

Detecting Patterns in 5' Untranslated Regions of Genes Involved in the Regulation of Blood Pressure and Oxidative Stress

**¹Maria Sanchez-Aguilar, ¹Luz Ibarra-Lara, ²Beatriz Zamora-Lopez,
¹Alicia Sanchez-Mendoza and ³Absalom Zamorano-Carrillo**

¹Department of Pharmacology, National Institute of Cardiology "Ignacio Chávez", Mexico City, Mexico

²Department of Neurosciences, Faculty of Medicine, National Institute of Perinatology,

"Isidro Espinosa de los Reyes" and Department of Mental Health, UNAM, Mexico City, Mexico

³Department of Molecular Biomedicine and Doctoral Program of Biotechnology,
Laboratory of Computational Biophysics and Biochemistry, ENMH-IPN, Mexico City, Mexico

Received 2013-08-26; Revised 2013-09-18; Accepted 2013-11-27

ABSTRACT

Hypertension is a chronic physical condition with multifactorial causes. Although genetic factors have been associated to this pathological condition, characterization of the sequence patterns in the regulatory sites, for example, in 3' or 5' Untranslated Regions (UTR), remains to be explored. In particular, the 5' UTR of genes associated with the regulation of blood pressure and oxidative stress are analyzed in this article. To gain insight into how certain DNA motifs are involved with high blood pressure, we decided to perform a study with sequences from genes reported as regulators of blood pressure and oxidative stress. 5' UTR gene sequences were submitted to patterns recognition by the Multiple Em for Motif Elicitation (MEME) software. Afterward, the motifs obtained were searched for in the Transcription Element Search System (TESS) and Consite platforms, in order to identify the submitted sequence as element response of some transcription factor described previously. Three different motifs in each group of vasorelaxing-and vasocontractile-related gene sequences were detected. In the vasorelaxing group, motif lengths were 39 to 50 nucleotides and were located from -361 to -167 bp before the Open Reading Frame (ORF). In turn, motifs in vasocontractile group sequences were located from -619 to -570 nucleotides, with a length of from 18 to 40 nucleotides. Regarding the nucleic acid content in the motifs found, adenine was more prevalent in vasorelaxing-related sequences with 45% of the average frequency, whereas guanine on those vasocontractile-related sequences with 38%. Distinct motif sequences and variations of nucleotide content in the promoter region of vasorelaxing and vasocontractile-activity related genes were detected. These motifs of each group of genes, with a putative antagonistic role between them, might be a differential cis-regulatory elements of transcriptional machinery.

Keywords: Blood Pressure, Genes, Motifs, Transcription

1. INTRODUCTION

Blood Pressure (BP) is a physiological parameter defined as the force exerted by blood against any unit area of the vessel wall. Elements such as Heart Rate (HR) and Vascular Resistance (VR) are key points to

control this parameter (Chen *et al.*, 2007; Mukkamala *et al.*, 2006).

Biological negative feedback systems regulate blood pressure by the activity of cardiac frequency, vascular resistance and blood volume. These physiological modulators are regulated by some hormones.

Corresponding Author: Absalom Zamorano-Carrillo, Department of Molecular Biomedicine and Doctoral Program of Biotechnology, Laboratory of Computational Biophysics and Biochemistry, ENMH-IPN, Mexico City, Mexico

Specifically, metabolites such as adrenaline, noradrenaline, vasopressin, angiotensin II, dopamine and Reactive Oxygen Species (ROS) promote the elevation of blood pressure. On the contrary, the metabolic activity of enzymes such as catalase, glutathione peroxidase, hemooxygenase and endothelial nitric oxide synthase reduce it. Also, the expression of some receptors (e.g., bradykinin B₂, adenosine A₂, peroxisome proliferator activated receptor gamma) and hormones (brain natriuretic peptide and acetylcholine) is associated to lower blood pressure. All these molecules maintain the blood pressure in homeostatic conditions (Feletou *et al.*, 2010; Vanhoutte, 2006).

Arterial Hypertension (HTA) is defined as the presence of systolic BP values ≥ 140 mmHg and/or diastolic BP values ≥ 90 mmHg. Along with other vascular diseases, HTA is a risk factor for cardiovascular disease (Peralta *et al.*, 2005; Velazquez-Monroy *et al.*, 2002).

Hypertension elicits two different kinds of diffuse structural changes in the systemic microcirculation. Although several causes have been related with increased BP, various research groups have focused their investigations on determining whether genetic mechanisms serve as initial triggers in the genesis of the disease (Renna *et al.*, 2013; Zenteno and Kofman, 2003). Considering the genetic susceptibility of HTA, it has been reported that the classical mendelian rules are not fully applied in the case of BP control, whereby the hypertension shows a complex model inheritance which requires different experimental and theoretical approaches for a better study. In this vein, it has been reported that the heritability of blood pressure or the genetic contribution is able to modulate the arterial tension in a 30-50% of its value (Doris, 2011; Timberlake *et al.*, 2001).

Currently, different molecular techniques allow identify genes associated with pathological conditions suggesting their physiological role. Experimental studies such as linkage studies, candidate genes analysis, punctual mutations, Single Nucleotide Polymorphisms (SNPs) detection and recently bioinformatical analysis provide information for these characterizations (Kunes and Zicha, 2009; Zenteno and Kofman, 2003).

Motif discovery in sequences of interest, as regulatory regions of the gene expression, has been a strategy to find patterns with high information content that play a physiological role. An analysis of Perera *et al.* (2006) proposed novel PPAR gamma target genes in adipocytes by microarray profiling and computational detection of motifs

in the 5' UTR of 182 related genes. Furthermore, they found 26 highly conserved sequences and identified a *drosophila* *serpent* homolog gene that has been reported as a participant in the fat body formation (Perera *et al.*, 2006). Similarly an Motif Discovery Scan (MDscan), Siersbaek *et al.* (2010) found that CCAAT/enhancer binding gen (c/EBP) has a binding motif target of PPAR gamma transcription factor, additional to DR-1 binding element. Suggesting that these adipogenic factors function together as common site to regulate genes (Siersbaek *et al.*, 2010). In this work, we performed an *in silico* analysis of the 5' UTR of genes encoding enzymes or proteins related with blood pressure regulation pathways. In other words, we identified patterns of sequences into the promoter regions of transcription factors that could be investigated as modulators of blood pressure.

2. MATERIALS AND METHODS

The Ensembl databases (http://www.ensembl.org/Homo_sapiens/Gene/) were the repository of the 5'UTR sequences. FASTA format files with sequences spanning 200 or 600 bp were obtained for the vasorelaxing and vasocontractile activity genes. For the first group, we considered catalase (*cat*), glutathione peroxidase (*gpx1*), superoxide dismutase (*sod1*), hemooxygenase types 1 and 2 (*hmox1*, *hmox2*), endothelial nitric oxide synthase (*nos3*), brain natriuretic peptide (*nppb*), beta adrenergic receptor type 2 (*adbr2*), bradykinin receptor type 2 (*bdkrb2*), adenosine receptor 2A (*adora2a*), muscarinic receptor type 3 (*chrm3*), dopamine receptor type 1 (*drd1*), the PPAR gamma isoform (*pparg*), adenine triphosphate-sensitive potassium channel (*kcnj8*) and small/intermediate conductance calcium-activated potassium channels (*kCNN3*, *kcnj1*). Regarding vasocontractile genes, we screened the motifs in xanthin dehydrogenase (*xdh*), lipoxygenase types 5 (*alox5*), 12 (*alox12*) and 15 (*alox15*), voltage-gated potassium channel (*cacng5*), nicotinamide adenine dinucleotide phosphate dehydrogenase (NAD[P]H) oxidase subunits Nox1 (*nox1*), Nox4 (*nox4*), Gp91 (*cybb*), p47 (*ing1*) and p22 (*dynclh1*) (all of these of 600 bp in lenght); also muscarinic receptor type 1 (*chrm1*), serotonin receptor type 2 (*htr2a*), alpha adrenergic receptor 2A (*adra2a*), dopamine receptor type 5 (*drd5*), endothelin receptor type 1 (*edn1*), vasopressin (*avpr1a*) and cyclooxygenase type 2 (*ptgs2*) (all of these 2000 bp in lenght).

With these sequences, we performed in Multiple Em for Motif Elicitation (MEME) program (<http://meme.sdsc.edu/meme/intro.html>) a search of patterns named motifs. This program analyzes and detects sequences with characteristics such as DNA binding or protein interaction domains. Default conditions offered by the web server were utilized; these parameters are minimum and maximum number of nucleotides in the motif. After obtaining the motifs, these were loaded on TESS database (www.ebil.upenn.edu/cgi-bin/tess/tess) to investigate if the motifs have been reported as transcription factors, DNA binding sites, or if they interact with protein functional domains. Similarly, this search was performed into Consite database (<http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite/>) with same purpose. Also, a quantification of nucleotide content in the motifs was carried out. GraphPad Prism program software was used to plot the results.

3. RESULTS

Bioinformatics analysis showed three highly relevant motifs with a high percentage of homology in both groups of genes. For vasorelaxing-activity-related-sequences, Motifs 1 and 2 were 50 nt in length and Motif 3 has a length of 39 nt. Positions of these motifs in reference to zero in the beginning of ORF was -167, -199 and -361 respectively (**Table 1**).

In the group of vasocontractile activity related sequences, the length for Motif 1 was 40 nt, for Motif 2 20 nt and for Motif 3, 18nt. Their positions were found at -570, -619 and -582 nt before the transcription site (**Table 1**). Afterward, we estimated the presence of each nucleotide (adenine, thymine, guanine and cytosine) in each motif found. In genes related with a vasorelaxing effect, adenine was present in 45%, guanine in 30%, thymine in 16% and cytosine in 9% (**Fig. 1A**). In contrast, in sequences of genes related to a vasocontractile effect, guanine appeared to be the most frequent nucleotide (38%) followed by cytosine (29%), thymine (31%) and finally, adenine (1%) (**Fig. 1B**).

According to these results the adenine content is the most evident factor that distinguishes the two studied groups.

To know whether motifs show a similarity with sequences reported as response elements of some characterized genes, the motifs were searched in TESS and Consite databases. With this analysis, a prediction about the affinity of some transcription factors for the

motifs can be obtained and as consequence proteins with a regulatory role in blood pressure can be proposed.

Among the motifs in vasorelaxing genes sequences, Motif 1 was found in several transcription factors (DBF4, ETF, TBP, TFIIA, TMF and YY1) corresponding to TESS database and the search in Consite database four more were included (SOX5, HFH-2, HMG-IY and Evi-1). Motif 2 was associated with the TCF family proteins and LEF-1 according TESS database, whereas using Consite SOX5, HFH-2, HMG-IY, HNF-1 and AML-1 were found. In particular, SOX5, HFH-2 and HMG-IY contain Motif 1 and 2 (**Table 2**) blue letters. Analysis on Motif 3 revealed its existence in MAZ, AP-2 α , SP and H4TF-1 using TESS, Consite was unable to advise similarity with response element contained into that database. These results could be suggested as new transcription factor participating in the vascular relaxation (**Table 2**).

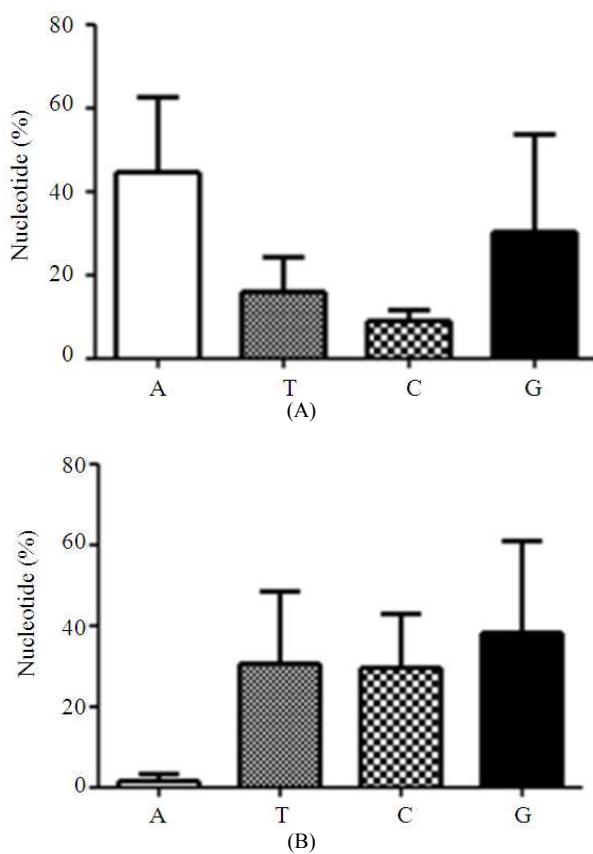


Fig. 1. Nucleotide content of the motifs found. (A) With a putative vasorelaxing activity; (B) with a putative vasocontractile activity. Bars represent mean +/- E.S

Table 1. Motifs detected by MEME program into 5' UTR region of genes studied

No. of motif	Motifs with a putative vasorelaxing activity	Initial position
1	AAAAAAAATTATAACAAGTAAAATAATTATAAGAAAAACAGTAAAAA	-167
2	AGAAAAAAAAAAAGTACCGAAATATGCTTCATTAAAATTATTCTAAAAT	-199
3	GGGGAGGGGGGGGGAGGGACGGGGGGGGAGGGCCGGGCC	-361
	Motifs with a putative vasocontractile activity	
1	TGCTGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	-570
2	TTTGTTCCTTTCTCAGC	-619
3	CCTGCCTCCCTCTCCGGG	-582

Table 2. Transcription factors found in the TESS and Consite databases containing the motifs found by MEME

Vasorelaxing Activity related Genes

Motif 1		Motif 2		Motif 3	
TESS	Consite	TESS	Consite	TESS	Consite
DBF4	SOX5	TCF-2 α	SOX5	MAZ	Any
ETF	HFH-2	TCF-1 α	HFH-2	AP2 α	
TBP	HMG-IY	TCF-1B	HMG-IY	SP	
TFIIA	Evi-1	LEF-1	HNF-1	H4TF-1	
TMF			AML-1		
YY1					
Motif 1	Motif 2	Motif 3			
SP-1	RREB-1	TCF-2 α	FREAC-2	Any	Any
AP-1		LEF-1	Irf-1		
ETF			SOX-5		
ER α			HMG-IY		

Regarding motifs found into genes related to vasocontractile responses, we discovered a similar pattern. TESS analysis showed that Motif 1 is present in SP1, AP-1, ER α and ETF, this last response element was also found in vasorelaxing Motif 1 (**Table 2**), on the top, green letters. Consite database search showed a correspondence with RREB-1 for this motif. TESS analysis, for the Motif 2, provided TCF-2 α and LEF-1, while Consite paired it with SOX5, HMG-IY, FREAC-2 and Irf-1. In the case of Motif 3, similarity was undetectable for both databases. Particularly, SOX5 and HMG-IY were another response elements that contain the Motif 3 obtained from vasorelaxing activity genes (**Table 2**) below, green letters. Therefore, we could suggest that these motifs are putative binding sites of transcription factors that might have a regulatory role in blood pressure.

4. DISCUSSION

In silico studies on biomolecules permit to detect information that could be relevant for their biochemical functions. Although, several three-dimensional models of DNA have been reported,

mainly DNA analysis has been almost limited to understand the combinations of nucleotides in sequences that contain instructions about the cellular machinery. The interest in the information content in key regions has driven the design of algorithms to excavate the mechanisms of data transmission between biomolecules. In particular, flanking regions of ORF are of interest due to their regulatory role. The study of characteristic patterns of group of sequences is a route to this end. In this work, we performed a search of patterns in 5'UTR sequences of vasorelaxing and vasocontractile genes, through the MEME program and a subsequent search into TESS and Consite databases.

The results of the present analysis show that MEME platform allows the identification putative response in genomic sequences that could be related to blood pressure. By discovering of sequence motif, a short pattern of nucleotides can be deemed to have biological significance (Spontaneo and Cercone, 2011). In addition, this platform is a tool that enables a better characterization of target genes (Bailey *et al.*, 2009). As an example, the study developed by Perera *et al.* (2006) permitted to predict novel target genes for PPAR gamma in human cultured adipocytes. They identified three

important genes for fat body composition in drosophila. Because some of them were not found in the transcription database TFBIND, the authors concluded that these motifs could represent novel transcription factors (Perera *et al.*, 2006). The length of DNA motifs sequences suggests that the information content in this number of nucleotides facilitates the spatial availability of DNA. However, some transcription factors exhibit flexibility to bind a few binding sites or bind DNA in some configuration (Reid *et al.*, 2010). In this work, we found that in vasorelaxing group, the motif length was higher (50 nt) with respect to those detected in the vasocontractile group (18 to 40 nt). Besides, due to vasorelaxing sequences are closer to the transcription starting point (-240 bp on average) than vasocontractile sequences (-590 bp on average), both could have different molecular crowding for the transcription, or distinct interaction modes with cis-regulation elements (Gruel *et al.*, 2011).

In our results, nucleotide content of regions was a parameter that distinguishes the vasorelaxing motifs from vasocontractile. Adenine showed the highest proportion (45%) in the first group, whereas guanine was more abundant (38%) in the second, possibly evidencing a regulatory function. Interestingly, it has been suggested that the high content of specific nucleotides could be a point for control of the function. Specifically, guanine-rich sequence in 5'-UTRs of RNA oncogenic can regulate negatively the activity of the telomerase through G-quadruplex formation at the telomeric end (Kaushik *et al.*, 2011; Patel *et al.*, 2007). Therefore, the presence of adenine-rich or guanine-rich regions in the motifs found might be a molecular signal for the differential transcription of genes in distinct contexts related to blood pressure.

5. CONCLUSION

The information in 5' UTR sequences could be a key factor in the gene expression of proteins that regulate the arterial pressure. Although more studies are required, the motifs reported in this work and the content adenine and guanine give rise to propose experimental and theoretical works that could take into account these findings.

6. ACKNOWLEDGEMENT

This research was funded by SIP-IPN (Project 20130466) and COFAA-IPN and SNI-CONACYT.

7. REFERENCES

- Bailey, T.L., M. Boden, F.A. Buske, M. Frith and C.E. Grant *et al.*, 2009. MEME suite: Tools for motif discovery and searching. Nucl. Acids Res., 37: W202-W208. DOI: 10.1093/nar/gkp335
- Chen, X., J.K. Kim, J.A. Sala-Mercado, R.L. Hammond, G. Swamy *et al.*, 2007. Identification of the total peripheral resistance baroreflex impulse response from spontaneous hemodynamic variability: Validation by chronic arterial baroreceptor denervation. Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 22-26, IEEE Xplore Press, Lyon, pp: 1051-1054. DOI: 10.1109/IEMBS.2007.4352475
- Doris, P.A., 2011. The genetics of blood pressure and hypertension: The role of rare variation. Cardiovascular Therapeut., 29: 37-45. DOI: 10.1111/j.1755-5922.2010.00246.x
- Feletou, M., R. Kohler and P.M. Vanhoutte, 2010. Endothelium-derived vasoactive factors and hypertension: Possible roles in pathogenesis and as treatment targets. Curr. Hypertens Res., 12: 267-275. PMID: 20532699
- Gruel, J., M. LeBorgne, N. LeMeur and N. Theret, 2011. Simple Shared Motifs (SSM) in conserved region of promoters: A new approach to identify co-regulation patterns. BMC Bioinform., 12: 365-378. DOI: 10.1186/1471-2105-12-365
- Kaushik, M., S. Kaushik, A. Bansal, S. Saxena, S. Kukreti, 2011. Structural diversity and specific recognition of four stranded G-quadruplex DNA. Curr. Mol. Med., 11: 744-769. PMID: 21999150
- Kunes, J. and J. Zicha, 2009. The interaction of genetic and environmental factors in the etiology of hypertension. Physiol. Res., 58: S33-S41. PMID: 20131935
- Mukkamala, R., J.K. Kim, Y. Li, J. Sala-Mercado, R.L. Hammond *et al.*, 2006. Estimation of arterial and cardiopulmonary total peripheral resistance baroreflex gain values: Validation by chronic arterial baroreceptor denervation. Am. J. Physiol. Heart Circ. Physiol., 290: H1830-H1836. PMID: 16306216
- Patel, D.J., A.T. Phan and V. Kuryavyi, 2007. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic Acids Res., 35: 7429-7455. DOI: 10.1093/nar/gkm711

- Peralta, M.R., A.L. Esqueda, G.P. Hernandez, O.V. Monroy and J.M. Reding *et al.*, 2005. National Re-survey of Arterial Hypertension (RENAHTA). mexican consolidation of the cardiovascular risk factors. national follow-up cohort. Arch. Cardiol. Mex., 75: 96-111. PMID: 15909748
- Perera, R.J., E.G. Marcusson, S. Koo, X. Kang and Y. Kim *et al.*, 2006. Identification of novel PPAR γ target genes in primary human adipocytes. Gene, 369: 90-99. DOI: 10.1016/j.gene.2005.10.021
- Reid, J.E., K.J. Evans, N. Dyer, L. Wernisch and S. Ott, 2010. Variable structure motifs for transcription factor binding sites. BMC Genom., 11: 30-48. DOI: 10.1186/1471-2164-11-30
- Renna, N.F., N. de Las Heras and R.M. Miatello, 2013. Pathophysiology of vascular remodeling in hypertension. Int. J. Hypertens. DOI: 10.1155/2013/808353
- Siersbaek, R., R. Nielsen and S. Mandrup, 2010. PPAR γ in adipocyte differentiation and metabolism-novel insights from genome-wide studies. FEBS Lett., 584: 3242-3249. PMID: 20542036
- Spontaneo, L. and N. Cercone, 2011. Correlating CpG islands, motifs and sequence variants in human chromosome 21. BMC Genom., 12: S10-S18. DOI: 10.1186/1471-2164-12-S2-S10
- Timberlake, D.S., D.T. O'Connor and R.J. Parmer, 2001. Molecular genetics of essential hypertension: Recent results and emerging strategies. Curr. Opin. Nephrol. Hypertens., 10: 71-79. PMID: 11195056
- Vanhoutte, P.M., 2006. Endothelial dysfunction and vascular pathology. Bull. Mem. Acad. R Med. Belg, 161: 529-536. PMID: 17503729
- Velazquez-Monroy, O., M. Rosas-Peralta, A. Lara-Esqueda, G. Pastelin-Hernandez and F. Attie *et al.*, 2002. Hipertensión arterial en Mexico: Resultados de la Encuesta Nacional de Salud (ENSA) 2000. Arch. Cardiol. Mexico, 72: 71-84. PMID: 11933703
- Zenteno, J.C. and S. Kofman-Alfaro, 2003. Aspectos genéticos de la hipertensión arterial. Rev. Med. Hospital General Mexico, 66: 218-223.