

Lysine richness in human snurps - possible sites for electrophilic attacks

Sanjay Kumar Dey*, Sayak Ganguli, Protip Basu, Paushali Roy, Abhijit Datta

DBT-Centre for Bioinformatics, Presidency College, Kolkata; Sanjay Kumar Dey E-mail: sanjaydey23@gmail.com

*Corresponding author.

Received September 10, 2009; accepted November 14, 2009; published March 31, 2010

Abstract:

Gene-expression strategies are remodeled following exposure to stress. The reactive oxidants and electrophiles generated after stress actually affects the structural and functional properties of different cellular proteins. It is also seen that lysine rich motifs of proteins play crucial role in electrophilic attack and modification. Therefore, this study revealing lysine richness in 5 main human snurps (Small Nuclear Ribonucleoproteins) indicates a possible mechanism of gene regulation under stress. This possibility is highly supported by the findings that surface residues of the molecules were full of lysine rich clusters. Lysine richness is also found to be a highly conserved pattern across the various domains of life indicative of stress adaptation in the prebiotic to biotic world transition. Moreover the modeled structures showed good all atom contacts and minimal outliers.

Keywords: Electrophiles, lysine, gene regulation, oxidative stress, snurps.

Background:

After the discovery of ribozymes [1, 2] extensive discussion of the role of RNA in the origin of life initiated and led to the coining of the phrase “the RNA world”[3]. At the dawn of the hypothetical RNA world, informational molecules and biocatalysts seem unlikely to have consisted of pure RNA, without proteinaceous helpers, as amino acids are formed much more easily than nucleotides in prebiotic simulation experiments, and are easily condensed into small peptides. Thus even long before the advent of information-directed protein synthesis, RNA would have functioned in concert with a host of random peptides present in its environment. The RNA world could have been a primitive RNP (ribonucleoprotein) world. RNA would have been evolved in the context of this background of peptides, some subset of which have eventually stabilized the RNA and extended its catalytic versatility and potency [4]. Ribonucleoproteins are defined as tight complexes of one or more proteins with a short RNA molecule, usually 60-300 nucleotides long, inhabit every compartment of the human cells [5]. The present study has been concentrated to human snRNPs, all of those whose function have been assigned to play role in gene expression, underscoring the pivotal participation of RNP molecules in the evolution of gene expression apparatus. Abiotic stress responses often result in the generation of oxidative stress condition in cells. Under such antagonistic conditions the cell experiences reduction in levels of gene-expression and production of reactive oxygen species [ROS] [6]. There have been several reports regarding these alterations in gene-expression patterns as the correct regulatory point for altering expression is still not known. However, ROS and Reactive Oxygen Intermediates (ROI) molecules possess unique electrophilic and nucleophilic properties which make them suitable scavengers for lysine rich proteins [7]. It has also been reported that some stress responses in economically weak populations consuming cereal-based diets can be improved with lysine fortification [8]. This work mainly focuses on the identification of such lysine rich stretches in protein components of snRNPs. Results indicate that, lysine richness is a conserved property amongst most snRNPs and thus we conclude that electrophilic and nucleophilic attack on this lysine rich stretches probably regulate gene-expression at the vital step of RNA processing thus leading to controlled production of mature mRNAs capable of translation.

Methodology:

Dataset:

The materials are existing database sequences available at the public databases such as SWISS-PROT and GenBank of NCBI. The accession numbers of the sequences used are provided in **Table 1 (see supplementary material)**. Sequences were derived in FASTA format and were aligned using an in house tool which provided the conservation and richness of the constituent amino acids. Here all the sequences showed richness in lysine residues followed by arginine. Accessible surface area of the individual molecules was derived from the ASAVIEW Database and it was found that the

lysine residues were clustered at the surface mostly. All the proteins were then modeled using modeler 9.2 and their base pair geometry was analyzed by constructing a Ramachandran Plot using MolProbity analyzer. The modeled structures also exhibited the presence of lysine at the surface. Numerous reference models were generated and each model was screened based on the presence of the number of residues in allowed regions. Those with statistically significant number of residues in the allowed regions were selected for further analysis. The Baum-Welch algorithm was then used for the analysis of lysine clusters and only clusters which showed a minimum Kullback - Leibler distance was screened for the analysis.

Results:

Sequences alignments using the in-house tool showed richness in lysine residues followed by arginine (5.04 to 11.57 % of lysine content). The least is shown by >sp|P08621 while >sp|O43395 shows maximum lysine content. Accessible surface area of the individual molecules revealed the lysine residues were clustered at the surface mostly. Modelled structures showed good all atom contacts and minimal outliers. The modeled structures also exhibited the presence of lysine at the surface. (**Table 2 in supplementary material**)

Discussion:

The Baum-Welch algorithm was used to compute maximum likelihood estimates and posterior mode estimates for the parameters (transition and emission probabilities) of HMMs representing the various state changes during the generation of models. The training data set comprised of 68 randomly generated models and each model was then screened for their relative entropies in a simulated environment. The minimum Kullback Leibler values were included with the parameters for scoring HMMs and the final set of five models was obtained. Thus these models were then subjected to further structural studies and association with stress parameters was also analyzed. In response to physical, chemical, abiotic and oxidative stresses that causes topological and physical changes of proteins and, thus, the disruption of normal cellular processes, cells alter gene-expression strategies to aid cellular recovery following exposure to stress. A number of reports say that pre-mRNA splicing emerges as a prime basis to integrate different stresses into gene-expression profiles [9]. Moreover it is also known that during stressed conditions a large number of endogenous reactive oxidants and electrophiles are generated in the cell. These chemical induce post-translational modification of certain critical proteins causing a change in structure /function representing the cellular response to chemical exposure [7]. Proteins rich in lysine content and also consisting of numerous stretches of lysine run-ons are presumed to be act as electrophile-binding motifs [7]. Therefore our finding that human snRNPs rich in lysine, sheds light on the possibility that they may also be involved in regulation of gene-expression following cellular exposure to stresses by possibly acting as electrophilic attack sites

which disrupts their 3D structure and impair their activities thus regulating gene-expression by toning down production of mature mRNAs through splicing. The electrophiles, generated during stress, are capable of damaging cellular constituents resulting in enhanced mutation rates, altered cell signaling etc. Protein carbonyls are generated by the oxidation of lysine amino acid side chain and by the formation of Michael adducts between nucleophilic residue (Lysine) and α,β -unsaturated aldehydes. Two molecules of acrolein react with the ϵ -amino group of lysine to form predominantly acyclic 3-formyl-3,4-dehydropiperidino adduct [6]. Malondialdehyde (MDA) mainly reacts with lysine residues by Michael addition, as well as Schiff base adducts are also formed only following reaction with the amino group on lysine residues (the nitrogen atom here contains a lone-pair electron, so it can act as a nucleophile). It may disrupt the coiled body formation by snRNPs and thereby altering the metabolism of nascent transcripts [10]. Hence, the production of mature, translatable mRNAs is most sensitive to stress due to the inhibition of messenger RNA splicing and alterations in the mRNA export from the nucleus. mRNAs accumulate in discrete cytoplasmic foci like processing bodies and stress granules following exposure to stress conditions [11]. These dynamic changes in RNA metabolism ensure the preferential production and export of heat-shock mRNAs, and the sequestration of general cellular mRNAs in the nucleus or in the cytoplasmic foci. It allows redirection of the translation machinery to encode stress-proteins which aid in cellular recovery following stress (established in Yeast sp.). The presence of lysine rich residues in regulatory proteins have been reported in lower eukaryotes as well as in plants [12, 13]. Most seed proteins of plants display the property of lysine richness and serve as effective supplement for human nutrition. Absence of lysine rich diet has been identified as a focal cause in patients suffering from anxiety and stress disorders [8]. These evidences further point out that lysine richness was an integral property of regulatory process associated protein (RPAP) which dawned during the prebiotic RNA world. The presence of such lysine run -ons in human snRNPs contribute largely to their establishment as RNA - RNP world intermediates and probably provides a putative insight to primitive gene-regulation. The

occurrence of most dynamic changes in RNA metabolism under conditions of stress was known but the molecular mechanism has yet not been elucidated. This study on human snRNPs revealing lysine richness possibly suggests a mechanism by which gene expression can be regulated at the post - transcriptional level. This finding is also consistent with the conservedness and primitiveness of snRNPs from the time of origin of life in the world when the first lives faced drastic environmental stresses.

Acknowledgement:

The authors acknowledge the financial assistance of DBT, Government of India.

References:

- [1] K Kruger et al., 1982. *Cell* 31: 147. [PMID: 6297745]
- [2] C Guerrier-Takada et al., 1983. *Cell* 35: 849. [PMID: 6197186]
- [3] W Gilbert, 1986. *Nature* 319: 618.
- [4] TR Cech et al., 2006. *The RNA World*, 3rd edn: 309, CSHLP.
- [5] KT Tycowski et al., 2006. *The RNA World*, 3rd edn: 327, CSHLP.
- [6] LJ Marnett et al., 2003. *J. of Clin. Investig.* 111: 583. [PMID: 12618510]
- [7] MT Labenski et al., 2009. *Drug Metabolism and Disposition*. vol. 37 no. 6: 1211-1218. [PMID: 19251821]
- [8] M Smriga et al., 2004. *PNAS*, Vol. 101, No. 22, pp. 8285-8288. [PMCID: PMC420386]
- [9] G Biamonti, JF Caceres. 2009. *Trends in Biochemical Science*, 34: 146. [PMID: 19208481]
- [10] M Carmo-Fonseca et al., 1992. *The Journal of Cell Biology*, 117: 1. [PMID: 1532583]
- [11] U Bond. 2006. *FEMS yeast research*: vol. 6, no2, pp. 160. [PMID: 16487339]
- [12] W Sun et al., 2005. *Plant and Cell Physiology*, 46: 975. [PMID: 15840645]
- [13] B Lee et al., 2006. *The Plant Cell*, 18: 1736. [PMID: 16751345]

Edited by P. Kanguane

Citation: Dey et al., Bioinformation 4(9): 409-411 (2010)

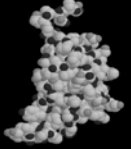
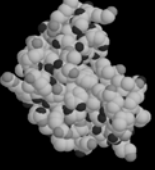

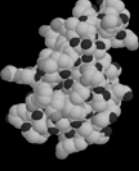
License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: The accession numbers of the sequences used are provided.

Sl. No.	Accn. No.	Description
1	>sp P08621	RU17_HUMAN U1 small nuclear ribonucleoprotein 70 kDa OS=Homo sapiens GN=SNRNP70 PE=1 SV=2
2	>sp P08579	RU2B_HUMAN U2 small nuclear ribonucleoprotein B" OS=Homo sapiens GN=SNRPB2 PE=1 SV=1
3	>SP Q15029	U5S1_HUMAN 116 KDA U5 SMALL NUCLEAR RIBONUCLEOPROTEIN COMPONENT OS=HOMO SAPIENS GN=EFTUD2 PE=1 SV=1
4	>sp Q9Y333	LSM2_HUMAN U6 snRNA-associated Sm-like protein LSm2 OS=Homo sapiens GN=LSM2 PE=1 SV=1
5	>sp O43395	PRPF3_HUMAN U4/U6 small nuclear ribonucleoprotein Prp3 OS=Homo sapiens GN=PRPF3 PE=1 SV=2

Table 2: Analytical parameters and modelled structures of Human snRNPs (Protein components only)

Sl. No.	Accn. No.	Description	Total no of amino acids	No. of lysine	% of lysine	Structure
1	>sp P08621	RU17_HUMAN U1 small nuclear ribonucleoprotein 70 kDa OS=Homo sapiens GN=SNRNP70 PE=1 SV=2	437	22	5.04	
2	>sp P08579	RU2B_HUMAN U2 small nuclear ribonucleoprotein B" OS=Homo sapiens GN=SNRPB2 PE=1 SV=1	225	25	11.11	
3	>SP Q15029	U5S1_HUMAN 116 KDA U5 SMALL NUCLEAR RIBONUCLEOPROTEIN COMPONENT OS=HOMO SAPIENS GN=EFTUD2 PE=1 SV=1	972	59	6.07	
4	>sp Q9Y333	LSM2_HUMAN U6 snRNA-associated Sm-like protein LSm2 OS=Homo sapiens GN=LSM2 PE=1 SV=1	95	8	8.42	
5	>sp O43395	PRPF3_HUMAN U4/U6 small nuclear ribonucleoprotein Prp3 OS=Homo sapiens GN=PRPF3 PE=1 SV=2	683	79	11.57	