings of 19 published papers and three domain-search tools (Additional file 1, Table S1). This resulted in the definition of 35 structural and 15 binding domains in the eDystrophin database (Figure 1A). All of the variants of each domain mentioned by the different authors are indexed in the database (See the “Knowledge” section), together with the original publication reference.

Two X-ray crystallographic structures of dystrophin domains have been reported: one for actin-binding domain 1 at the N-terminal end [PDB: 1DXX] [51] and the other for a WW domain and EF hand-region fragment complexed with a  $\beta$ -dystroglycan peptide [PDB: 1EG3] [52]. We recently used homology modeling to propose the three-dimensional structure models of the 24 central rod domain repeats [53]. The database gives models for isolated repeats (Figure 1B) and for tandem repeats (Figure 1C). Each repeat consists of three  $\alpha$ -helices: A, B, and C, and A', B', and C' for the following repeat. These two structures and all the structure models can be visualized and downloaded from the eDystrophin database (from the Explore database section).

### Data for mutated dystrophin

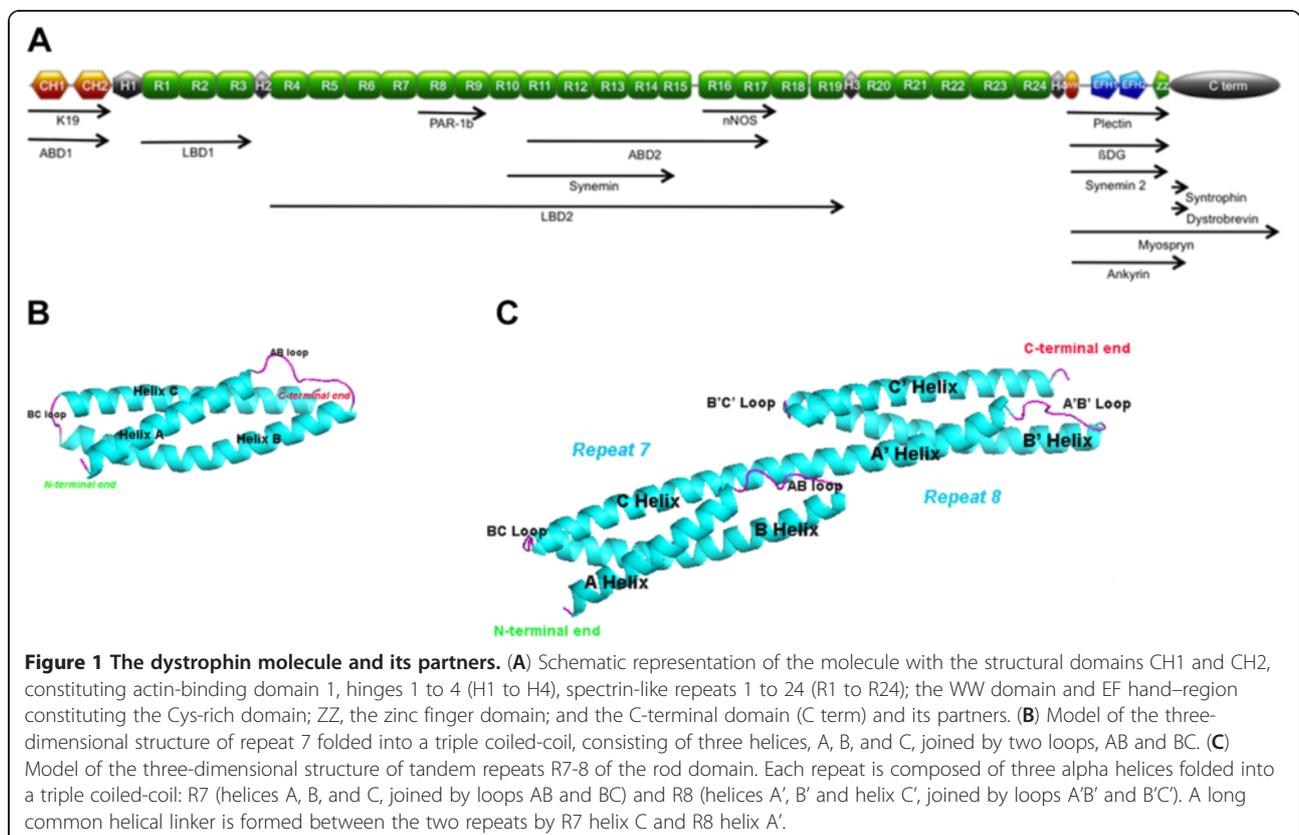
Our principal goal was to provide information about the structure of the protein in cases of *DMD* in-frame mutation, as a valuable tool for exon skipping therapy. We collaborated with one of the two existing databases — the French UMD-DMD database [40] — resulting in the inclusion of all patients carrying in-frame mutations from the largest French cohort, for whom detailed genetic and molecular investigations had been carried out. Data from published studies reporting well characterized exon deletions/duplications were also included. The eDystrophin database compiles 209 different in-frame mutations from a total of 945 patients. One hundred of these mutations were described in previous studies, 67 were provided by the *Laboratoire de Biochimie et Génétique Moléculaire* (LBGM, Cochin Hospital, APHP, Paris) and 42 were present in both sources (Additional file 2, Table S2). The mutations are named according to Human Genome Variation Society nomenclature recommendations [54]. For each mutation, the cDNA sequence was predicted and is available from eDystrophin (See the “Explore database” section). As mRNA studies are rarely performed in cases of exon deletion/duplication, we did not consider mRNA levels in eDystrophin. For duplications, we cannot understand events at the protein level unless we know how the duplication is arranged at the gene level. We therefore assumed that repeats were in tandem and not in opposing directions, in which case a stop codon might occur, resulting in the absence of dystrophin and a DMD phenotype. The eDystrophin database also provides some general

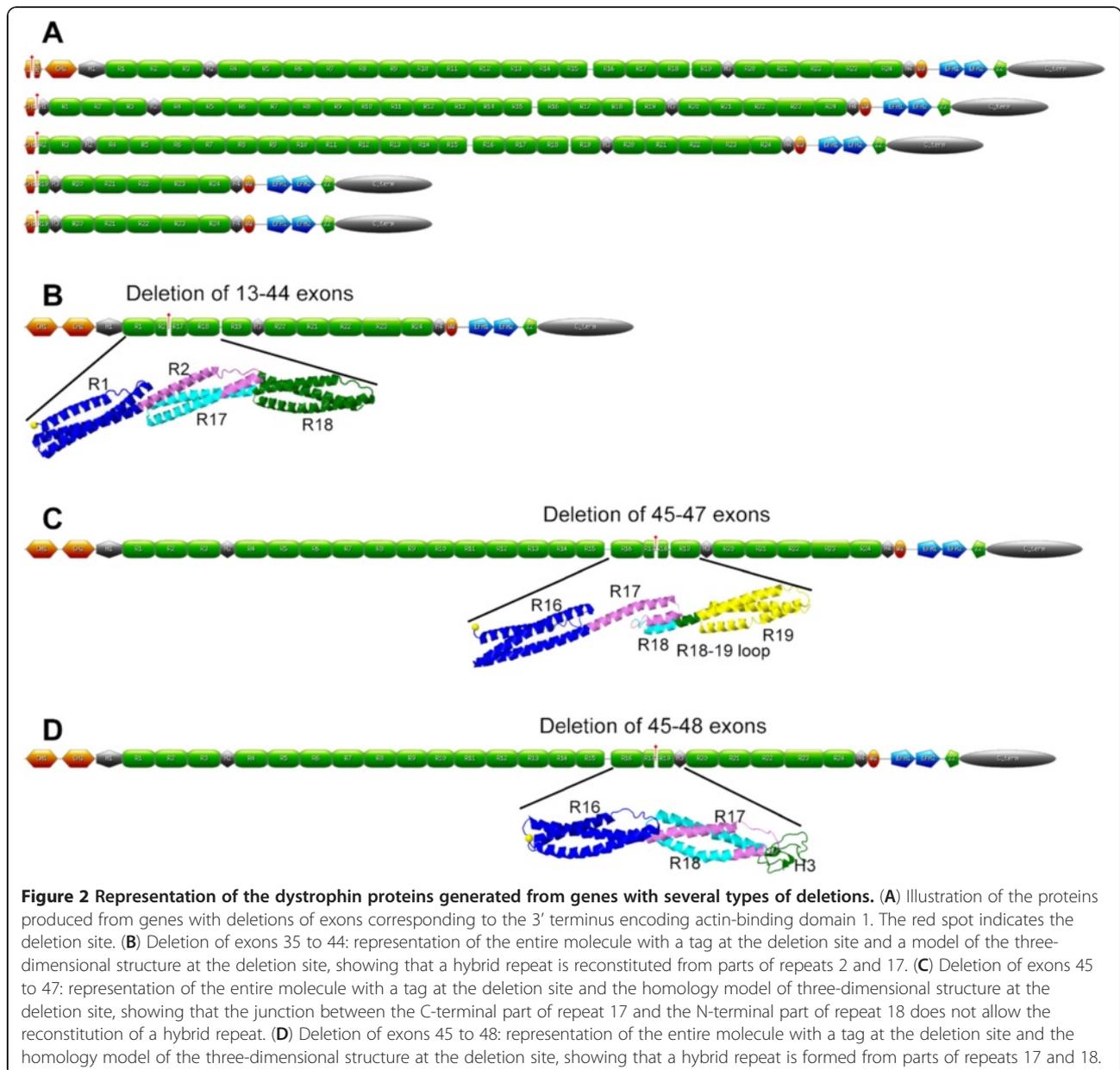
**Table 1 DMD transcript variants and their tissue expression**

Promoters	Publication	Tissue specificity	Alternative splicing	mRNA RefSeq	Protein RefSeq
dp427m	(Koenig <i>et al.</i> , 1989)[16]	Skeletal muscle, heart muscle, glial cells	Not referenced	NM_004006.2	NP_003997.1
dp427c	(Chelly <i>et al.</i> , 1990)[46]	Brain, retina	Not referenced	NM_000109.3	NP_000100.2
dp427p	(Chelly <i>et al.</i> , 1990)[46]	Purkinje cells, muscle	Not referenced	NM_004009.3	NP_004000.1
dp260	(D'Souza <i>et al.</i> , 1995)[47]	Retina	No but two alternative exon 1	1: NM_004011.3 2: NM_004012.3	1: NP_004002.2 2: NP_004003.1
dp140	(Lidov <i>et al.</i> , 1995)[48]	Brain, kidney, retina	No splicing -> dp140 Exons 71 & 78 -> dp140ab Exon 78 -> dp140b Exons 71 to 74 & 78 -> dp140bc Exons 71 to 74 -> dp140c	NM_004013.2 NM_004022.2 NM_004021.2 NM_004023.2 NM_004020.3	NP_004004.1 NP_004013.1 NP_004012.1 NP_004014.1 NP_004011.2
dp116	(Byers <i>et al.</i> , 1993)[49]	Schwann cells	Not referenced	NM_004014.2	NP_004005.1
dp71	(Hugnot <i>et al.</i> , 1992)[50]	Everywhere except skeletal muscle	No splicing -> dp71 Exons 71 -> dp71a Exon 71 & 78 -> dp71ab Exon 78 -> dp71b Stop at Exon 70 -> dp40	NM_004015.2 NM_004017.2 NM_004018.2 NM_004016.2 NM_004019.2	NP_004006.1 NP_004008.1 NP_004009.1 NP_004007.1 NP_004010.1

clinical and protein information (see below) for each of the in-frame *DMD* mutations. For each of the 209 protein sequences derived from these in-frame mutations, a map of the conserved and altered structural and binding

domains of dystrophin was produced, including deletions of exons at the 3' end encoding actin-binding domain 1 (Figure 2A) and deletions of exons 13–44, 45–47 and 45–48 (Figure 2B, C, D).





**Figure 2 Representation of the dystrophin proteins generated from genes with several types of deletions.** (A) Illustration of the proteins produced from genes with deletions of exons corresponding to the 3' terminus encoding actin-binding domain 1. The red spot indicates the deletion site. (B) Deletion of exons 35 to 44: representation of the entire molecule with a tag at the deletion site and a model of the three-dimensional structure at the deletion site, showing that a hybrid repeat is reconstituted from parts of repeats 2 and 17. (C) Deletion of exons 45 to 47: representation of the entire molecule with a tag at the deletion site and the homology model of three-dimensional structure at the deletion site, showing that the junction between the C-terminal part of repeat 17 and the N-terminal part of repeat 18 does not allow the reconstitution of a hybrid repeat. (D) Deletion of exons 45 to 48: representation of the entire molecule with a tag at the deletion site and the homology model of the three-dimensional structure at the deletion site, showing that a hybrid repeat is formed from parts of repeats 17 and 18.

For each deletion of exons encoding part of the central rod domain (exons 10–61), homology modeling was used to determine the structure of the protein at the junctions on either side of the deletion. Homology modeling was performed on the automated server, I-TASSER [55,56], as in our previous study on native repeats [53]. I-TASSER combines various techniques, including threading, *ab initio* modeling and structure-refinement approaches, for the prediction of protein structures. For each submitted sequence, one to five models are produced based on homology to the spectrin repeats, the crystallographic structures of which have been resolved. Briefly, the C-score in I-TASSER estimates the quality of

the predicted model, based on the significance of the threading alignments and the convergence of the simulations. C-score values typically lie in the range of -5 to 2, with higher scores indicative of a better model. Only the model with the best C-score is retained in the eDystrophin database. Furthermore, three-dimensional structure models of isolated hinges were submitted to I-TASSER and were incorporated into models as necessary. These new structures, made of separated blocks, were minimized twice, in water and 50mM NaCl. The models were further analyzed by graphical display with PyMOL [57] and evaluated with PROCHECK (to check the stereochemistry information supplied by ProSA-web

[58,59]) and Verify3D [60,61] (for structure validation). Ramachandran plots showing the amino-acid distribution are provided by eDystrophin, indicating the energetically favorable regions for peptide bond torsion angles for each amino-acid of the protein. The structural model is considered of higher quality if the distribution of amino-acid torsions is restricted to the regions allowed in the plot.

The models can be visualized online with Jmol [45], using a color code for each region of the protein. Comparison of the structure of the mutated protein with native repeat folding is provided by eDystrophin, through a static view of a native three-repeat fragment of dystrophin in a parallel box. Static images and PDB files for the wild-type and mutated models can be downloaded.

### **Clinical data**

In total, 945 clinical records (531 provided by the LBG and 414 from published studies) were included in eDystrophin. The database provides a brief description of the disease phenotype corresponding to the most commonly used clinical classification. The DMD and BMD phenotypes are attributed to cases in which the patient loses the ability to walk before the age of 13 years and after the age of 16 years, respectively. The intermediate muscular dystrophy (IMD) phenotype is used to describe patients who stopped walking between the ages of 13 and 16 years. The DCM phenotype is attributed to the subgroup of patients with isolated cardiomyopathy without skeletal muscle involvement, whereas the “pending” subgroup corresponds to patients with insufficient clinical data for correct classification of their phenotype. Finally, the “asymptomatic” phenotype is assigned to patients with no myopathic or cardiac symptoms at their most recent check-up.

The results of dystrophin immunostaining and/or western blot studies with three specific antibodies against different domains of dystrophin were available for 360 patients. The Dys-1 antibody is specific to repeats 8–9 in the central rod region; the Dys-2 antibody is specific to the C-terminus (residues 3669–3685); and the Dys-3 antibody is specific to the N-terminus (residues 321–494). For the LBG patients, data from immunostaining studies are presented according to staining regularity (normal, regular, irregular, mosaic, no signal with revertant fibers, and no signal) and intensity (high, medium, and low). Western blots with the same three antibodies were used to assess the quantity (high, medium or low) and size (increased, normal, reduced or undetected) of the three major regions of the mutated protein. Thus, for each of the recorded in-frame mutations, the eDystrophin database displays the percentage of patients carrying the mutation with resulting in the production of low, normal or high levels of dystrophin proteins of small, normal or large

size. In addition, for cases in which clinical data are available, the presence of cardiomyopathy or mental retardation at the last clinical examination is also displayed. The aim of eDystrophin was not to provide detailed descriptions of clinical symptoms, dystrophin immunofluorescence and western blot status for each patient, but a global view for each subgroup of patients carrying the same in-frame *DMD* mutation.

## **Results**

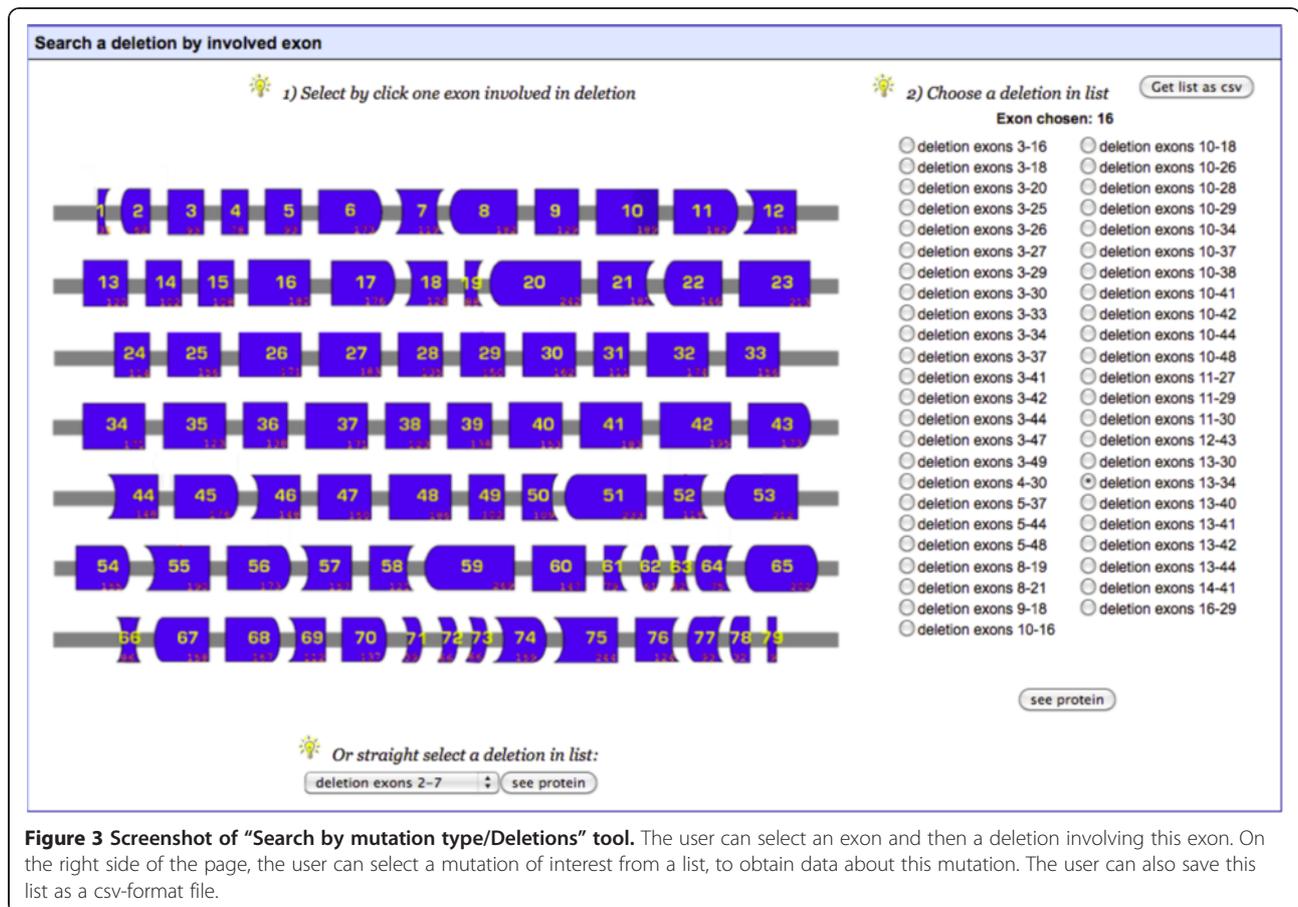
### **Website organization**

The eDystrophin website contains four distinct sections: “Knowledge”, “Explore database”, “Statistics” and “Links”. The “Knowledge” pages provide background information about the dystrophin gene and protein, the diseases associated with mutations of the gene and current cell- and gene-therapy strategies. The “Explore database” pages contain data that can be downloaded, and this section is divided into two parts: ‘Wild-type dystrophin’, and ‘Mutated dystrophin’. The “Statistics” section provides a brief summary of statistics for eDystrophin content. Finally, the “Links” section provides some useful links and a list of previous publications reporting well characterized exon deletions/duplications implemented in eDystrophin.

The ‘Wild-type dystrophin’ pages provide all of the wild-type sequences and the corresponding three-dimensional structures of the domains. The cDNA and protein sequences of the 16 isoforms can be obtained by clicking on the “Isoform full-length sequences” tab. A diagram of the organization of the cDNA can be obtained by clicking on the “Exon sequences” tab, and the sequences of the 79 exons can be obtained by clicking on the chosen exon. The sequences of the various domains, with all of the versions reported in published studies, can be obtained by clicking on the “Structural domain sequences” tab, and a diagram of dystrophin with the functional domains reported in published studies can be obtained by clicking on the “Binding domain sequences” tab. By clicking on the domain, the user can download the sequence. Similarly, the three-dimensional structures found in PDB are available via the “3D-structure models” tab, as are the homology-based models of all single and tandem repeats, which can be visualized with the Jmol visualization tool [45].

The “Mutated dystrophin” pages are dedicated to the 209 in-frame mutations of the human *DMD* gene included in eDystrophin. It is possible to search for mutations according to mutation type, phenotype or the domain involved.

The “Search by mutation type” tool allows the user to search for deletions, duplications or substitutions. If “Deletions” or “Duplications” are selected, a diagram of the exons of the *DMD* gene is displayed (as in Figure 3). This allows the user to select an exon, for which a list of all deletions or



**Figure 3** Screenshot of “Search by mutation type/Deletions” tool. The user can select an exon and then a deletion involving this exon. On the right side of the page, the user can select a mutation of interest from a list, to obtain data about this mutation. The user can also save this list as a csv-format file.

duplications affecting this exon can then be obtained, or to select a mutation of interest from the list at the bottom of the page. If “Substitutions” is selected, a list of mutations is provided. The “Search by phenotype” tool provides a list of mutations “affecting at least one patient” or “affecting no patient” with a specific phenotype. These phenotypes include pending or asymptomatic, BMD, IMD, DMD or DCM. The “Search by involved domain” tool provides a representation of the dystrophin protein, together with its known partners. The names of the structural domains and partners of dystrophin are active buttons. Clicking on these buttons brings up a list of mutations affecting the chosen domain or partner. All of these mutation lists are easy to save down as a csv-formatted data file.

Regardless of the way in which the list of mutations was obtained, clicking on the “see protein” button opens a new page showing a summary table with a global view of the effects of the mutations and four horizontal tabs displaying details (Figure 4A). By clicking on the ‘Clinical data’ tab, the user can obtain the number of patients listed in the database carrying a given mutation, and information about the overall distribution of phenotypes, the severity of the disease in BMD patients, the presence or absence of cardiomyopathy and mental retardation

and the size and amount of dystrophin (when such data are available) (Figure 4B).

Clicking on the “Protein structural domains” tab provides a map of the protein domains modified by the mutation and a table summarizing the consequences of the deletion for the four shorter dystrophin isoforms (Figure 4C). Similarly, clicking on the “Protein binding domains” tab brings up a map of modifications to the binding domains of the mutated protein (Figure 4D). Finally, clicking on the “3D-structure model of the mutation site” tab provides the user with a model of the three-dimensional structure of the regions on either side of the mutation in cases in which exons encoding part of the central rod domain are deleted (Figure 4E). The user can freely download a PDF summary file including the clinical data described above, information about structural and binding domains and the three-dimensional structure model. A final comment about the impact of the mutation on the filamentous structure of the protein is also provided.

## Overview of the content of the database

### Statistics

The 209 mutations recorded in the eDystrophin database include 128 large deletions of one or several exons

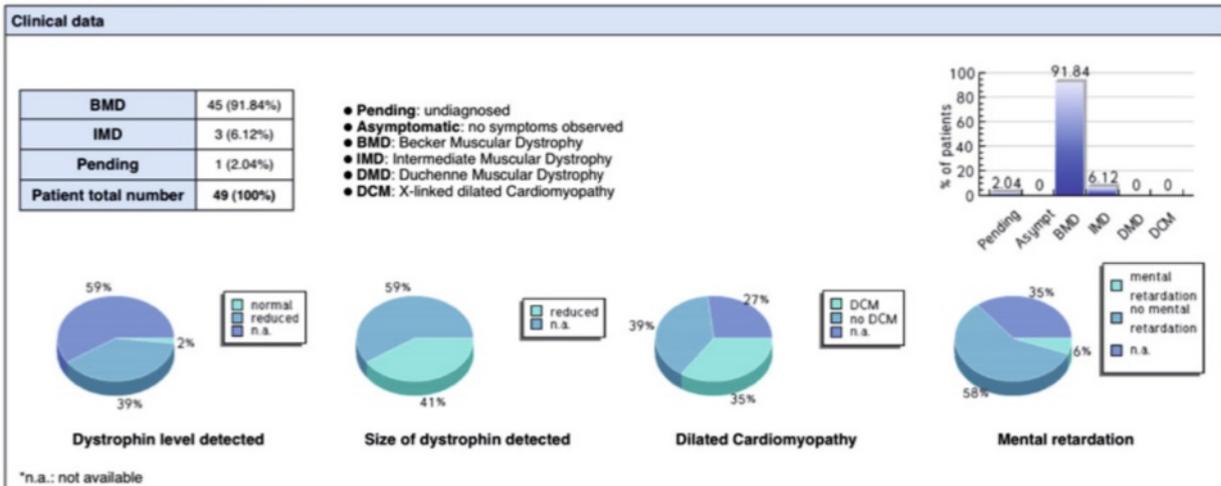
deletion exons 45-49 \*

**A**

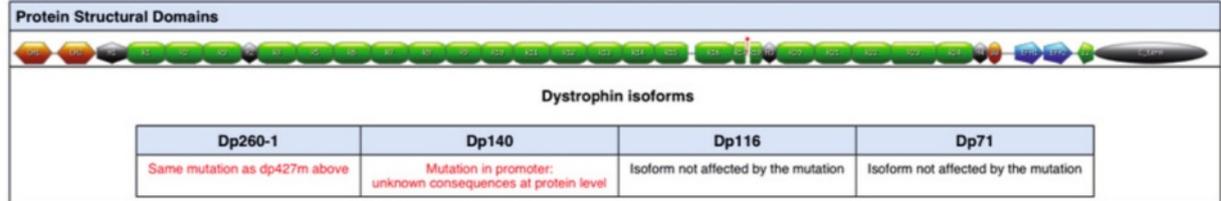
<b>cDNA Mutation</b>	c.6439-7_7200+7del	<b>Protein Mutation</b>	p.Glu2147_Lys2400del
<b>cDNA Size</b>	10293 nt	<b>Protein Size</b>	3431 aa
<b>Phenotype</b>	45 BMD ; 3 IMD ; 1 Pending	<b>Protein Weight</b>	397539 Da

\* mutation found in [Flanigan.2009](#) (BMD) and [Morandi.1995](#) (BMD) and [Beggs.1991](#) (BMD) and [Taylor.2007](#) (BMD) and [Comi.1994](#) (BMD) and [Dent.2005](#) (BMD) and Cochin Hospital, Paris (BMD, IMD, Pending) and [Kaspar.2009](#) (BMD)

**B**



**C**



**D**

**Protein Binding Domains**

<b>K19</b>	present	<b>ABD1</b>	present	<b>LBD1</b>	present	<b>LBD2</b>	partial, center	<b>PAR-1b</b>	present
<b>ABD2</b>	partial, N-term	<b>Synemin</b>	present	<b>nNOS</b>	partial, N-term	<b>Plectin</b>	present	<b>BDG</b>	present
<b>Syntrophin</b>	present	<b>Dystrobrevin</b>	present	<b>Synemin2</b>	present	<b>Myospryn</b>	present	<b>Ankyrin</b>	present

**E**

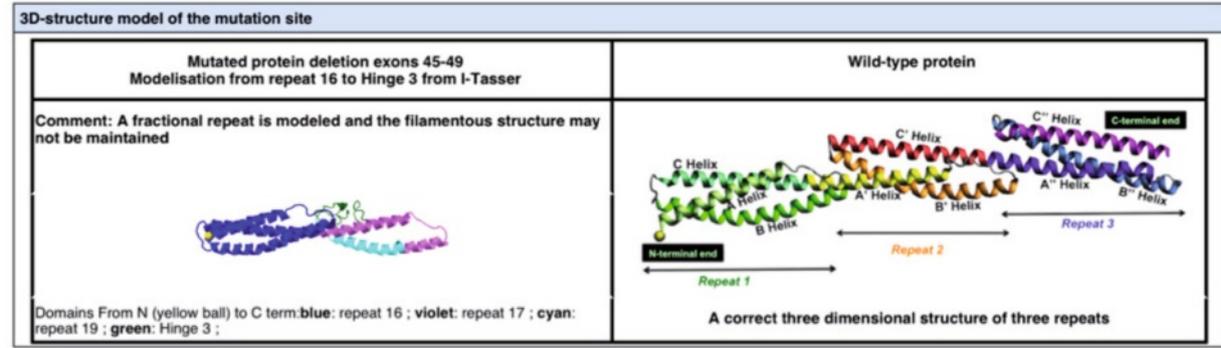


Figure 4 (See legend on next page.)

(See figure on previous page.)

**Figure 4 Screenshot of data available for the deletion of exons 45 to 49.** (A) A general view of the webpage after loading. Description of the mutation at the nucleotide and protein levels, cDNA and protein size, the molecular weight of the mutated protein, a link to cDNA and protein sequences and a list of references reporting patients carrying the mutation are available in table form. Detailed information about clinical data, structural and binding domains and models of three-dimensional structure can be obtained by clicking on the boxes below. (B) The "Clinical data" tab on the first line provides access to a table showing the distribution of phenotypes. In the second line, pie-charts showing the amount and size of dystrophin, as determined by western blotting, the presence of cardiomyopathy and mental retardation are given. (C) The "Structural domains" tab provides a schematic representation of the mutated dystrophin. The sequences of each protein domain are available. (D) The "Binding domains" tab indicates, in red, the changes to the binding domains caused by the mutations listed in the table. (E) The "3D-structure model of the mutation site" tab shows the model of the three-dimensional structure of the domains around the mutation junctions (here R16, R17, R19 and H3). The secondary structure predicted by I-TASSER is displayed above the model. The PDB file and the Ramachandran plot are also available. The modeling tab is available only for the deletion of exons encoding part of the central rod domain. All information can be saved down in the form of PDF files.

(61% of the mutations listed) and 50 large duplications of one or several exons (24% of the mutations). There are also 23 missense mutations (11%), and 8 small in-frame deletions (4%) (Figure 5A).

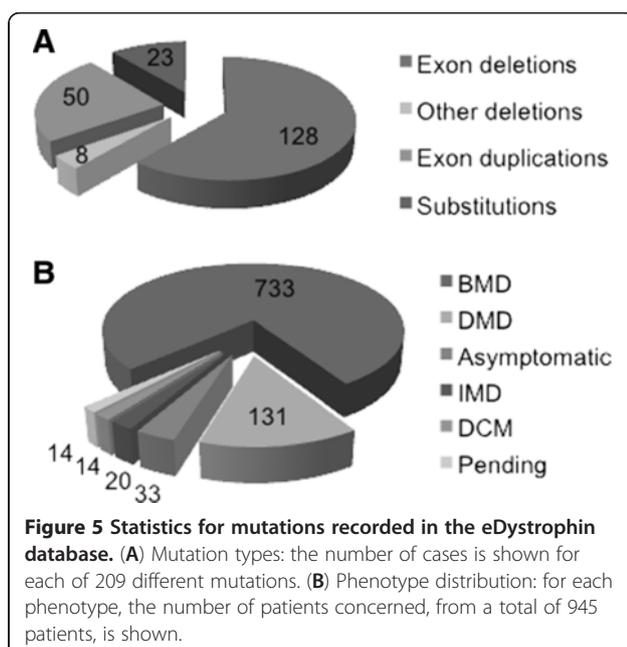
Figure 5B shows the phenotype distribution for all the patients. Of the 945 patients, 733 (78%) had the BMD phenotype, 131 (14%) had the DMD phenotype (most of these patients had mutations towards the 3' end of the *DMD* gene), and 20 patients (2%) had the IMD phenotype. Thus, 16% of the patients registered in the eDystrophin database are exceptions to the Monaco rule [26], as they carry in-frame mutations associated with the severe DMD/IMD phenotypes. Deletions of exons 60 to 79 are generally observed only in DMD patients, but deletions of exons 45–79 and of exon 60 have been reported in at least one BMD patient (Additional file 3 Figure S1A, B). This is consistent with previous reports showing that deletions involving the Cys-rich domain are more deleterious than mutations involving the central rod domain. Similarly, 18

mutations starting at exon 3 have been found in at least one DMD patient, whereas only nine such mutations have been found in at least one BMD patient. As previously reported, deletions affecting ABD1 were generally found to be more deleterious than those affecting the central domain [23,40]. Twenty-nine of the deletions involving the central domain were found in at least one DMD patient, whereas 60 were found in at least one BMD patient. Thirty-one of the duplication mutations were observed in at least one DMD patient (Additional file 3, Figure S1C,D), and 24 were observed in at least one BMD patient. These findings indicate that duplications are generally more deleterious than deletions.

#### Case studies

We illustrate the use of the eDystrophin database for analysis of the consequences of a specific in-frame *DMD* mutation, by studying two exon deletions: deletion of exons 13–44 (c.1483-? 6438 + ?del; p.Val495\_Lys2146del) which is a proximal large deletion with a low frequency, and deletion of exons 45–47 (c.6439-?\_6912 + ?del; p.Glu2147\_Lys2304del), a very frequent, distal, relatively short deletion. The easiest way to proceed is to select one of the mutated exons from the "Mutated dystrophin" page and then to click on the "Search by mutation type"/"deletions" tabs. The desired mutation is chosen and the "see protein" tab is clicked to view the results page.

As shown in the summary at the top of the new page, the deletion of exons 13–44 has been observed in only one BMD patient ("Clinical data" tab). This deletion leads to the production of a 2033-residue dystrophin protein with a molecular weight of 234kDa, which is smaller than its wild-type counterpart (3685 residues and a molecular weight of 427kDa) and was originally reported in [22]. All the data on the page can be downloaded as a pdf file, by clicking on "Get pdf file". The "Protein structural domains" tab shows that the region from the C-terminal part of repeat 2 to the N-terminal part of repeat 17 (including hinge 2), is missing from the mutated protein (Figure 2B). Furthermore, this deletion



affects the promoter of the Dp260 and Dp140 isoforms but has no effect on Dp116 and Dp71. The “Protein binding domains” tab shows that the deletion encompasses the entire PAR-1b and synemin-binding domains and partly modifies the LBD1, LBD2, ABD2, and nNOS binding domains. Clicking on the “3D-structure model of the mutation site” tab brings up the I-TASSER-built three-dimension structural model, showing that a long helix is reconstituted at the mutation site junction between the N-terminal part of repeat 2 and the C-terminal part of repeat 17 (Figure 2B). Thus, the reconstitution of a triple coiled-coil, as in the wild-type three-repeat model, may occur. This I-TASSER model has a C-score of 0.66, indicating a correct fold, as confirmed by Verify3D, ProSA-web and Procheck, all of which indicated a high overall quality for the model. The Ramachandran plot showed that 91.7% of the residues were in the most favored regions, with 1.7% of residues in disallowed regions. These disallowed residues are located in loops, which are often poorly defined. These values are compatible with a crystal structure. As shown in the box, we can therefore conclude that a correct filamentous three-dimensional structure, in the form of a hybrid repeat generated by the concatenation of two truncated repeats, is reconstituted at the new junction. Such a hybrid repeat was hypothesized for the deletion of exon 41–42 in an *in vitro* experiment. The author of the study concerned concluded that this hybrid repeat is viable and has some of the properties of the native repeat [62].

The second case study is that of the deletion of exons 45–47. Unlike the first example, this deletion is very frequent, having been observed in 223 patients (23.6% of the eDystrophin cohort). The resulting protein is 3527 residues long, with a calculated molecular weight of 409kDa. The “Clinical data” tab shows that the observed phenotypes correspond to BMD in 96% of patients, associated cardiomyopathy in 19% of patients, and features suggesting mental retardation in 2% of patients. Dystrophin protein levels were reported to be lower than normal in 30% of the patients and normal or high in 1% of patients, with no data available for the remaining patients. The protein was small in 30% of patients and of normal size in 3% of patients, with no data available for the remaining patients. Thus, data concerning the amount and size of the dystrophin protein are missing for 67% of the patients. The “Protein structural domains” tab shows that the deletion eliminates the C-terminal end of repeat 17 and the N-terminal end of repeat 18 (Figure 2C). Dp260 is deleted, along with Dp427m and the Dp140 promoter is affected by the deletion, but the wild-type Dp116 and Dp71 isoforms are unaffected. The “Protein binding domains” tab shows that the ABD2, LBD2, and nNOS binding domains are partially modified by this deletion. As can be seen by clicking on the “3D-structure model of the mutation site” tab, the junction

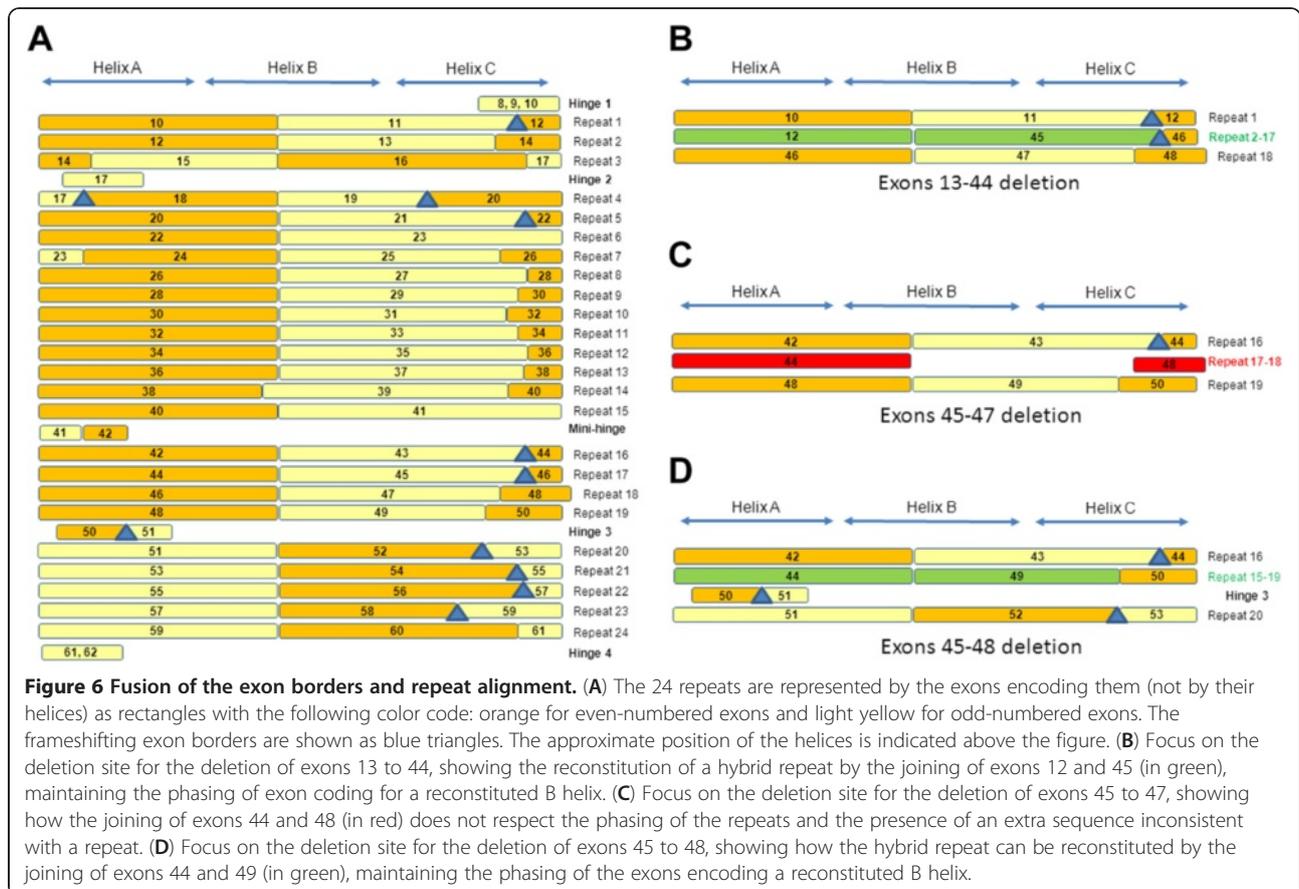
region on either side of the mutation is correctly folded, consistent with the findings of Verify 3D and ProSA-web analysis. The Ramachandran plot shows that 90.2% of the residues are in the most favored regions, and 1.6% residues are in disallowed regions, specifically in loops between helices. However, the model also includes small helices that do not reconstitute a triple coiled-coil (Figure 2C). This implies that no hybrid repeat reconstitution occurs at the junction, with a fractional repeat formed instead and resulting in an incorrect filamentous 3D structure. The term “fractional repeat” has previously been used to describe the joining of two truncated repeats without the reconstitution of a triple coiled-coil similar to that observed in native repeats [62,63]. This conclusion is again displayed in the box.

### Exon phasing versus repeat phasing

The triple coiled-coil structure of the wild-type dystrophin repeats requires the amino-acid sequences to have a seven-residue pattern (the heptad), with apolar residues located alternately three and four residues apart. The  $\alpha$ -helices assemble such that they are tilted and coiled around one another, each in an opposite direction to the other helices. This generates the multi-stranded structure of the coiled-coil [64,65]. Such structures may be formed from two or three helices. Like all spectrin-like repeats, dystrophin repeats consist of three helices. As shown above, exon deletions may or may not allow the reconstitution of a triple coiled-coil at the junction between the sequences on either side of the deletion.

Comparison of the heptad pattern of the repeat alignment obtained by Winder [7] with the exon boundaries showed that all the B helices were encoded by two successive in-frame coding exons (Additional file 3, Figure S1). A simplified diagram of this global organization of exons and repeats in the central rod domain of dystrophin is shown in Figure 6A. Each line represents a repeat or a hinge. The exons encoding the repeats are represented by rectangles, with alternating colors for clarity: even-numbered exons are shown in orange and odd-numbered exons are shown in light yellow. In all the repeats other than repeat 14, the B helices appear to be encoded by two successive in-frame exons with the boundaries precisely aligned with the third heptad of the B helices (Additional file 4, Figure S2).

Thus, an in-frame deletion of two successive exons, the first of which encodes the C-terminal part of a B helix, would be expected to result in deletion of the end of this B helix and of the C helix of the first repeat, together with the beginning of the B helix of the following repeat. This deletion results in a concatenation of the first part of the B helix of the first repeat with the second part of the B helix of the second repeat. Consequently, the heptad pattern remains similar to that in wild-type dystrophin, and



the domain can fold like a native repeat, thereby constituting a hybrid repeat. This observation confirms the previous report of an observation of a hybrid repeat [62]. A similar phenomenon may occur for many deletions involving two such exons or multiples of two exons. In situations in which the formation of a hybrid repeat is not possible, fractional repeats, in which the heptad pattern of alignment is not respected at the new junction between the sequences on either side of the deletion, may form, as previously suggested [63].

### Discussion

The eDystrophin database is a new biomedical resource for clinicians and researchers working on human dystrophin diseases. This dedicated database for the dystrophin protein specifically aims to provide information about in-frame *DMD* mutations and their consequences for the dystrophin protein. It provides a framework for the analysis of such mutations, by presenting a large body of information for both wild-type and mutated dystrophin proteins, including findings relating to the structure of these proteins and their interactions with known partners. Although eDystrophin is a locus-specific database, it was not constructed with an existing database

system, such as LOVD [66,67] or UMD [68,69]. Indeed, such systems are more useful for DNA variant databases and are not suitable for the construction of a protein-based database like eDystrophin.

In human dystrophin diseases, the ratio of the frequency of DMD to that of BMD is approximately 2/3 – 1/3 [40,70]. Most cases of DMD are caused by frame-shift mutations, whereas BMD is generally caused by in-frame mutations, although exceptions have been reported [26]. Documented cases of in-frame mutations are largely underrepresented in existing databases, and the primary aim of the eDystrophin database project was to redress the balance, by developing a dedicated information source for in-frame mutations. Unlike frame-shift mutations, in-frame mutations lead to the production of proteins with various degrees of functionality. The secondary goal of eDystrophin was therefore to determine and show the predicted consequences of these mutations for the composition and structure of the encoded proteins and their clinical consequences. In this first version of eDystrophin, patients and in-frame mutations were obtained from one of the major French contributors to the UMD-DMD database and from published studies. Evidently, the database could be expanded in the near future by

including mutations and patients from around the world, which would probably yield more accurate phenotype-genotype correlations.

However, dystrophin is a large protein, and it is a challenge to investigate the consequences of mutations of its gene. The dystrophin protein has two principal roles: as a scaffolding protein for several interacting partners and as a filamentous protein with a mechanical and structural function, providing resistance to the stress of muscle contraction [8,13]. Any mutation altering the structure of dystrophin may therefore affect both these functions (and potentially other minor functions of the protein as well) simultaneously. Our database provides an overview of the effects of mutations on protein function. In particular, it provides the user with information about changes to interactions and about the maintenance or disruption of the filamentous structure of the mutated dystrophin protein.

Several binding partners of dystrophin have been identified, and the eDystrophin database infers changes to their binding to a mutated dystrophin variant by considering whether the interacting domains remain intact and unmodified. Based on these inferences and previous observations, deletions affecting the Cys-rich or ABD1 domains appear to be much more deleterious than those affecting the central domain [23,40]. However, we detected several mutations affecting the central rod domain and causing a DMD phenotype in a substantial number of patients. In these patients, mRNA levels may have been low and/or unstable, accounting for the presence of little or no protein. Careful re-examination of the boundaries of the mutation is also necessary for these patients. Indeed, Taylor *et al.* (2007, PhD thesis) re-examined a large cohort of DMD patients with in-frame deletions affecting the central rod domain and found that most were frame-shift mutations, consistent with the Monaco reading frame rule. Furthermore, we cannot entirely exclude the possibility of two mutations occurring in the same gene. For the other DMD patients carrying in-frame mutations, uncertainties remain concerning the levels or stability of the corresponding mRNA.

We obtained models of the three-dimensional structure of the new junctions created between the sequences on either side of the deletions in the central rod domain, as previously described [53]. The database provides a computational model for each in-frame deletion collected. An analysis of the structural features of these new junctions showed that two outcomes were possible: the reconstitution of a hybrid repeat and the formation of a fractional repeat in situations in which it was not possible to form a hybrid repeat. The likelihood of hybrid repeat formation depends on the phasing of the exon boundaries with the center of the B helix of the

repeats. The reconstitution of a hybrid repeat can be assumed to occur because the major factor controlling this folding pattern is the presence of a heptad pattern. As this pattern is respected in cases in which the deletion creates a new junction between the first half of one B helix and the second half of the next, from two truncated repeats, coiled-coil folding similar to that in native repeats would be expected [64,65]. By contrast, in fractional repeats, the  $\alpha$ -helices can fold correctly, but the heptad pattern is not respected and a three-dimensional coiled-coil structure therefore cannot be obtained. This may result in a less stable deletion site than for native and hybrid repeats. The hypothesis that repeats phasing in truncated dystrophins is essential to ensure a high level of protein function has already been tested. Transgenic *mdx* mice were produced with several types of truncated dystrophin, some with correct and others with incorrect phasing of the repeats. However, in this previous study, native repeats were either entirely conserved or entirely lost [71]. These findings led to the “mini-dystrophin” concept for DeltaH2-R19, in which the rod domain was decreased in size by a deletion encompassing the amino acids from hinge 2 to repeat 19. By contrast the “micro-dystrophin” DeltaR4-R23 had a deletion extending from repeat 4 to repeat 23. Constructs encoding these proteins proved to be among the best therapeutic constructs for *mdx* mouse rescue. In BMD patients, phasing is not as described in these previous experiments and only hybrid repeats may be reconstituted. However, the demonstration of beneficial effects of phasing in the *mdx* mouse suggests that the presence of hybrid repeats may be associated with a milder phenotype than the presence of fractional repeats [62,63]. Such a correlation between the structural features of mutated dystrophin and clinical severity in a cohort of BMD patients has been reported for cardiomyopathy [24]. The authors constructed models of the mutated dystrophin for deletions involving exons 45 to 49 and investigated the phasing of spectrin repeats. They concluded that the absence of hinge 3 delayed the onset of dilated cardiomyopathy.

However, it should be stressed that the presence of a hybrid repeat does not itself imply a better conservation of dystrophin function than the presence of a fractional repeat. Indeed, mRNA instability or changes to protein-protein interactions may also affect the function of the mutated dystrophin, and it is not currently possible to predict these effects. Investigations of the correlation between the presence of a hybrid repeat and the severity of clinical symptoms are now required. However, the eDystrophin database can be used as a predictive tool for exon skipping-based therapy. The choice of the exon to be deleted to restore the reading frame could be based on careful consideration of the likelihood of reconstituting a hybrid repeat.

## Conclusions

The eDystrophin database is a new tool providing an overview of the proteins generated by *DMD* genes carrying in-frame mutations. It provides information about the consequences of these mutations for protein production and folding and for phenotype-genotype correlations. This database, through these features, is thus a valuable tool for predicting the efficacy of exon-skipping therapy for DMD patients.

## Supporting data

The supporting datasets are provided within the article and the additional files.

## Additional files

**Additional file 1: Table S1.** provides the references from which information about the structural and binding domains of dystrophin described in eDystrophin was obtained [7,51,52,72–90].

**Additional file 2: Table S2.** provides the origin of the mutations described in the eDystrophin database [18,20,22,24,91–102].

**Additional file 3: Figure S1.** Statistics for mutations included in eDystrophin. Exon deletions (A, B) and duplications (C, D) associated with at least one DMD (A, C) or one BMD (B, D) patient. Each line represents a type of exon deletion (A and B) or duplication (C and D).

**Additional file 4: Figure S2.** Sequence alignment of the 24 spectrin-like repeats of dystrophin. Repeats were aligned by ClustalW, using default parameters, as described by Winder *et al.* [7]. The alignment was visualized in Jalview. In the first line, heptad motifs are indicated, showing the hydrophobic residues in the (a) and (d) positions. The repeat numbers and the number of residues per repeat are indicated at the start of the repeat sequence. The presence of absence of hinges 2 and 3 is indicated in separate lines. There are two extra sequences at the ends of repeats 15 and 18 not aligned with the heptad pattern. The presence of these sequences is indicated at the end of the corresponding lines. At the end of each line, the numbers of the exons encoding the repeat are indicated in parentheses. The repeat sequences are highlighted alternately in orange (even-numbered exons) and light yellow (odd-numbered exons). The rectangle indicates the middle of the B helices of the repeats. If the successive exons are not in frame, the residues are shown in red. The alignment reported by Koenig *et al.* [76] is also mentioned: the residues at the start of the repeats are underlined when they differ from Winder's alignment.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

ELR, FBH, RBY and JC initiated and supervised the project. AN, CLM and FBH created the database. AN, RBY, FL and ELR monitored data collection. AN and ELR created and analyzed the structural models. ALL authors participated in the writing of the manuscript and approved its submission.

## Author's information

Frédérique Barloy-Hubler: IFR GFAS - <http://ifr140.univ-rennes1.fr/plates-formes/Amadeus/>.

## Acknowledgments

We thank the "Association Française contre les Myopathies" and "IFR 140" for providing funding for this work. We also thank the referring physicians both within and outside the French CORNEMUS network (*Coordination des centres de Référence Neuromusculaires*), for their valuable collaboration and for providing the clinical information used. We thank Michael Buckley for interesting discussions and for providing us Peter Taylor's thesis. We also

thank the Prosite team for modifying the MyDomains tool for use in our database, and Biogenouest, which hosts the database. AN was supported by the CNRS.

## Author details

<sup>1</sup>Université de Rennes 1 Rennes France. <sup>2</sup>UMR CNRS 6026, Equipe RMN-ILP, Faculté de Médecine, CS 34317 Rennes Cedex 35043, France. <sup>3</sup>Université Européenne de Bretagne 5, Boulevard Laënnec, Rennes 35000, France. <sup>4</sup>UMR CNRS 6026, Equipe SP@RTE Campus Beaulieu, Rennes cedex 35042, France. <sup>5</sup>Département de recherche clinique, Institut de Myologie, GH Pitié-Salpêtrière Paris, France. <sup>6</sup>Laboratoire de Biochimie et Génétique Moléculaire-Hôpital Cochin, Paris France. <sup>7</sup>Institut Cochin, CNRS UMR 8104, Inserm U 1016, Université Paris Descartes Paris, France. <sup>8</sup>Faculté de Médecine Paris Descartes, CHU Cochin Paris France. <sup>9</sup>IFR 140 Biosit, Plate-forme Amadeus, Université de Rennes1 Rennes France. <sup>10</sup>UMR CNRS 6290, Institut de Génétique et Développement, CS 34317 Rennes cedex 35043, France.

Received: 11 April 2012 Accepted: 18 June 2012

Published: 9 July 2012

## References

1. Mendelian Inheritance in Man. [http://www.ncbi.nlm.gov/omim]
2. Hoffman EP, Brown RH, Kunkel LM: **Dystrophin: the protein product of the Duchenne muscular dystrophy locus.** *Cell* 1987, **51**:919–928.
3. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM: **Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals.** *Cell* 1987, **50**:509–517.
4. Roberts RG, Bobrow M: **Dystrophins in vertebrates and invertebrates.** *Hum Mol Genet* 1998, **7**:589–595.
5. Segalat L: **Dystrophin and functionally related proteins in the nematode *Caenorhabditis elegans*.** *Neuromuscul Disord* 2002, **12**(Suppl 1):S105–S109.
6. Koenig M, Monaco AP, Kunkel LM: **The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein.** *Cell* 1988, **53**:219–226.
7. Winder SJ, Gibson TJ, Kendrick-Jones J: **Dystrophin and utrophin: the missing links!** *FEBS Lett* 1995, **369**:27–33.
8. Le Rumeur E, Winder SJ, Hubert JF: **Dystrophin: More than just the sum of its parts.** *Biochim Biophys Acta* 2010, **1804**:1713–1722.
9. Blake DJ, Weir A, Newey SE, Davies KE: **Function and genetics of dystrophin and dystrophin-related proteins in muscle.** *Physiol Rev* 2002, **82**:291–329.
10. Campbell K, Kahl S: **Association of dystrophin and an integral membrane glycoprotein.** *Nature* 1989, **338**:259–262.
11. Chamberlain JS, Corrado K, Rafael JA, Cox GA, Hausser M, Lumeng C: **Interactions between dystrophin and the sarcolemma membrane.** *Soc Gen Physiol Ser* 1997, **52**:19–29.
12. Petrof BJ, Shrager JB, Stedmann HH, Kelly AM, Sweeney HL: **Dystrophin protects the sarcolemma from stresses developed during muscle contraction.** *Proc Natl Acad Sci, USA* 1993, **90**:3710–3714.
13. Deconinck N, Dan B: **Pathophysiology of duchenne muscular dystrophy: current hypotheses.** *Pediatr Neurol* 2007, **36**:1–7.
14. Batchelor CL, Winder SJ: **Sparks, signals and shock absorbers: how dystrophin loss causes muscular dystrophy.** *Trends Cell Biol* 2006, **16**:198–205.
15. Desguerre I, Christov C, Mayer M, Zeller R, Becane HM, Bastuji-Garin S, Leturcq F, Chiron C, Chelly J, Gherardi RK: **Clinical heterogeneity of duchenne muscular dystrophy (DMD): definition of sub-phenotypes and predictive criteria by long-term follow-up.** *PLoS One* 2009, **4**:e4347.
16. Koenig M, Beggs A, Moyer M, Scherpf S, Heindrich K, Bettecken T, Meng G, Muller C, Lindlof M, Kaariainen H: **The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion.** *Am J Hum Genet* 1989, **45**:498–506.
17. Magri F, Govoni A, D'Angelo MG, Del Bo R, Ghezzi S, Sandra G, Turconi AC, Sciacco M, Ciscato P, Bordoni A, *et al*: **Genotype and phenotype characterization in a large dystrophinopathic cohort with extended follow-up.** *J Neurol* 2011, **258**:1610–1623.
18. Beggs A, Hoffman E, Snyder J, Arahata K, Specht L, Shapiro F, Angelini C, Sugita H, Kunkel L: **Exploring the molecular basis for variability among patients with Becker muscular dystrophy: dystrophin gene and protein studies.** *Am J Hum Genet* 1991, **49**:54–67.
19. Bushby KM, Gardner-Medwin D, Nicholson LV, Johnson MA, Haggerty ID, Cleghorn NJ, Harris JB, Bhattacharya SS: **The clinical, genetic and**

- dystrophin characteristics of Becker muscular dystrophy. II. Correlation of phenotype with genetic and protein abnormalities. *J Neurol* 1993, **240**:105–112.
20. Comi GP, PELLE A, Bresolin N, Moggio M, Bardoni A, Gallanti A, Vita G, Toscano A, Ferro MT, Bordoni A, et al: **Clinical variability in Becker muscular dystrophy. Genetic, biochemical and immunohistochemical correlates.** *Brain* 1994, **117**(Pt 1):1–14.
  21. Nigro V, Nigro G, Esposito MG, Comi LI, Molinari AM, Puca GA, Politano L: **Novel small mutations along the DMD/BMD gene associated with different phenotypes.** *Hum Mol Genet* 1994, **3**:1907–1908.
  22. Morandi L, Mora M, Confalonieri V, Barresi R, Di Blasi C, Brugnioni R, Bernasconi P, Mantegazza R, Dworzak F, Antozzi C, et al: **Dystrophin characterization in BMD patients: correlation of abnormal protein with clinical phenotype.** *J Neurol Sci* 1995, **132**:146–155.
  23. Beroud C, Tuffery-Giraud S, Matsuo M, Hamroun D, Humbertclaude V, Monnier N, Moizard MP, Voelckel MA, Caemard LM, Boisseau P, et al: **Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63 % of patients with Duchenne muscular dystrophy.** *Hum Mutat* 2007, **28**:196–202.
  24. Kaspar RW, Allen HD, Ray WC, Alvarez CE, Kissel JT, Pestronk A, Weiss RB, Flanigan KM, Mendell JR, Montanaro F: **Analysis of dystrophin deletion mutations predicts age of cardiomyopathy onset in becker muscular dystrophy.** *Circ Cardiovasc Genet* 2009, **2**:544–551.
  25. Cunniff C, Andrews J, Meaney FJ, Mathews KD, Matthews D, Cafaloni E, Miller TM, Bodensteiner JB, Miller LA, James KA, et al: **Mutation analysis in a population-based cohort of boys with Duchenne or Becker muscular dystrophy.** *J Child Neurol* 2009, **24**:425–430.
  26. Monaco A, Bertelson C, Liechti-Gallati S, Moser H, Kunkel L: **An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus.** *Genomics* 1988, **2**:90–95.
  27. Aartsma-Rus A, Van Deutekom JC, Fokkema IF, Van Ommen GJ, Den Dunnen JT: **Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule.** *Muscle Nerve* 2006, **34**:135–144.
  28. Winnard AV, Klein CJ, Covert DD, Prior T, Papp A, Snyder P, Bulman DE, Ray PN, McAndrew P, King W, et al: **Characterization of translational frame exception patients in Duchenne/Becker muscular dystrophy.** *Hum Mol Genet* 1993, **2**:737–744.
  29. Hattori N, Kaido M, Nishigaki T, Inui K, Fujimura H, Nishimura T, Naka T, Hazama T: **Undetectable dystrophin can still result in a relatively benign phenotype of dystrophinopathy.** *Neuromuscul Disord* 1999, **9**:220–226.
  30. Tuffery-Giraud S, Saquet C, Thorel D, Disset A, Rivier F, Malcolm S, Claustres M: **Mutation spectrum leading to an attenuated phenotype in dystrophinopathies.** *Eur J Hum Genet* 2005, **13**:1254–1260.
  31. Gualandi F, Neri M, Bovolenta M, Martoni E, Rimessi P, Fini S, Spitali P, Fabris M, Pane M, Angelini C, et al: **Transcriptional behavior of DMD gene duplications in DMD/BMD males.** *Hum Mutat* 2009, **30**:E310–E319.
  32. Anthony K, Cirak S, Torelli S, Tasca G, Feng L, Arechavala-Gomez V, Armaroli A, Guglieri M, Straathof CS, Verschuuren JJ, et al: **Dystrophin quantification and clinical correlations in Becker muscular dystrophy: implications for clinical trials.** *Brain* 2011, **134**:3544–3556.
  33. van Deutekom JC, Janson AA, Ginjaar IB, Frankhuizen WS, Aartsma-Rus A, Bremmer-Bout M, den Dunnen JT, Koop K, van der Kooij AJ, Goemans NM, et al: **Local dystrophin restoration with antisense oligonucleotide PRO051.** *N Engl J Med* 2007, **357**:2677–2686.
  34. Kinali M, Arechavala-Gomez V, Feng L, Cirak S, Hunt D, Adkin C, Guglieri M, Ashton E, Abbs S, Nihoyannopoulos P, et al: **Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study.** *Lancet Neurol* 2009, **8**:918–928.
  35. Cirak S, Arechavala-Gomez V, Guglieri M, Feng L, Torelli S, Anthony K, Abbs S, Garraza ME, Bourke J, Wells DJ, et al: **Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study.** *Lancet* 2011, **378**:595–605.
  36. Goemans NM, Tulinius M, van den Akker JT, Burm BE, Ekhart PF, Heuvelmans N, Holling T, Janson AA, Platenburg GJ, Sipkens JA, et al: **Systemic administration of PRO051 in Duchenne's muscular dystrophy.** *N Engl J Med* 2011, **364**:1513–1522.
  37. *Leiden Muscular Dystrophy pages.* [http://www.dmd.nl/nmdb/home.php?select\_db=DMD]
  38. White SJ, den Dunnen JT: **Copy number variation in the genome; the human DMD gene as an example.** *Cytogenet Genome Res* 2006, **115**:240–246.
  39. *UMD-DMD France.* [http://www.umd.be/DMD/W\_DMD/index.html]
  40. Tuffery-Giraud S, Beroud C, Leturcq F, Yao RB, Hamroun D, Michel-Calemard L, Moizard MP, Bernard R, Cossee M, Boisseau P, et al: **Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase.** *Hum Mutat* 2009, **30**:934–945.
  41. Campbell KP: **Three muscular dystrophies: loss of cytoskeleton-extracellular matrix linkage.** *Cell* 1995, **80**:675–679.
  42. *MAMP.* [http://www.mamp.info/en/index.html]
  43. *BioGenouest Platform.* [http://www.genouest.org]
  44. *MyDomains.* [http://prosite.expasy.org/mydomains/]
  45. *Jmol.* [http://www.jmol.org]
  46. Chelly J, Hamard G, Koulakoff A, Kaplan JC, Kahn A, Berwald-Netter Y: **Dystrophin gene transcribed from different promoters in neuronal and glial cells.** *Nature* 1990, **344**:64–65.
  47. D'Souza VN, Nguyen TM, Morris GE, Karges W, Pillers DA, Ray PN: **A novel dystrophin isoform is required for normal retinal electrophysiology.** *Hum Mol Genet* 1995, **4**:837–842.
  48. Lidov HG, Selig S, Kunkel LM: **Dp140: a novel 140 kDa CNS transcript from the dystrophin locus.** *Hum Mol Genet*, **4**:329–335.
  49. Byers TJ, Lidov HG, Kunkel LM: **An alternative dystrophin transcript specific to peripheral nerve.** *Nat Genet*, **4**:77–81.
  50. Hugnot JP, Gilgenkrantz H, Vincent N, Chafey P, Morris GE, Monaco AP, Berwald-Netter Y, Koulakoff A, Kaplan JC, Kahn A, et al: **Distal transcript of the dystrophin gene initiated from an alternative first exon and encoding a 75-kDa protein widely distributed in nonmuscle tissues.** *Proc Natl Acad Sci U S A*, **89**:7506–7510.
  51. Norwood F, Sutherland-Smith A, Keep N, Kendrick-Jones J: **The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy.** *Structure* 2000, **8**:481–491.
  52. Huang X, Poy F, Zhang R, Joachimiak A, Sudol M, Eck MJ: **Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan.** *Nat Struct Biol* 2000, **7**:634–638.
  53. Legrand B, Giudice E, Nicolas A, Delalande O, LeRumeur E: **Computational study of the human dystrophin repeats: interaction properties and molecular dynamics.** *PLoS One* 2011, **6**:e23819.
  54. *Human Genome Variation Society.* [http://www.hgvs.org/].
  55. Roy A, Kucukural A, Zhang Y: **I-TASSER: a unified platform for automated protein structure and function prediction.** *Nat Protoc* 2010, **5**:725–738.
  56. Zhang Y: **I-TASSER server for protein 3D structure prediction.** *BMC Bioinforma* 2008, **9**:40.
  57. DeLano WL: *(Ed): The PyMOL user's manual.* San Carlos, CA: DeLano Scientific; 2002.
  58. Sippl MJ: **Recognition of errors in three-dimensional structures of proteins.** *Proteins* 1993, **17**:355–362.
  59. Wiederstein M, Sippl MJ: **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.** *Nucleic Acids Res* 2007, **35**:W407–W410.
  60. Bowie JU, Luthy R, Eisenberg D: **A method to identify protein sequences that fold into a known three-dimensional structure.** *Science* 1991, **253**:164–170.
  61. Luthy R, Bowie JU, Eisenberg D: **Assessment of protein models with three-dimensional profiles.** *Nature* 1992, **356**:83–85.
  62. Menhart N: **Hybrid spectrin type repeats produced by exon-skipping in dystrophin.** *Biochim Biophys Acta* 2006, **1764**:993–999.
  63. Yokota T, Duddy W, Patridge T: **Optimizing exon skipping therapies for DMD.** *Acta Myologica* 2007, **26**:179–184.
  64. Lupas A: **Coiled coils: new structures and new functions.** *Trends Biochem Sci* 1996, **21**:375–382.
  65. Parry DAD, Fraser RDB, John M, Squire JM: **Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure.** *J Struct Biol* 2008, **163**:258–269.
  66. *Leiden Open Variation Database.* [http://www.lovd.nl/2.0/]
  67. Fokkema IF, den Dunnen JT, Taschner PE: **LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach.** *Hum Mutat* 2005, **26**:63–68.

68. UMD. [http://www.umd.be/]
69. Beroud C, Hamroun D, Colod-Beroud G, Boileau C, Soussi T, Claustres M: **UMD (Universal Mutation Database): 2005 update.** *Hum Mutat* 2005, **26**:184–191.
70. Bushby KM, Thambyayah M, Gardner-Medwin D: **Prevalence and incidence of Becker muscular dystrophy.** *Lancet* 1991, **337**:1022–1024.
71. Harper SQ, Hauser MA, DelloRusso C, Duan D, Crawford RW, Phelps SF, Harper HA, Robinson AS, Engelhardt JF, Brooks SV, Chamberlain JS: **Modular flexibility of dystrophin: implications for gene therapy of Duchenne muscular dystrophy.** *Nature Med* 2002, **8**:253–261.
72. Banuelos S, Saraste M, Djinić Carugo K: **Structural comparisons of calponin homology domains: implications for actin binding.** *Structure* 1998, **6**(11):1419–1431.
73. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, *et al*: **CDD: specific functional annotation with the Conserved Domain Database.** *Nucleic Acids Res* 2009, **37**(Database issue):D205–D210.
74. Letunic I, Doerks T, Bork P: **SMART 6: recent updates and new developments.** *Nucleic Acids Res* 2009, **37**(Database issue):D229–D232.
75. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin QL, Gunasekaran P, Ceric G, Forslund K, *et al*: **The Pfam protein families database.** *Nucleic Acids Res* 2010, **38**(Database issue):D211–D222.
76. Koenig M, Kunkel LM: **Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility.** *J Biol Chem* 1990, **265**(8):4560–4566.
77. Ishikawa-Sakurai M, Yoshida M, Imamura M, Davies KE, Ozawa E: **ZZ domain is essentially required for the physiological binding of dystrophin and utrophin to beta-dystroglycan.** *Hum Mol Genet* 2004, **13**(7):693–702.
78. Hnia K, Zouiten D, Cantel S, Chazalotte D, Hugon G, Fehrentz JA, Masmoudi A, Diment A, Bramham J, Mornet D, Winder SJ: **ZZ domain of dystrophin and utrophin: topology and mapping of a beta-dystroglycan interaction site.** *Biochem J* 2007, **3**(3):667–677. 3.
79. Koenig M, Monaco AP, Kunkel LM: **The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein.** *Cell* 1988, **53**(2):219–228.
80. Stone MR, O'Neill A, Catino D, Bloch RJ: **Specific interaction of the actin-binding domain of dystrophin with intermediate filaments containing keratin 19.** *Mol Biol Cell* 2005, **16**(9):4280–4293.
81. Legardinier S, Raguene-Nicol C, Tascon C, Rocher C, Hardy S, Hubert JF, Le Rumeur E: **Mapping of the lipid-binding and stability properties of the central rod domain of human dystrophin.** *J Mol Biol* 2009, **389**(3):546–558.
82. Yamashita K, Suzuki A, Satoh Y, Ide M, Amano Y, Masuda-Hirata M, Hayashi YK, Hamada K, Ogata K, Ohno S: **The 8th and 9th tandem spectrin-like repeats of utrophin cooperatively form a functional unit to interact with polarity-regulating kinase PAR-1b.** *Biochem Biophys Res Commun* 2010, **391**(1):812–817.
83. Amann KJ, Renley BA, Ervasti JM: **A cluster of basic repeats in the dystrophin rod domain binds F-actin through an electrostatic interaction.** *J Biol Chem* 1998, **273**(43):28419–28423.
84. Bhosle RC, Michele DE, Campbell KP, Li Z, Robson RM: **Interactions of intermediate filament protein synemin with dystrophin and utrophin.** *Biochem Biophys Res Commun* 2006, **346**(3):768–777.
85. Lai Y, Thomas GD, Yue Y, Yang HT, Li D, Long C, Judge L, Bostick B, Chamberlain JS, Terjung RL, Duan D: **Dystrophins carrying spectrin-like repeats 16 and 17 anchor nNOS to the sarcolemma and enhance exercise performance in a mouse model of muscular dystrophy.** *J Clin Invest* 2009, **119**(3):624–635.
86. Rezniczek GA, Konieczny P, Nikolic B, Reipert S, Schneller D, Abrahamsberg C, Davies KE, Winder SJ, Wiche G: **Plectin 1f scaffolding at the sarcolemma of dystrophic (mdx) muscle fibers through multiple interactions with betadystroglycan.** *J Cell Biol* 2007, **176**(7):965–977.
87. Newey SE, Benson MA, Ponting CP, Davies KE, Blake DJ: **Alternative splicing of dystrobrevin regulates the stoichiometry of syntrophin binding to the dystrophin protein complex.** *Curr Biol* 2000, **10**(20):1295–1298.
88. Sadoulet-Puccio HM, Rajala M, Kunkel LM: **Dystrobrevin and dystrophin: an interaction through coiled-coil motifs.** *Proc Natl Acad Sci U S A* 1997, **94**(23):12413–12418.
89. Reynolds JG, McCalmon SA, Donaghey JA, Naya FJ: **Deregulated protein kinase A signaling and myospryn expression in muscular dystrophy.** *J Biol Chem* 2008, **283**(13):8070–8074.
90. Ayala G, Davis JQ, Scotland PB, Bennett V: **An ankyrin-based mechanism for functional organization of dystrophin and dystroglycan.** *Cell* 2008, **135**(7):1189–1200.
91. Flanigan KM, Dunn DM, von Niederhausern A, Soltanzadeh P, Gappmaier E, Howard MT, Sampson JB, Mendell JR, Wall C, King WM, *et al*: **Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort.** *Hum Mutat* 2009, **30**(12):1657–1666.
92. Taylor PJ, Maroulis S, Mullan GL, Pedersen RL, Baumli A, Elakis G, Piras S, Walsh C, Prosper-Gutierrez B, De La Puente-Alonso F, *et al*: **Measurement of the clinical utility of a combined mutation detection protocol in carriers of Duchenne and Becker muscular dystrophy.** *J Med Genet* 2007, **44**(6):368–372.
93. Dent KM, Dunn DM, von Niederhausern AC, Aoyagi AT, Kerr L, Bromberg MB, Hart KJ, Tuohy T, White S, den Dunnen JT, *et al*: **Improved molecular diagnosis of dystrophinopathies in an unselected clinical cohort.** *Am J Med Genet A* 2005, **134**(3):295–298.
94. Taylor PJ, Betts GA, Maroulis S, Gilissen C, Pedersen RL, Mowat DR, Johnston HM, Buckley MF: **Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy.** *PLoS One* 2009, **5**(1):8803.
95. Nevo Y, Muntoni F, Sewry C, Legum C, Kutai M, Harel S, Dubowitz V: **Large in-frame deletions of the rod-shaped domain of the dystrophin gene resulting in severe phenotype.** *Isr Med Assoc J* 2003, **5**(2):94–97.
96. Carsana A, Frisso G, Tremolaterra MR, Lanzillo R, Vitale DF, Santoro L, Salvatore F: **Analysis of dystrophin gene deletions indicates that the hinge III region of the protein correlates with disease severity.** *Ann Hum Genet* 2005, **69**(Pt 3):253–259.
97. Tsukamoto H, Inui K, Fukushima H, Nishigaki T, Taniike M, Tanaka J, Okada S: **Molecular study of Duchenne and Becker muscular dystrophies in Japanese.** *J Inher Metab Dis* 1991, **14**(5):819–824.
98. Janssen B, Hartmann C, Scholz V, Jauch A, Zschocke J: **MLPA analysis for the detection of deletions, duplications and complex rearrangements in the dystrophin gene: potential and pitfalls.** *Neurogenetics* 2005, **6**(1):29–35.
99. Buzin CH, Feng J, Yan J, Scaringe W, Liu Q, den Dunnen J, Mendell JR, Sommer SS: **Mutation rates in the dystrophin gene: a hotspot of mutation at a CpG dinucleotide.** *Hum Mutat* 2005, **25**(2):177–188.
100. Ramelli GP, Joncourt F, Luetsch J, Weis J, Tolnay M, Burgunder JM: **Becker muscular dystrophy with marked divergence between clinical and molecular genetic findings: case series.** *Swiss Med Wkly* 2006, **136**(11–12):189–193.
101. Fajkusova L, Lukas Z, Tvrdikova M, Kuhrova V, Hajek J, Fajkus J: **Novel dystrophin mutations revealed by analysis of dystrophin mRNA: alternative splicing suppresses the phenotypic effect of a nonsense mutation.** *Neuromuscul Disord* 2001, **11**(2):133–138.
102. Prior TW, Bridgeman SJ: **Experience and strategy for the molecular testing of Duchenne muscular dystrophy.** *J Mol Diagn* 2005, **7**(3):317–326.

doi:10.1186/1750-1172-7-45

**Cite this article as:** Nicolas *et al.*: Assessment of the structural and functional impact of the in-frame mutations of the *DMD* gene, using the tools included in the eDystrophin online database. *Orphanet Journal of Rare Diseases* 2012 **7**:45.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit





La dystrophine est une protéine impliquée dans les myopathies de Duchenne (DMD) et de Becker (BMD). Malgré les nombreux travaux de recherche fondamentale et clinique effectués sur ces pathologies, le rôle moléculaire précis de la dystrophine est largement méconnu. A ce jour, la corrélation entre génotype et degré de gravité de ces pathologies est difficile à établir. Cette connaissance s'avère cependant indispensable au développement de nouvelles thérapies. L'objet de cette thèse est l'étude bioinformatique de la fonctionnalité des formes mutées de dystrophine exprimées chez les patients BMD avec l'objectif d'une mise en corrélation avec les phénotypes cliniques. Nous avons dans un premier temps créé la base de données *eDystrophin* qui apporte des informations sur l'ensemble des mutations BMD en lien avec les phénotypes associés. Des modèles obtenus par homologie ont été construits et une vue d'ensemble des conséquences de chaque mutation sur les propriétés d'interactions et la structure 3D de la protéine est également proposée. La majorité des mutations BMD répertoriées sont des délétions d'exons entiers, ayant pour conséquences des troncatures d'une partie du domaine central de la dystrophine composé de 24 répétitions homologues à la spectrine. Dans la deuxième partie de la thèse, l'utilisation combinée des informations répertoriées dans *eDystrophin*, et des résultats de modélisation et de dynamique moléculaire a permis de mettre en évidence l'existence de deux catégories de protéines tronquées potentiellement présentes chez les patients BMD : 1) des protéines portant à la jonction de la délétion des répétitions dites « hybrides » qui conservent une structure en faisceau de trois hélices proche des répétitions natives et 2) des protéines portant des répétitions dites « fracturées », de nature hétérogène mais qui ne présentent jamais le motif canonique en triple hélice. L'analyse des dynamiques moléculaires réalisées sur une série d'exemples indique que les mutations qui génèrent des répétitions fracturées sont globalement plus délétères que celles qui génèrent des répétitions hybrides. Une première corrélation a pu être établie avec les phénotypes cliniques d'une cohorte de patients.

**Mots-clés** : dystrophine, myopathie de Becker, base de données, structure de protéine, dynamique moléculaire, modélisation par homologie

---

Dystrophin is involved in Duchenne (DMD) and Becker (BMD) Muscular Dystrophies. A lot of clinical and therapeutic researches are published on DMD but precise molecular role of dystrophin is largely unknown, and consequently the correlation between genotype and phenotype is difficult to establish. However, this relation is essential to offer new therapies to patients. That is why we propose to study function of BMD patient mutated dystrophin to correlate with clinical phenotypes. The database *eDystrophin* provides an overview of phenotypes associate with these mutations and consequences of each mutation on function and 3D-structures of mutated protein through homology models. The great majority of these mutations are exon deletions located in the central rod domain composed by 24 spectrin-like repeats. The use of *eDystrophin*, models and molecular dynamics highlights two types of structures at the deletion junction: hybrid repeats that reconstitute a triple coiled-coil like native repeats and fractional repeats that do not reconstitute this structure. Molecular dynamics analysis reveals that fractional repeats may be more deleterious than hybrid repeats. A first correlation between clinical phenotypes and the protein structure is established.

**Keywords**: dystrophin, Becker Muscular Dystrophy, database, protein structure, molecular dynamics, homology modeling