



Insilco Characterization and Predication of Mutation Host Pots I ACTA1 and Designing the Mutable Vector for Carrying the Peptide

*Shany George

U.C college-Aluva

Phone: 9746198277, Email: *shanygeorge87@gmail.com

Received: July 20th 2013, Accepted: Aug 13th 2013, Published: Sep 1st 2013

Abstract:

Nemaline myopathy is a disease that primarily affects skeletal muscles due to mutations in the Actin Protein family. The purpose of analysis was to determine the potential regions on the ACTA1 protein that have maximum probability of mutations. The identification of the disordered regions was performed using various tools like **DISEMBL**, **GLOB PLOT** and **RONN**. From the results of the disorder prediction the regions are identified which are further processed for the identification of sites within the regions that have maximum chances of mutability and thus termed as mutational hotspots. To identify the mutational hotspots RONN results were considered followed by the use of the tools like Polyphen and I Mutant. The effect of point substitution mutations on the stability of the protein was calculated using **Imutant**. **Polyphen** would detect the effect of the mutations on the function of the proteins. The complete mutational site annotation and structure prediction was carried out. The work was extended to identify the restriction sites present in the gene sequence of the study protein. Once the restriction sites were identified using **Nebcutter** the results were employed for the construction of Plasmid vector that can carry the mutant or wild peptide so as to express the activity of the peptide. This data can be of immense use in the application of rDNA technology in the analysis of this protein.

Keywords: *Nemaline myopathy, Mutation Analysis, ACT1 Protein, ACTA1 protein, r DNA Technology.*

Introduction:

Myopathy is the neuromuscular disorders in which the primary symptom is the muscle weakness ^[1] caused due to dysfunction of muscle fiber. Myopathy has been classified in to many disorders; they are

Inflammatory Myopathies, Muscular dystrophies, Myotonic Syndromes, Congenital myopathies, metabolic myopathies, Endocrine myopathies, Endocrine myopathies etc. Myopathic disorders generally produce proximal arm and leg weakness a so called limb girdle pattern ^[2]. Congenital myopathies are muscle diseases present at birth (congenital) that result from genetic defects in our muscle development. Congenital myopathies may result in muscle weakness, breathing and eating problems. Congenital Myopathies include Central core disease, Centronuclear myopathies, congenital fiber type disproportion myopathy, Nemaline myopathy, Multiminicore disease, Myotubular myopathy. Congenital myopathies has led to the recognition that diverse disorders are often associated with the same causative gene, and is challenging traditional classifications ^[3]. Missense mutations of the skeletal muscle alpha-actin gene (ACTA1) often results in the onset of Ccngenital myopathies ^[4]. In view of the importance of this protein in the onset of the disease the current work aims to study the mutational hot spot regions within the ACTA1 protein followed by the gene analysis and construction of suitable vector to carry the gene using Insilco methodology.

Materials and Methods:

Sequence Retrieval of ACTA1 from NCBI:

To study the protein ACTA1 responsible for the functioning of the muscle cells it is essential to have the protein sequence that would code for the functioning of the protein. The NCBI Primary database has been used for the retrieval of the protein coding amino acid sequence of ACTA1 of Human beings.

Conservation studies using BOX SHADE:

To study the conservation pattern of the protein with respect to the other organisms BOX SHADE ^[5] tool has been employed that would align the sequences using Global alignment method. This tool specifically

performs multiple sequence alignment in order to identify and analyze the conservation pattern and evolutionary strong sites present within the sequences.

Domain identification using SMART:

SMART^[6] is an online tool which performs the internal blast of the input sequence with the sequences in the CDD database to identify the total number of domain present in the given sequence along with the information regarding the functions of each domain. The tool also provides information about the other elements like low sequence repeats, signal peptides etc.

Disorder Analysis using Insilco tools:

Disorder analysis was performed using several tools like **DISEMBL**^[7], **GLOBPLOT**^[8] and **RONN**^[9]. The first two tools will enable the user to select or identify the regions of high mutability or disorderness based upon the input sequence. RONN will not only specify the regions of mutability but it also provides the disorder probability for each residue based upon which the main target site having maximum mutability can be selected.

Analysis effect of Mutations on the stability and Functionality of the proteins:

The tools like **I Mutant**^[10] and **Poly Phen**^[11] were employed to analyze the effect of the substitutions of wild type amino acid with the new amino acid. One tool would study the functional effect whereas the other is useful to calculate the effect of substitution mutations on the energy and stability of the protein sequence. Based upon the results of the disorder prediction tools mentioned above the regions of mutability were identified followed by the selection of the sites responsible for major mutations. Once the sites are identified the effect of substitution of these sites with any of the other 19 amino acids was studied by the I mutant and Poly Phen tools.

Identification of Binding sites in the Structure:

To identify the pockets/cavities present in the user provided structure CAST P^[12] has been used. This tool would provide the information about the total number of pockets present in the given structure and also the details like the total number of amino acids involved in the formation of the pocket, pocket area and volume etc. This tool also provides a direct

access to the JMOL visualization of the protein structure.

Construction of the peptides using ARGUS LAB:

The selected peptides are constructed in the ARGUS LAB software. Here the peptides are constructed based on the known sequence of the amino acids. The constructed peptides are further made stable by the process of optimization. Once the optimization is performed the energy evaluation is performed.

Reverse translation of the protein to DNA sequence:

The protein sequence obtained from the NCBI is reverse translated to DNA sequence using the online tool Reverse Translate. This tool accepts a protein sequence as input and uses a codon usage table to generate a DNA sequence representing the most likely non-degenerate coding sequence. A consensus sequence derived from all the possible codons for each amino acid is also returned.

Identification of Restriction sites using NEBCUTTER

The tool is used to identify the restriction sites and the recognition regions within the given sequence for a specified restriction enzyme. By default, only enzymes available from NEB^[13] are used, but other sets may be chosen. Just enter sequence and "submit". Further options will appear with the output. The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.

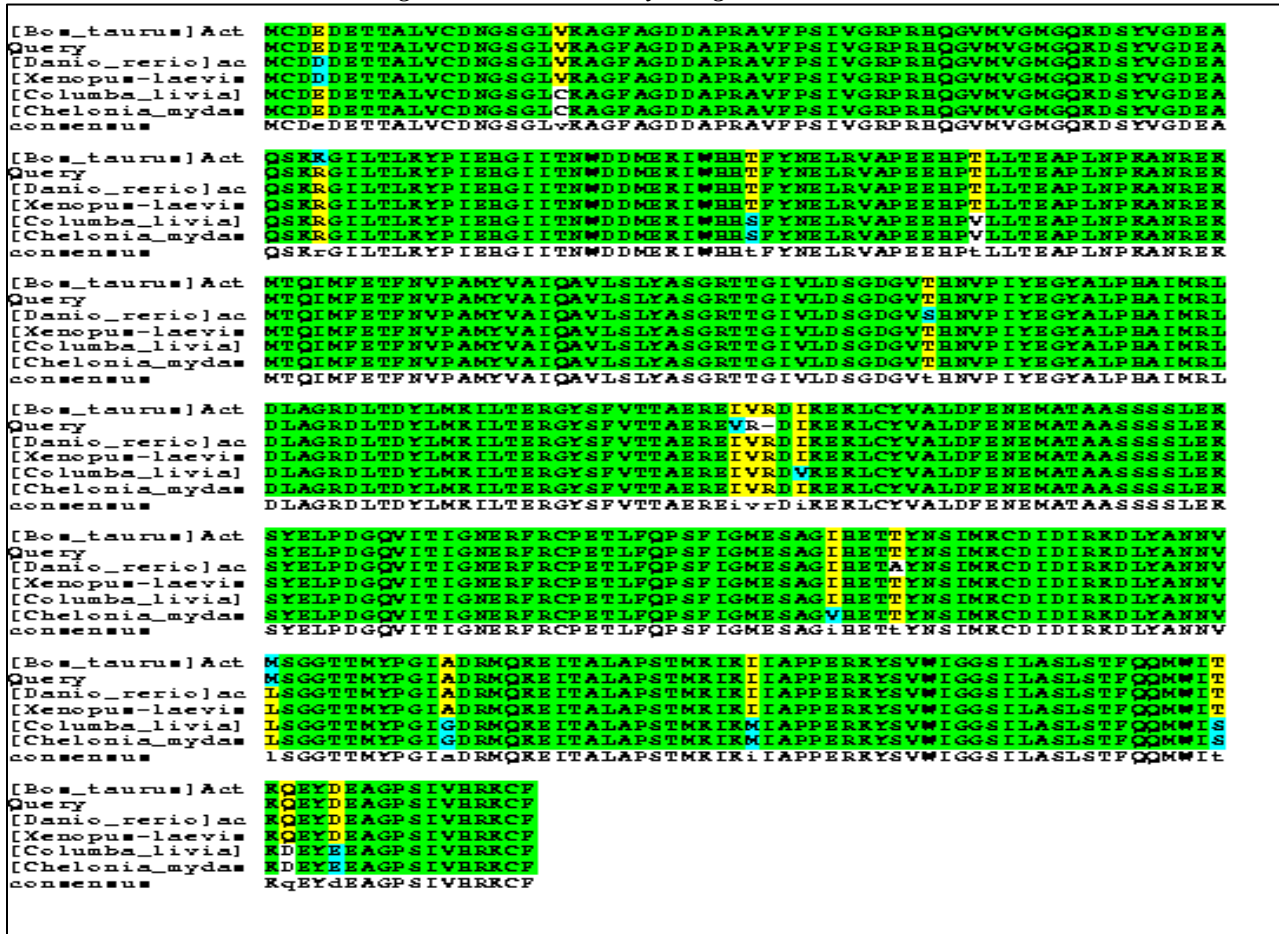
Construction of a suitable Vector to carry the gene of interest:

Once the restriction sites are identified the information can be used to construct the suitable carrier vector to carry the gene of interest. PlasMapper^[14] can be used for this purpose, which is an automatic vector designer than can provide the carrier vector for the given sequence.

Results and Discussion:

The protein sequence of ACTA1 of Homo sapiens has been retrieved from the NCBI data base using the selected Boolean operator search. The length of the sequence was found to be 376 amino acids.

Fig 1: Conservation study using BOXSHADE

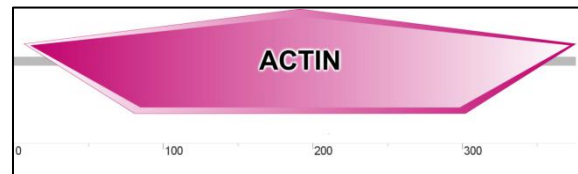


From the Fig 1 of BOX SHADE analysis it can be concluded that all the sequence of the ACTA1 protein among the selected organisms has been conserved and they show maximum identity

Functional analysis using SMART:

Based on the results of smart it can be observed that there is only one functional domain named ACTIN extending from 7 to 376 positions in the amino acid sequence. The major function of this ACTIN domain is the formation of filaments which are the major components of the cytoskeleton.

Fig 2: Picture showing the domain region of the ACTA1 protein as displayed by SMART tool



Disorder analysis using DISEMBL, GLOBPLOT and RONN:

From the results of the above mentioned tools the Region 1- 72 was found to be common among the regions shown by all the tools and has been selected for further analysis.

Table 1: Summarized Results of Disorder prediction

TOOLS	POSITION
RONN	27-31, 37-59 107-117, 226-245
DISEMBL	1-84, 95-118 149-174, 229-308 321-340, 361-376
GLOBPLOT	54-72, 190-203 337-346

The maximum probability of disorder was found to be in the region MGQKDSY from 49- 55 positions.

Table 2: Hydropathicity and Disorder probability for the selected region 49-55

Amino acid	Position	probability	Hydropathicity
M (MET)	49	0.60	-0.10
G (GLY)	50	0.61	-0.4
Q (GLN)	51	0.62	-3.5
K (LYS)	52	0.61	-3.9
D (ASP)	53	0.61	-3.5
S (SER)	54	0.61	-0.8
Y (TYR)	55	0.60	-1.3

Effect of Substitution mutations at 54 positions on Function and Stability of the protein:

Using the online tools like I Mutant and PolyPhen the effect of the substitutions on the Stability and Functionality has been calculated.

Table 3: Effect of substitutions on stability and function

Position	Wild type	New	Stability (I Mutant)	Effect (Poly Phen)
54	S	G	Decrease	Tolerated
54	S	M	Increase	Deleterious
54	S	A	Decrease	Tolerated
54	S	I	Increase	Tolerated
54	S	Q	Decrease	Deleterious
54	S	K	Increase	Tolerated
54	S	D	Increase	Tolerated
54	S	Y	Increase	Deleterious
54	S	V	Increase	Tolerated
54	S	H	Decrease	Deleterious
54	S	P	Increase	Tolerated
54	S	E	Increase	Tolerated
54	S	L	Increase	Tolerated
54	S	R	Increase	Tolerated
54	S	F	Increase	Deleterious
54	S	W	Increase	Deleterious
54	S	C	Decrease	Tolerated
54	S	N	Increase	Tolerated
54	S	T	Increase	Tolerated

Fig 3: Showing the Mutable site (hot spot site) on the 3d structure of the ACTA1 protein

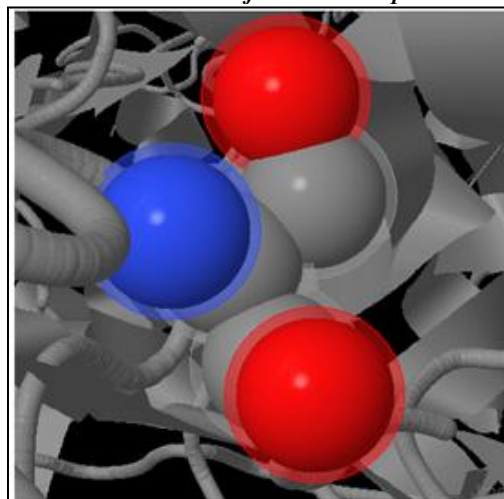
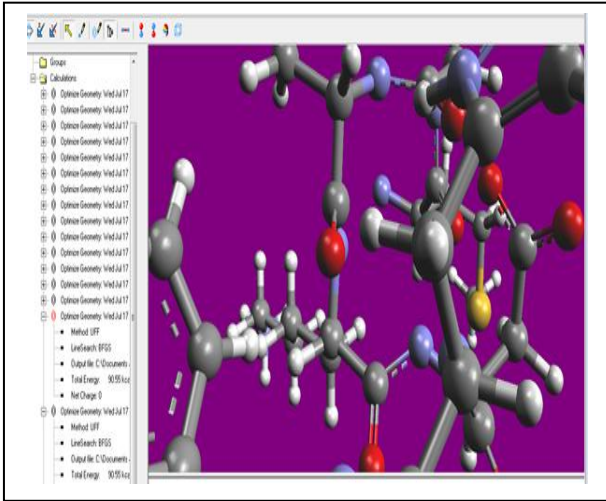


Fig 4: Construction of Mutant and Wild peptides using Argus Lab:

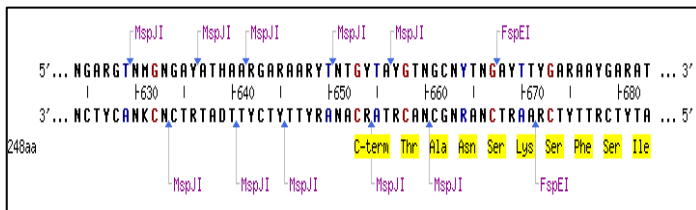


All the alternative peptides with the position 54 replaced by the other amino acids were constructed in Argus lab and the energies were calculated. The energy obtained for the peptide with wild type amino acid is 53.67 where as all the other amino acids in that position gave a high increase in the energy value. This indicated that the mutant peptides have low stability than the wild type amino acid

Protein to DNA Reverse Translate:

The online tool name Reverse Translate has been used to convert the given Protein sequence into Nucleic acid sequence so as to identify the restriction sites present in the given sequence.

Fig 5: NEB CUTTER to identify the restriction site and Recognition regions:

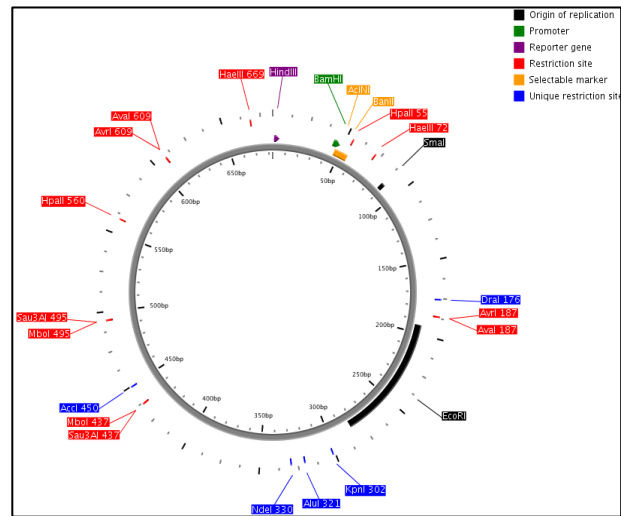


The above figure shows the total restriction sites along with the restriction enzyme for cutting the ACTA1 gene. This information can further be used to design the suitable carrier for carrying this gene in case of genetic transformations.

PLASMAPPER to construct the carrier vector to carry the ACTA1 gene:

PLASMAPPER an online automated tool has been used to construct the carrier for the selected gene. The results can be further employed in undertaking the gene transformation experiments with the query gene.

Fig 6: Showing the carrier vector for the ACTA1 gene



The above figure 6 shows the total restriction site, enzymes and other related information along with the Vector designed to carry the gene ACTA1 in case of transformational experiments.

Conclusion:

The current study aimed to analyze the complete protein of ACTA1 mutations in which results in the nemaline myopathy. The work involves the analysis of the normal sequence structure and functional properties of the protein from Homo sapiens. Once the properties of the native proteins were studied the work focused on the identification of the probable mutational regions present in the sequence based on the tools like GLOB PLOT, DISEMBL. Once the region was identified the study was extended to select the specific sites base on the tools like RONN and parameters like hydrophaticity and mutation probability. The site 54 has been identified as the mutational hot spot for the ACTA1 protein from Homo sapiens. The effect of substitution of wild amino acid by any of the other 19 possible amino

acids at the position 54 was analyzed using various tools. I mutant was used to analyze the effect on the stability of the protein where as Polyphen was used to analyze the effect on the functionality of the protein. Finally the protein was reverse translated to obtain the corresponding gene sequence which was used to assess the restriction sites in the gene. The results of nebcutter for restriction site identification were used to design the vector for carrying the ACTA1 gene. This was done using the tool PLASMAPPER. The final outcome of the work is the identification of the mutational hot spot and study the effect of mutation in this site. This was followed by designing a suitable vector that can act as a carrier for the query gene sequence.

Acknowledgement:

I express my humble gratitude to Ms. Jyothsna Gundlapally Senior Research Associate, BioAxis DNA Research center who guided me in the complete research of the paper. I would like to extend my gratitude to my teachers.

References:

- 1) NINDS Myopathy Information Page: National Institute of Neurological disorders and stroke <http://www.ninds.nih.gov/disorders/myopathy/myopathy.htm>.
- 2) Disral Myopathies, Chapter 12: Richard J. Barohn & Antony A. Amato <http://www.med.nyu.edu/neuromuscular/assets/ch12a.pdf>
- 3) Current Opinion in Neurology: October 2008 - Volume 21 - Issue 5 - p 569-575 http://journals.lww.com/co-neurology/Abstract/2008/10000/Congenital_myopathies.10.aspx
- 4) Missense mutations of ACTA1 cause dominant congenital myopathy with cores: A M Kaindl *et al*, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1735626/pdf/v041p00842.pdf>
- 5) BOXSHADE tool from MSA, from SDSC BIOLOGY WORK BENCH
- 6) SMART online tool for domain analysis: Schultz *et al*. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5857-

5864, Letunic *et al*. (2012) *Nucleic Acids Res*, doi:10.1093/nar/gkr931

7) R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson and R.B. Russell "Protein disorder prediction: implications for structural proteomics *Structure* Vol 11, Issue 11, 4 November 2003

8) GlobPlot: exploring protein sequences for globularity and disorder, *Nucleic Acid Res* 2003 - Vol. 31, No.13 (OpenAccess)

9) Yang, Z.R., Thomson, R., McMeil, P. and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins *Bioinformatics* 21: 3369-3376

10) *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W306-10., I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure

11) Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. *Nat Methods* 7(4):248-249 (2010). PubMed PDF Supplemental Information

12) Joe Dundas, Zheng Ouyang, Jeffery Tseng, Andrew Binkowski, Yaron Turpaz, and Jie Liang. 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl. Acids Res.*, 34:W116-W118.

13) Vincze, T., Posfai, J. and Roberts, R.J. NEBcutter: a program to cleave DNA with restriction enzymes, *Nucleic Acids Res.* 31: 3688-3691 (2003)

14) Xiaoli Dong, Paul Stothard, Ian J. Forsythe, and David S. Wishart "PlasMapper: a web server for drawing and auto-annotating plasmid maps" *Nucleic Acids Res.* 2004 Jul 1;32(Web Server issue):W660-4.
