Muscular Dystrophy

Richard J. L. F. Lemmers, *et al.*
Science **329**, 1650 (2010);
DOI: 10.1126/science.1189044

within a habitat (*5*–*7*). Population density–dependent olfactory plasticity may serve as one of such adaptive mechanisms. Our results indicate that *nep-2* regulates olfactory plasticity by acting in a cell-nonautonomous manner. Mammals possess seven neprilysin family proteins: neprilysin, *endothelin-converting enzyme 1* (ECE1), ECE2, phosphate-regulating neutral endopeptidase (PHEX), neprilysin-2, damage-induced neuronal endopeptidase (DINE), and Kell (*16*). Although little is known about the loss-of-function phenotypes of the neprilysin family proteins, recent biochemical studies revealed that neprilysin scavenges amyloid β peptide efficiently and with this mechanism can prevent Alzheimer's disease (*17*). On the basis of our observations, we propose a model in which SNET-1 peptide is secreted as an environmental signal that prevents olfactory plasticity and NEP-2 turns off the SNET-1 signals by degrading the excess peptides, creating a balance sensitive to environmental signals. By acting on an unknown receptor in the olfactory sensory circuit, SNET-1 negatively regulates olfactory plasticity (Fig. 4G and fig. S14). Given that *daf-22*, *nep-2*, and *snet-1* regulate the plasticity of the response to several types of odorants (fig. S3), the pheromonal information may also regulate other behaviors through the peptide signaling. Our identification of the SNET-1 pathway that regulates olfactory plasticity sheds light on the complexity of the regulatory network underlying simple sensory behavior in *C. elegans* and suggests similar regulation might be present in other animals.

References and Notes

1. C. I. Bargmann, E. Hartwieg, H. R. Horvitz, *Cell* **74**, 515 (1993).
2. H. A. Colbert, C. I. Bargmann, *Neuron* **14**, 803 (1995).
3. T. Hirotsu, Y. Iino, *Genes Cells* **10**, 517 (2005).
4. W. M. Nuttley, K. P. Atkinson-Leadbeater, D. Van Der Kooy, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12449 (2002).
5. J. M. J. Travis, D. J. Murrell, C. Dytham, *Proc. Biol. Sci.* **266**, 1837 (1999).
6. H. J. Poethke, T. Hovestadt, *Proc. Biol. Sci.* **269**, 637 (2002).
7. D. E. Bowler, T. G. Benton, *Biol. Rev. Camb. Philos. Soc.* **80**, 205 (2005).
8. P. J. Hu, "Dauer," in *Worm Book*, D. L. Riddle, Ed., 10.1895/wormbook.1.144.1 (2007).
9. R. A. Butcher *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1875 (2009).
10. R. A. Butcher, M. Fujita, F. C. Schroeder, J. Clardy, *Nat. Chem. Biol.* **3**, 420 (2007).
11. P. Y. Jeong *et al.*, *Nature* **433**, 541 (2005).
12. J. Srinivasan *et al.*, *Nature* **454**, 1115 (2008).
13. R. A. Butcher, J. R. Ragains, E. Kim, J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14288 (2008).
14. K. Yamada, T. Hirotsu, M. Matsuki, H. Kunitomo, Y. Iino, *Genetics* **181**, 1347 (2009).
15. Materials and methods are available as supporting material on *Science* Online.
16. A. J. Turner, R. E. Isaac, D. Coates, *Bioessays* **23**, 261 (2001).
17. N. Iwata, M. Higuchi, T. C. Saido, *Pharmacol. Ther.* **108**, 129 (2005).
18. A. Devault *et al.*, *FEBS Lett.* **231**, 54 (1988).
19. R. Taussig, R. R. Kaldany, R. H. Scheller, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4988 (1984).
20. T. R. Zahn *et al.*, *Traffic* **5**, 544 (2004).
21. W. S. Schackwitz, T. Inoue, J. H. Thomas, *Neuron* **17**, 719 (1996).
22. W. Li, S. G. Kennedy, G. Ruvkun, *Genes Dev.* **17**, 844 (2003).
23. E. L. Peckol, E. R. Troemel, C. I. Bargmann, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11032 (2001).
24. K. Kim *et al.*, *Science* **326**, 994 (2009).

25. The authors acknowledge J. Ragains for synthesis of the dauer pheromone ascarosides. Y.I. and T.I. were supported by a Grant-in-aid for Scientific Research, J.C. by NIH grant CA24487, and R.A.B. by an NIH K99 Pathway to Independence Award (GM087533). We thank S. Mitani and P. Sengupta for *srbc-64* and *srbc-66* strains.

Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5999/1647/DC1
Materials and Methods
Figs. S1 to S14
References

10 May 2010; accepted 26 July 2010
10.1126/science.1192020

A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy

Richard J. L. F. Lemmers,[1] Patrick J. van der Vliet,[1] Rinse Klooster,[1] Sabrina Sacconi,[2] Pilar Camaño,[3,4] Johannes G. Dauwerse,[5] Lauren Snider,[6] Kirsten R. Straasheijm,[1] Gert Jan van Ommen,[1] George W. Padberg,[7] Daniel G. Miller,[8] Stephen J. Tapscott,[6] Rabi Tawil,[9] Rune R. Frants,[1] Silvère M. van der Maarel[1]*

Facioscapulohumeral muscular dystrophy (FSHD) is a common form of muscular dystrophy in adults that is foremost characterized by progressive wasting of muscles in the upper body. FSHD is associated with contraction of D4Z4 macrosatellite repeats on chromosome 4q35, but this contraction is pathogenic only in certain "permissive" chromosomal backgrounds. Here, we show that FSHD patients carry specific single-nucleotide polymorphisms in the chromosomal region distal to the last D4Z4 repeat. This FSHD-predisposing configuration creates a canonical polyadenylation signal for transcripts derived from *DUX4*, a double homeobox gene of unknown function that straddles the last repeat unit and the adjacent sequence. Transfection studies revealed that *DUX4* transcripts are efficiently polyadenylated and are more stable when expressed from permissive chromosomes. These findings suggest that FSHD arises through a toxic gain of function attributable to the stabilized distal *DUX4* transcript.

Autosomal dominant FSHD (FSHD1; OMIM 158900) is a common form of muscular dystrophy, affecting 1 in 20,000 people, that is characterized by progressive and often asymmetric weakness and wasting of facial, shoulder girdle, and upper arm muscles (*1*). The disorder is most often caused by contraction of the D4Z4 macrosatellite repeat array in the subtelomeric region of chromosome 4q35 (*2*). This polymorphic macrosatellite repeat normally consists of 11 to 100 D4Z4 units, each 3.3 kb in size and ordered head-to-tail. Patients with FSHD1 have one repeat array of 1 to 10 units (Fig. 1A). At least one unit of D4Z4 is required to develop FSHD (*3*).

D4Z4 contraction needs to occur on a specific chromosomal background to cause FSHD. The chromosome 10q subtelomere contains an almost identical repeat array, but contractions on this chromosome are nonpathogenic (Fig. 1A). Translocated copies of the chromosome 4 and chromosome 10 repeat units are frequently encountered on either chromosome end (*4*). This complex genetic situation, in which genetically almost identical repeat units can be exchanged between both chromosomes with apparently discordant pathological consequences, has long hampered the identification of the disease mechanism.

Disease models were postulated in which D4Z4 repeat contractions cause chromatin remodeling and transcriptional deregulation of genes close to D4Z4. Indeed, contracted D4Z4 repeat arrays show partial loss of DNA methylation and of heterochromatic histone 3 lysine 9 trimethylation and heterochromatin protein 1γ markers consistent with a more open chromatin structure (*5*, *6*). Transcriptional up-regulation of genes proximal to D4Z4 was reported in FSHD1 patients (*7*) but could not be confirmed (*8*, *9*).

Exchanges between repeat units of chromosomes 4 and 10 occur much less frequently than anticipated: Most translocated repeat units are relicts of ancient translocation events between chromosomes 4q and 10q (*10*). Of the two distal chromosome 4q configurations, 4qA and 4qB, only contractions of the 4qA form led to FSHD1

[1]Department of Human Genetics, Leiden University Medical Center, 2333 ZA Leiden, Netherlands. [2]Centre de Reference pour les Maladies Neuromusculaires and CNRS UMR6543, Nice University, 06202 Nice, France. [3]Department of Neurosciences, BioDonostia Health Research Institute, Hospital Donostia, 20014 San Sebastián, Spain. [4]CIBERNED, Instituto de Salud Carlos III, 28029 Madrid, Spain. [5]Department of Clinical Genetics, Leiden University Medical Center, 2333 ZA Leiden, Netherlands. [6]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [7]Department of Neurology, Radboud University Medical Center Nijmegen, 6500 HC Nijmegen, Netherlands. [8]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. [9]Department of Neurology, University of Rochester Medical Center, Rochester, NY 14642, USA.

*To whom correspondence should be addressed. E-mail: maarel@lumc.nl

(*11*). Genetic follow-up studies unveiled consistent polymorphisms in the FSHD locus, resulting in the recognition of at least 17 genetic variants of distal 4q (*10*). Contractions in the common variant 4A161 cause FSHD1, whereas contractions in many other variants, such as the common 4B163, do not cause FSHD1 (Fig. 1A) (*12*). Thus, it appears that chromosome 4A161–specific sequence variants are causally related to FSHD.

Because at least one D4Z4 unit is necessary to cause disease, we reasoned that the minimal pathogenic region might reside in the first or the last unit. The distal unit of the D4Z4 repeat was recently shown to have a transcriptional profile that differs from internal units (*13*, *14*). Although the major transcript in each unit is the *DUX4* gene, which codes for a double homeobox protein, none of these transcripts seem to be stable,

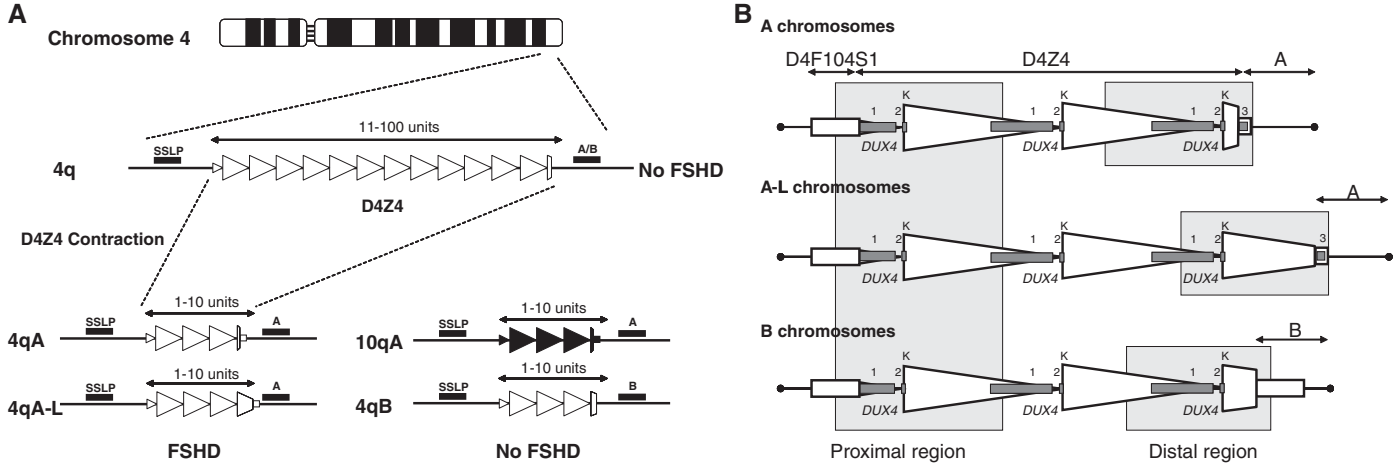

Fig. 1. Schematic overview showing the minimal genetic requirement for FSHD1. (**A**) The D4Z4 repeat array on chromosome 4q35 (open triangles) and its homolog on chromosome 10q26 (closed triangles), indicating the location of the simple sequence length polymorphism (SSLP) and 4qA/4qB polymorphisms that define the genetic background of the repeat. Patients with FSHD1 have a D4Z4 repeat array size of 1 to 10 units on 4qA but not on 4qB or 10q chromosomes. (**B**) Schematic of the D4Z4 repeat and flanking sequences on A, A-L (with an extended distal D4Z4 repeat unit), and B chromosomes. Each D4Z4 unit is defined by the *Kpn*I restriction site (K). The proximal and distal regions that were sequenced are indicated. The exons of *DUX4* are indicated as gray boxes numbered 1 to 3.

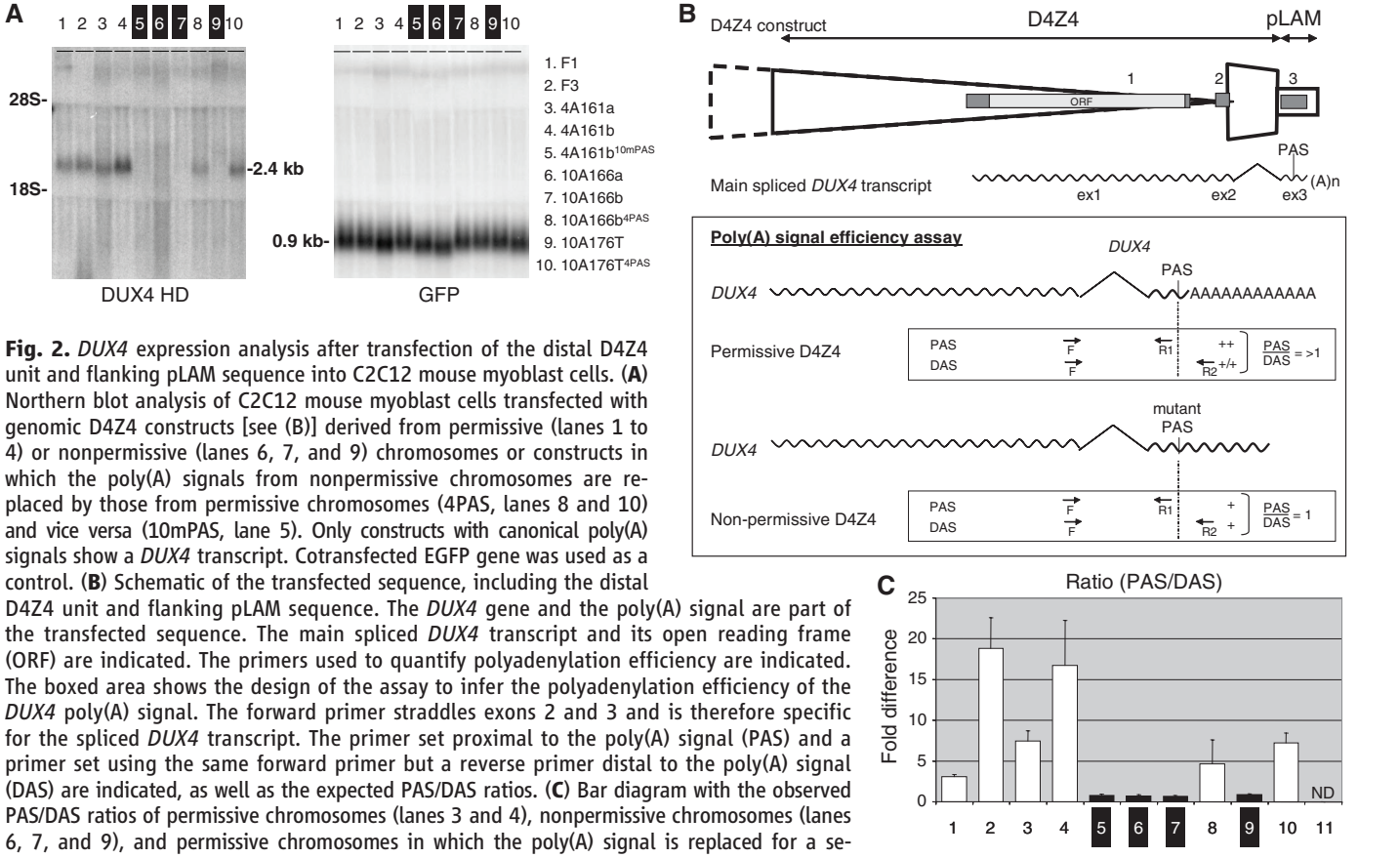

Fig. 2. *DUX4* expression analysis after transfection of the distal D4Z4 unit and flanking pLAM sequence into C2C12 mouse myoblast cells. (**A**) Northern blot analysis of C2C12 mouse myoblast cells transfected with genomic D4Z4 constructs [see (B)] derived from permissive (lanes 1 to 4) or nonpermissive (lanes 6, 7, and 9) chromosomes or constructs in which the poly(A) signals from nonpermissive chromosomes are replaced by those from permissive chromosomes (4PAS, lanes 8 and 10) and vice versa (10mPAS, lane 5). Only constructs with canonical poly(A) signals show a *DUX4* transcript. Cotransfected EGFP gene was used as a control. (**B**) Schematic of the transfected sequence, including the distal D4Z4 unit and flanking pLAM sequence. The *DUX4* gene and the poly(A) signal are part of the transfected sequence. The main spliced *DUX4* transcript and its open reading frame (ORF) are indicated. The primers used to quantify polyadenylation efficiency are indicated. The boxed area shows the design of the assay to infer the polyadenylation efficiency of the *DUX4* poly(A) signal. The forward primer straddles exons 2 and 3 and is therefore specific for the spliced *DUX4* transcript. The primer set proximal to the poly(A) signal (PAS) and a primer set using the same forward primer but a reverse primer distal to the poly(A) signal (DAS) are indicated, as well as the expected PAS/DAS ratios. (**C**) Bar diagram with the observed PAS/DAS ratios of permissive chromosomes (lanes 3 and 4), nonpermissive chromosomes (lanes 6, 7, and 9), and permissive chromosomes in which the poly(A) signal is replaced for a sequence derived from a nonpermissive chromosome (10mPAS, lane 5) or vice versa (4PAS, lanes 8 and 10), or of pathogenic chromosomes derived from families F1 and F3 (lanes 1 and 2). Lane 11 is a negative control. ND, not detected. Error bars represent values of quadruple experiments with standard errors of the mean.

probably due to the absence of a polyadenylation signal in internal D4Z4 units. Spliced and unspliced transcripts of the *DUX4* gene in the last unit, however, use a unique 3′ untranslated region (UTR) in the pLAM region (*15*), which is immediately distal to this last unit (Figs. 1B and 2B) and which contains a poly(A) signal that presumably stabilizes this distal transcript (*13*, *14*). The *DUX4* transcript of the distal D4Z4 unit encompasses two facultative introns in the 3′UTR. When expressed in C2C12 muscle cells, DUX4 causes a phenotype compatible with molecular observations in FSHD (*16*). This distal *DUX4* transcript can be observed in FSHD1 myotubes but not in control myotubes (fig. S1) (*17*).

To investigate why the 4A161 chromosome is permissive for disease, we compared the sequence of the 4A161 chromosome with that of common, nonpermissive 4B163 and 10A166 chromosomes. We could not identify a sequence signature in the proximal D4Z4 unit of the repeat array that explained the permissiveness of the 4A161 chromosome (fig. S2). However, immediately distal to D4Z4, in the adjacent pLAM sequence, we found a polymorphism potentially affecting polyadenylation of the distal *DUX4* transcript. The *DUX4* poly(A) signal ATTAAA, which is commonly used in humans (*18*), is present on the permissive 4A161 chromosome, whereas the corresponding ATCAAA sequence on chromosome 10q is not known to be a poly(A) signal (fig. S2). Nonpermissive 4qB chromosomes, like 4B163, lack pLAM altogether, including this poly(A) site (Fig. 1B). Another nonpermissive 10qA chromosome (10A176T) (*10*) carries ATTTAA at this position, which is also not known as a poly(A) signal (figs. S2 and S3). In silico poly(A) signal prediction programs (*19*, *20*) also recognized the *DUX4* poly(A) signal in 4A161 but failed to identify potential poly(A) signals in nonpermissive chromosomes 10A166 and 10A176T.

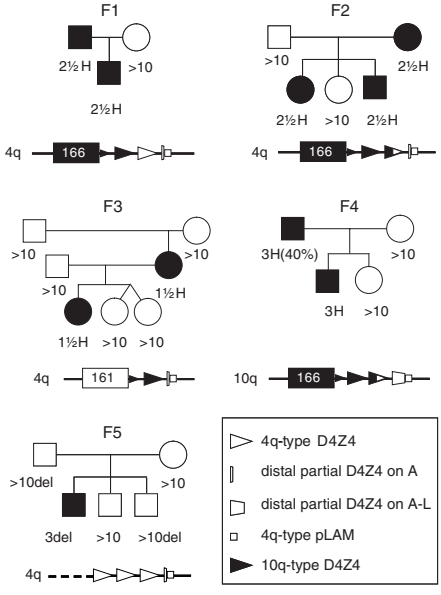

Fig. 3. Pedigrees of FSHD1 families with complex pathogenic chromosomes. Families F1 to F4 all carry a hybrid D4Z4 repeat (H) that commences with chromosome 10-type repeat units (closed triangles) but end with 4-type repeat units (open triangles). In family F3, a meiotic rearrangement between chromosomes 4 and 10 generated a short hybrid repeat structure on 4A161. In family F4, this pathogenic repeat is located on chromosome 10 and originates from a mitotic D4Z4 contraction in the mildly affected father that is transmitted to his affected son. Family F5 represents a disease chromosome in which, in addition to partial deletion of the D4Z4 repeat, the region proximal to the D4Z4 repeat is also deleted (del).

To explore whether these polymorphisms affect the distal *DUX4* transcript, we transfected the last D4Z4 unit and flanking pLAM sequence of permissive and nonpermissive chromosomes in C2C12 cells and assessed the stability of the distal *DUX4* transcript by Northern blot analysis (Fig. 2A). We also examined the relative potency of the poly(A) signals on the permissive and nonpermissive chromosomes in directing polyadenylation of the distal *DUX4* transcript. We studied polyadenylation site usage indirectly by using a quantitative reverse transcription polymerase chain reaction (*21*) in which we compared *DUX4* transcript levels proximal and distal of the poly(A) site (Fig. 2B). The use of the predicted poly(A) signal was verified by 3′ rapid amplification of cDNA ends (fig. S4). We also transfected constructs in which the poly(A) signal of permissive chromosomes was replaced by those of nonpermissive chromosomes, and vice versa. We found that *DUX4* transcripts were stable (Fig. 2A) and efficiently polyadenylated (Fig. 2C) when we used constructs from permissive chromosomes or when the poly(A) signal of a permissive chromosome was introduced on constructs derived from nonpermissive chromosomes. Consistently, when constructs derived from nonpermissive chromosomes were transfected, no *DUX4* transcripts could be detected on Northern blot and polyadenylation was inefficient. *DUX4* stability and polyadenylation efficiency decreased when the poly(A) signal of permissive constructs was replaced by nonpermissive sequences. Altogether, constructs with a bona fide poly(A) signal produced stable transcripts and showed polyadenylation efficiency that was 4 to 16 times as high as constructs with a mutation in the poly(A) signal. This suggests that increased polyadenylation, and hence stability, of the distal *DUX4* transcript may be centrally involved in FSHD pathogenesis.

We next studied FSHD1 patients with unusual hybrid D4Z4 repeat array structures that contain mixtures of 4-type and 10-type units. We identified four families (F1 to F4) with one or more individuals with FSHD1, carrying a contracted D4Z4 repeat array that commences with 10-type units and ends with 4-type units (Fig. 3). In family F3, we identified a patient with a de novo meiotic rearrangement between chromosomes 4q and 10q, leaving one and a half 10-type repeat units on a permissive 4A161 chromosome. In family F4, the mildly affected father is a mosaic FSHD1 patient (*22*) due to a mitotic contraction of such hybrid repeat array. The mosaic pathogenic repeat starts with two and a half 10-type D4Z4 units and ends with one and a half 4-type repeat units. This repeat array in the father was transmitted to his affected son, demonstrating its pathogenicity, and, unexpectedly, it was found to reside on chromosome 10 (fig. S6). Only the distal end of the D4Z4 repeat array was transferred to chromosome 10q, so that none of the FSHD candidate genes located proximal to the D4Z4 repeat array were cotransferred to chromosome 10 (fig. S6). This report of a FSHD1 family linked to chromosome 10 apparently precludes a key role for proximal 4q genes in the pathogenesis of FSHD. Altogether, all unusual FSHD1-causing repeat arrays reported here thus share the commonality of a terminal 4qA repeat unit with a directly adjacent pLAM sequence.

We also analyzed other disease-permissive chromosome 4 variants (fig. S7): 4A161L was previously described (*10*, *15*), whereas 4A159 and 4A168 are newly discovered uncommon permissive variants from a survey of >300 independent patients with FSHD. In addition, we studied >2000 control individuals and identified additional nonpermissive chromosome variants: 4B168, 10A164, and 10B161T (fig. S3). Thus, D4Z4 contractions on 4A161, 4A161L, 4A159, and 4A168 chromosomes are pathogenic, and D4Z4 contractions on 4B163, 4B168, 10A166, 10A164, 10B161T, and 10A176T chromosomes are nonpathogenic.

We sequenced the first and last D4Z4 units and flanking sequences in these newly identified permissive and nonpermissive chromosomes. (Fig. 1B and fig. S2). In support of our earlier data, there is no common sequence in the proximal D4Z4 region that unifies FSHD-permissive chromosomes. At the distal end, all permissive chromosomes differed very little in sequence and all contained a canonical *DUX4* poly(A) signal, whereas nonpermissive chromosomes showed much more sequence variation relative to the permissive chromosomes. The only exception, 4B163, has a D4Z4 sequence highly identical to 4A161 but, importantly, lacks the pLAM sequence (Fig. 1). The permissive 4A161L chromosome is identical to 4A161 but carries an extended D4Z4 sequence, preceding an identical pLAM sequence (Fig. 1B and fig. S2). Sequence analysis of the distal D4Z4-pLAM region of the pathogenic chromosome in our four families with complex repeat array structures showed a sequence identical to the permissive 4A161 sequence. Transfection experiments with D4Z4-pLAM sequences derived from the disease chromosomes of families F1 and F3 showed transcript stabilities and polyadenylation efficiencies of the distal *DUX4* transcript comparable to 4A161 chromosomes (Fig. 2B). This demonstrates that *DUX4* can also be efficiently produced from these chromosomes. Altogether, our study demonstrates that all patients with FSHD1 that came to our attention have an

identical sequence in the last D4Z4 unit and immediately flanking pLAM sequence, and it shows that specific sequence variants unique to the permissive haplotypes confer pathogenicity to the repeat irrespective of its chromosomal localization (fig. S8).

Finally, this distal pLAM region is also preserved in individuals with FSHD1 in whom the deleted region extends proximally to the D4Z4 repeat array (F5 in Fig. 3), as well as in FSHD2 patients, who have a classical FSHD phenotype but show a similar local chromatin relaxation on a 4A161 chromosome independent of D4Z4 repeat array contraction (*6*, *23*).

Our study puts forward a plausible genetic model for FSHD. In this model, two polymorphisms create a polyadenylation site for the distal *DUX4* transcript, located in the pLAM sequence. In combination with the chromatin relaxation of the repeat, this leads to increased *DUX4* transcript levels. FSHD may arise through a toxic gain of function attributable to the stabilized distal DUX4 transcript. Our study thus not only explains the striking chromosome specificity of the disorder but also provides a genetic mechanism that may unify the genetic observations in patients with FSHD.

References and Notes

1. R. Tawil, S. M. Van Der Maarel, *Muscle Nerve* **34**, 1 (2006).
2. C. Wijmenga *et al.*, *Nat. Genet.* **2**, 26 (1992).
3. R. Tupler *et al.*, *J. Med. Genet.* **33**, 366 (1996).
4. J. C. de Greef, R. R. Frants, S. M. van der Maarel, *Mutat. Res.* **647**, 94 (2008).
5. P. G. van Overveld *et al.*, *Nat. Genet.* **35**, 315 (2003).
6. W. Zeng *et al.*, *PLoS Genet.* **5**, e1000559 (2009).
7. D. Gabellini, M. R. Green, R. Tupler, *Cell* **110**, 339 (2002).
8. R. Klooster *et al.*, *Eur. J. Hum. Genet.* **17**, 1615 (2009).
9. P. S. Masny *et al.*, *Eur. J. Hum. Genet.* **18**, 448 (2010).
10. R. J. Lemmers *et al.*, *Am. J. Hum. Genet.* **86**, 364 (2010).
11. R. J. Lemmers *et al.*, *Nat. Genet.* **32**, 235 (2002).
12. R. J. Lemmers *et al.*, *Am. J. Hum. Genet.* **81**, 884 (2007).
13. M. Dixit *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18157 (2007).
14. L. Snider *et al.*, *Hum. Mol. Genet.* **18**, 2414 (2009).
15. J. C. van Deutekom *et al.*, *Hum. Mol. Genet.* **2**, 2037 (1993).
16. D. Bosnakovski *et al.*, *EMBO J.* **27**, 2766 (2008).
17. Materials and methods are available as supporting material on *Science* Online.
18. E. Beaudoing, S. Freier, J. R. Wyatt, J. M. Claverie, D. Gautheret, *Genome Res.* **10**, 1001 (2000).
19. F. Ahmed, M. Kumar, G. P. Raghava, *In Silico Biol.* **9**, 13 (2009).
20. H. Liu, H. Han, J. Li, L. Wong, *Bioinformatics* **21**, 671 (2005).
21. S. H. Ahn, M. Kim, S. Buratowski, *Mol. Cell* **13**, 67 (2004).
22. S. M. van der Maarel *et al.*, *Am. J. Hum. Genet.* **66**, 26 (2000).
23. J. C. de Greef *et al.*, *Hum. Mutat.* **30**, 1449 (2009).

24. We thank all patients and family members for their participation. This study was supported by the Fields Center for FSHD and Neuromuscular Research, the Netherlands Organization for Scientific Research NWO 917.56.338, a Breakthrough Project Grant by the Netherlands Genomics Initiative NWO 93.51.8001, the National Institutes of Health P01NS069539, the Muscular Dystrophy Association, the Shaw Family Foundation, a Marjorie Bronfman Fellowship grant from the FSH Society, the Dutch FSHD Foundation, the Pacific Northwest Friends of FSH Research, Centro Investigación Biomédica en Red para Enfermedades Neurodegenerativas (CIBERNED), the Basque Government (Fellowship grant, 2008111011), and Instituto Carlos III, ILUNDAIN Fundazioa.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1189044/DC1
Materials and Methods
Figs. S1 to S8
References

3 March 2010; accepted 21 July 2010
Published online 19 August 2010;
10.1126/science.1189044
Include this information when citing this paper.

The Genetic and Molecular Basis for Sunscreen Biosynthesis in Cyanobacteria

Emily P. Balskus and Christopher T. Walsh*

Ultraviolet UV-A and UV-B radiation is harmful to living systems, causing damage to biological macromolecules. An important strategy for dealing with UV exposure is the biosynthesis of small-molecule sunscreens. Among such metabolites, the mycosporine and mycosporine-like amino acids (MAAs) are remarkable for their wide phylogenetic distribution and their unique chemical structures. Here, we report the identification of a MAA biosynthetic gene cluster in a cyanobacterium and the discovery of analogous pathways in other sequenced organisms. We have expressed the cluster in a heterologous bacterial host and characterized all four biosynthetic enzymes in vitro. In addition to clarifying the origin of the MAAs, these efforts have revealed two unprecedented enzymatic strategies for imine formation.

Mycosporines and mycosporine-like amino acids (MAAs) are ultraviolet (UV)–absorbing small molecules (λ_{max} = 310 to 360 nm) (*1*) that are synthesized by cyanobacteria, fungi, and eukaryotic micro- and macroalgae (*2*–*5*). These metabolites are also ingested and accumulated by numerous marine animals. Discovered and named for mediating light-induced fungal sporulation (*6*–*8*), mycosporines and MAAs are involved in many other biological processes, including UV photoprotection of organisms (*2*–*5*, *9*) and their embryos (*10*), osmotic regulation (*11*), and defense against oxidative stress (*12*). Additionally, their ability to prevent UV-induced skin damage in vivo (*13*) has led to their commercialization as Helioguard 365, a formulation containing MAAs shinorine and porphyra-334 that is used in skin care and cosmetic products.

Structurally this large natural-product family is divided into two groups: the mycosporines, which contain a single proteogenic or modified amino acid residue linked to a cyclohexenone core, and the MAAs, which have two such substituents (Fig. 1A). The amino acids are attached to the central ring through imine linkages, an arrangement best represented as a combination of resonance tautomers (Fig. 1B). This extensive conjugation facilitates absorption of UV (Fig. 1C). Fungi produce only mycosporines, whereas cyanobacteria, algae, and marine organisms can synthesize both mycosporines and MAAs.

The occurrence, distribution, and biology of mycosporines and MAAs have been subjects of intense investigation, but elucidation of the precise biochemical mechanisms underlying their functions has been hampered by a lack of knowledge regarding the molecular and genetic basis for their biosynthesis. Previous feeding studies with cyanobacteria and fungi revealed that the amino acid substituents were added in the later stages of biosynthesis and suggested, but did not prove, that the mycosporine core might originate from shikimate pathway intermediate 3-dehydroquinate (*14*, *15*). Despite these efforts, no genes have been definitively linked to mycosporine or MAA production in any organism.

We attempted to locate candidate biosynthetic gene clusters using a genome mining approach prompted by the discovery of genes encoding shikimate pathway enzymes in the genome of sea anemone *Nematostella vectensis* (*16*), including a dehydroquinate synthase homolog (DHQS) adjacent to an *O*-methyltransferase (O-MT). Recognizing that a methyltransferase was likely involved in mycosporine assembly, we considered the DHQS and O-MT to be candidate biosynthetic genes. BLAST searches with both protein sequences revealed clustered homologs in genomes of dinoflagellates, cyanobacteria, and fungi (*17*).

Examining the hits in their genomic contexts revealed that open reading frames (ORFs) encoding DHQSs and O-MTs were accompanied by a third conserved ORF, usually annotated as a hypothetical protein. A conserved domain search using this gene revealed homology to the adenosine triphosphate (ATP)–grasp superfamily. As these enzymes catalyze peptide bond formation, we reasoned that this gene product might be involved in amino acid installation. We also noted genetic variations between cyanobacterial and fungal clusters consistent with the structural differences between cyanobacterial MAAs and fungal mycosporines; cyanobacteria had a second

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: christopher_walsh@hms.harvard.edu