



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# *In silico* analyses of dystrophin Dp40 cellular distribution, nuclear export signals and structure modeling



Alejandro Martínez-Herrera<sup>a,1</sup>, Jorge Aragón<sup>a,1</sup>, Rosa Ma. Bermúdez-Cruz<sup>a</sup>, Ma. Luisa Bazán<sup>a</sup>, Gabriela Soid-Raggi<sup>a</sup>, Víctor Ceja<sup>a</sup>,  
 Andrea Santos Coy-Arechavaleta<sup>a</sup>, Víctor Alemán<sup>b</sup>,  
 Francisco Depardón<sup>a,2</sup>, Cecilia Montañez<sup>a,\*</sup>

<sup>a</sup> Departamento de Genética y Biología Molecular, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México, D. F., Mexico

<sup>b</sup> Departamento de Fisiología, Biofísica y Neurociencias, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México, D. F., Mexico

## ARTICLE INFO

## Article history:

Received 21 May 2015

Received in revised form

16 June 2015

Accepted 16 June 2015

Available online 25 June 2015

## Keywords:

*In silico*

Dystrophin Dp40

Cellular distribution

## ABSTRACT

Dystrophin Dp40 is the shortest protein encoded by the DMD (Duchenne muscular dystrophy) gene. This protein is unique since it lacks the C-terminal end of dystrophins. In this data article, we describe the subcellular localization, nuclear export signals and the three-dimensional structure modeling of putative Dp40 proteins using bioinformatics tools. The Dp40 wild type protein was predicted as a cytoplasmic protein while the Dp40n4 was predicted to be nuclear. Changes L93P and L170P are involved in the nuclear localization of Dp40n4 protein. A close analysis of Dp40 protein scored that amino acids <sup>93</sup>LEQEHNNLV<sup>101</sup> and <sup>168</sup>LLLHDSIQI<sup>176</sup> could function as NES sequences and the scores are lost in Dp40n4. In addition, the changes L93/170P modify the tertiary structure of putative Dp40 mutants. The analysis showed that changes of residues 93 and 170 from leucine to proline allow the nuclear localization of Dp40 proteins. The data described here

DOI of original article: <http://dx.doi.org/10.1016/j.neulet.2015.05.038>

\* Corresponding author.

E-mail address: [cecim@cinvestav.mx](mailto:cecim@cinvestav.mx) (C. Montañez).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Deceased.

\*Corresponding author.

<http://dx.doi.org/10.1016/j.dib.2015.06.007>

2352-3409/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

are related to the research article entitled “EF-hand domains are involved in the differential cellular distribution of dystrophin Dp40” (J. Aragón et al. *Neurosci. Lett.* 600 (2015) 115–120) [1].

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications table

Subject area	Biology
More specific subject area	Dystrophin bioinformatics analyses
Type of data	Table, image
How data was acquired	<i>In silico</i> analyses: PSORT II program ( <a href="http://www.genscript.com/psort/psort2.html">http://www.genscript.com/psort/psort2.html</a> ) was used to predict the protein localizations; NetNES 1.1 server ( <a href="http://www.cbs.dtu.dk/services/NetNES/">http://www.cbs.dtu.dk/services/NetNES/</a> ) to search nuclear export signal; and SWISS-MODEL and I-TASSER programs ( <a href="http://zhanglab.cmb.med.umich.edu/I-TASSER/">http://zhanglab.cmb.med.umich.edu/I-TASSER/</a> ) for modeling protein structures. The comparative modeling was made in: <a href="http://www.pymol.org/funding.html">http://www.pymol.org/funding.html</a>
Data format	Analyzed
Experimental factors	N/A
Experimental features	N/A
Data source location	Departamento de Genética y Biología Molecular, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México, D. F., México.
Data accessibility	Dp40 mRNA sequence, accession number: KF154977, <a href="http://www.ncbi.nlm.nih.gov/nuccore/KF154977">http://www.ncbi.nlm.nih.gov/nuccore/KF154977</a> . Dp40 protein sequence, <a href="http://www.ncbi.nlm.nih.gov/protein/543869031">http://www.ncbi.nlm.nih.gov/protein/543869031</a> .

## Value of the data

- Bioinformatics tools permit to search putative dystrophin Dp40 protein domains and/or functions;
- PSORT II program is an alternative tool to screen for the subcellular localization of Dp40 proteins;
- NetNES 1.1 server allows to identify putative nuclear export signals of dystrophin Dp40;
- Comparative modeling analysis between Dp40 and Dp40 mutants identify differences in protein structure.

## 1. Data, experimental design, materials and methods

### 1.1. Prediction of the subcellular localization of putative Dp40 and mutant proteins

To predict the subcellular localization, the amino acids sequence of Dp40 protein (Protein ID: AGV74356.1) was analyzed using PSORT II software (<http://www.genscript.com/psort/psort2.html>) [2]. The Dp40 protein was predicted as a cytoplasmic protein while the predicted localization of Dp40n4, carrying changes L93P and L170P (L93/170P) into the EF1 and EF2 hand domains, was nuclear (Table 1). Additional changes are present in Dp40n4 (M288T and D303G); however, none of these changes modified the predictions of the subcellular localization and NES score. Interestingly, the replacement of proline to leucine residues in Dp40n4, either Dp40n4-P93L, Dp40n4-P170L or Dp40n4-P93/170L, was scored as cytoplasmic (Table 1). In addition, the replacement of leucines 93 and 170 to proline residues in Dp40 (Dp40-L93/170P), was predicted to have a nuclear localization (Table 1) which was confirmed by site-directed mutagenesis [1].

### 1.2. Identification of putative nuclear export signals in Dp40 amino acid sequence

To identify possible nuclear export signals (NES), a close analysis of Dp40 and Dp40n4 amino acid sequences was carried out using the NetNES 1.1 server (<http://www.cbs.dtu.dk/services/NetNES/>) [3].

Analyses of Dp40 amino acids from T87 to C106 and C110 to L230 scored that amino acids  $^{93}\text{LEQEHNNLV}^{101}$  and  $^{168}\text{LLLHDSIQI}^{176}$  could function as NES sequences (Fig. 1A and B, respectively) and the NES score was lost when Dp40n4 and Dp40-L93/170P amino acids ( $^{93}\text{PEQEHNNLV}^{101}$  and  $^{168}\text{LLPHDSIQI}^{176}$ ; changes L93P and L170P are underlined) were analyzed (Fig. 1C and D, respectively).

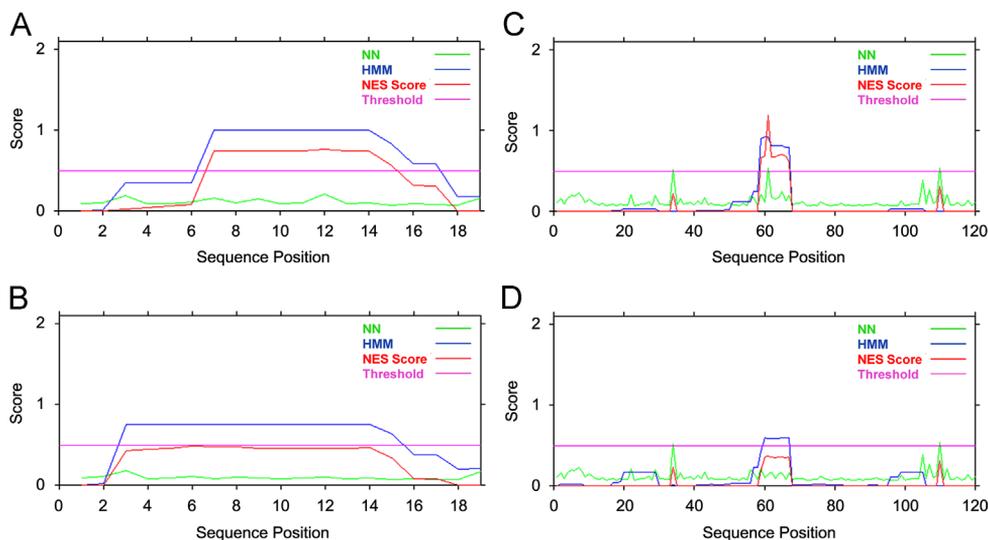
### 1.3. Comparative modeling analyses between Dp40 and mutant proteins

To identify whether mutations L93P and L170P modify the tertiary structure of Dp40 proteins, structure modeling of Dp40 and Dp40 mutants were analyzed using SWISS-MODEL and I-TASSER programs [4,5]. Fig. 2 shows the tertiary structure modeling of putative Dp40 proteins. Differences are observed between Dp40 and Dp40 mutants as well as among Dp40 mutants.

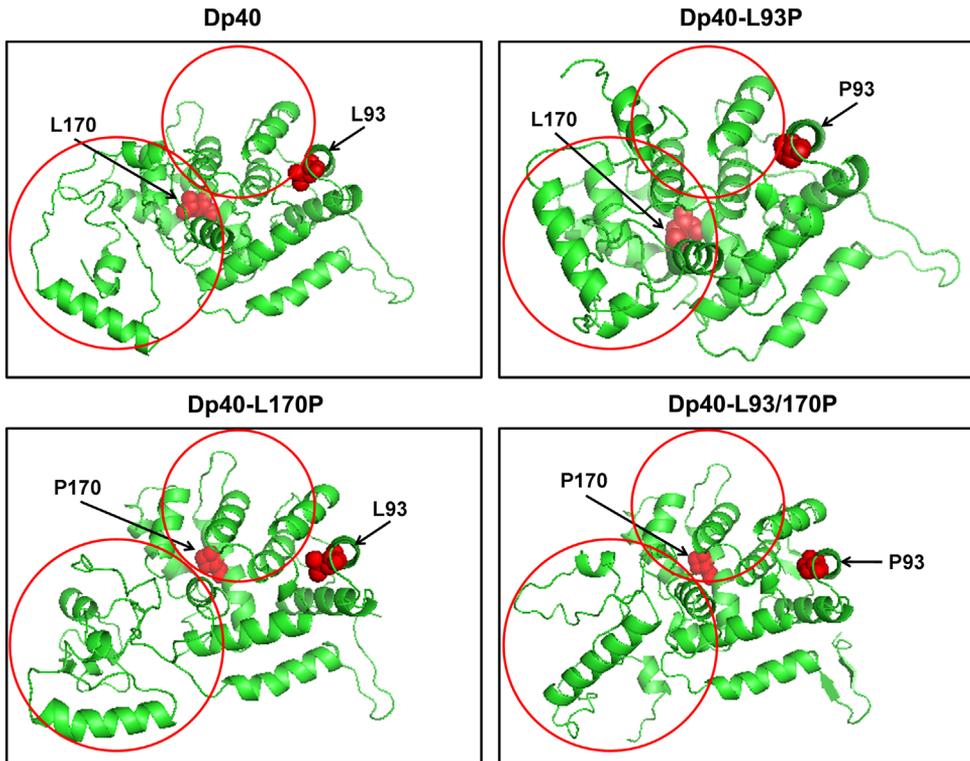
**Table 1**  
Prediction of the subcellular localization (%) of putative Dp40 proteins.

Protein	Outside the nucleus <sup>a</sup>	Nuclear
Dp40	69.6	30.4
Dp40-L93P	69.6	30.4
Dp40-L170P	69.6	30.4
Dp40-L93/170P	39.1	60.9
Dp40n4	39.1	60.9
Dp40n4-P93L	69.6	30.4
Dp40n4-P170L	69.6	30.4
Dp40n4-P93/170L	69.6	30.4

<sup>a</sup> Includes cytoplasmic, mitochondrial and peroxisomal localization.



**Fig. 1.** Prediction of putative nuclear export signals in Dp40 proteins. Analyses of Dp40 amino acids were performed using the NetNES 1.1 server [3]. (A) Analysis of Dp40 amino acids T87 to C106. (B) Analysis of Dp40 amino acids C110 to L230. (C) Analysis of Dp40n4 and Dp40-L93/170P amino acids T87 to C106. (D) Analysis of Dp40n4 and Dp40-L93/170P amino acids C110 to L230. Putative nuclear export signals (NES score), amino acids  $^{93}\text{LEQEHNNLV}^{101}$  and  $^{168}\text{LLLHDSIQI}^{176}$ , were identified in Dp40 (A and B, respectively) and values of NES score were lost in Dp40n4 and Dp40-L93/170P (C and D, respectively). NN, Neural Network; HMM, Hidden Markov Model.



**Fig. 2.** Comparative modeling of Dp40, Dp40-L93P, Dp40-L170P and Dp40-L93/170P proteins. Structure modeling was carried out using SWISS-MODEL and I-TASSER programs [4,5]. Circles indicate differences between Dp40 and Dp40 mutants. Arrows indicate leucine and/or proline residues 93 and 170 (red).

### Conflict of interest

The authors declare that there is no conflict of interest.

### Acknowledgments

This work was supported by CONACyT Grants, CB-2009-127600 and SEP-CONACyT-ECOS-ANUIES M11-S02 B000/064/12, to C. Montañez, a doctoral scholarship, 173195, to A. Martínez-Herrera, and a post-doctoral grant, 76927, to G. Soid-Raggi.

### References

- [1] J. Aragón, A. Martínez-Herrera, R.M. Bermúdez-Cruz, M.L. Bazán, G. Soid-Raggi, V. Ceja, A. Santos Coy-Arechavaleta, V. Alemán, F. Depardón, C. Montañez, EF-hand domains are involved in the differential cellular distribution of dystrophin Dp40, *Neurosci. Lett.* 600 (2015) 115–120. In press.
- [2] K. Nakai, P. Horton, PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization, *Trends Biochem. Sci.* 24 (1999) 34–35.
- [3] T. la Cour, L. Kiemer, A. Molgaard, R. Gupta, K. Skriver, S. Brunak, Analysis and prediction of leucine-rich nuclear export signals, *Protein Eng. Des. Sel.* 17 (2004) 527–536.
- [4] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis* 18 (1997) 2714–2723.
- [5] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat. Protoc.* 5 (2010) 725–738.