Human Immunology 74 (2013) 473–477

Contents lists available at SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Considerations on regulatory sequences of the distal promoter region of the *HLA-G* gene

José Samuel da Silva, Renata Slowik, Maria da Graça Bicalho *

Laboratório de Imunogenética e Histocompatibilidade-LIGH, Departamento de Genética, Centro Politécnico, Universidade Federal do Paraná-UFPR, Av. Coronel Francisco H. dos Santos, Jardim das Américas. Curitiba, Paraná, CEP 81530990, Caixa Postal 19071, Brazil

ARTICLE INFO

Article history:
Received 3 October 2012
Accepted 28 November 2012
Available online 7 December 2012

ABSTRACT

Gene expression in eukaryotic cells is accomplished via association of transcription factors, some of which directly bind to DNA regulatory sequences.

HLA-G codes for an immunoregulatory protein with tissue-specific expression, its unique promoter regulatory region is responsible for this feature. The aim of the present study was to explore motif composition as well as identify haplotypes in the *HLA-G* 5′ distal promoter region. The sample was composed by 176 euro-descendents individuals genotyped by Sequence Based Typing of *HLA-G* distal promoter, encompassing 16 SNPs. Haplotypes were inferred by the expectation maximization algorithm. Only haplotypes with frequency higher than 1% were aligned to check for similarities and differences and thirteen haplotypes remained. For a better understanding of the nucleotide diversity of the analyzed region our approach was to split the whole sequence into two regions. Two contrasting haplotype groups were found in both regions, allowing us to suggest the existence of different transcription factors capable of binding *cis* elements while the intra-group variations suggest the intensity modulation of binding with regulatory factors.

1. Introduction

Gene expression in eukaryotic cells depends on transcriptional and post-transcriptional regulatory signals. Transcriptional regulatory signals are responsible for transcription initiation, that is accomplished via association of certain transcription factors, some of which bind directly to *cis* acting elements within the 5′ upstream regulatory region (5′URR) of coding sequences [1,2]. On the other hand, post-transcriptional regulatory signals regulate the translation of the nascent polypeptide [2]. The *HLA-G* gene encodes a tolerogenic protein initially described in the context of semi-allogenic embryo acceptance during human pregnancy [3–5]. More recently, it has been associated with kidney and heart allograft acceptance, reduced rejection episodes in transplantation [6–8] and with cancer antitumor response [9].

The promoter region of this locus exhibits a structural *cis* element organization which differs from other HLA Class I genes [10]. This unique feature may be responsible for the *HLA-G* gene expression in specific tissues such as the trophoblast or thymic epithelial cells and renders its non-expression under non-pathological conditions [3,11]. The basal promoter region of this locus contains the same *cis* elements that are part of the basal transcription initiation complex of all genomic genes such as CCAAT and the TATA box slightly modified to TCTTAA in the case of *HLA-G* [10,12]. Nevertheless, other *cis* regulatory sequences spread out in the proximal and distal promoter regions, detailed bellow, both with distinctive structural characteristics that make the *HLA-G* promoter region different from its counterparts in other Class I genes [10].

The exact boundaries of the distal and proximal *HLA-G* promoter regions are not yet fully elucidated. The proximal promoter lays assumedly between the −100 and the −200 nucleotide positions [13], where two main regulatory modules can be found: the SXY module and the Enhancer A. The SXY module is composed of four distinct sequences: S, X1, X2 and Y. While S and X1 are conserved in *HLA-G*, X2 and Y are disrupted. As a consequence, the binding of several nuclear factors and the *HLA-G* response to Class II transactivator (CIITA) are absent [14–17]. The Enhancer A is located next to the −200 position (similar to other Class I genes), and contains modified κβ1 and κβ2 sequences, which impact the

Abbreviations: CIITA, class II trans activator; EM, expectation maximization algorithm; GAS, interferon-gama activated site; HLA, human leucocyte antigen; HSE, heat shock element; HSF-1, heat shock factor 1; HWF>1%, haplotypes with frequency higher than 1%; INF, interferon; ISRE, interferon stimulated regulatory element; LCR, locus control region; MHC, major histocompatibility complex; SNP, single nucleotide polymorphism; URR, upstream regulated region; UTR, untranslated region.

* Corresponding author. Fax: +55 4133611729.
E-mail addresses: samuel-bio@hotmail.com (J.S Silva), renata_slowik@ufpr.br (R. Slowik), ligh@ufpr.br (M.G Bicalho).

0198-8859/$36.00 - see front matter © 2012 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.humimm.2012.11.027

binding affinity of NF-κβ [18]. One Interferon–Stimulated Regulatory Element (ISRE), partially deleted in *HLA-G*, is present adjacent to the Enhancer A [19]. This sequence controls the constitutive and interferon-induced expression of Class I genes. The absence of ISRE in the promoter region of *HLA-G* may turn it unresponsive to INF-γ [16,19].

In the distal promoter, the next module is a *cis* Heat Shock Element (HSE), found in the −480 position, that binds the Heat Shock Factor-1 (HSF-1), which is a regulatory stress-induced protein [20]. GAS (interferon Gamma Activated Site) is another regulatory sequence located at the −734 nucleotide position and it is reported to be unresponsive in *HLA-G* [21]. Close to GAS in the position −746 there is another ISRE which is responsive to IFN- [22].

The locus control region (LCR) of *HLA-G* is a long-range *cis*-acting DNA segment, harbored at −1350 to −1100 nucleotide positions [23,24]. Similar to the LCRs present in other mammal genes, it controls the expression of linked genes in a tissue-specific manner, through interactions with various *cis* regulatory elements [25].

Several polymorphic sites have been described in the *HLA-G* 5′ URR region but relatively little attention has been given to how these variations can influence *HLA-G* expression. Binding efficiency of transcriptional and post-transcriptional factors depends on the interplay between protein and nucleotide sequence, so that changes in the promoter region can directly affect the affinity of transcription factors and therefore directly influence protein expression [26,27]. Having this in mind, the aim of this work was to identify specific nucleotide patterns within the promoter region of *HLA-G* through the alignment and comparison of previously sequenced data [28].

2. Methods

2.1. Population assessed and HLA-G 5′URR genotyping

We assessed SNP data of the *HLA-G* regulatory 5′URR region in 176 non-related individuals of European ancestry that had been previously sequenced in an association study by our group [28]. While Costa and colleagues focused their work on the association between *HLA-G* 5'URR SNPs with cases of embryo implantation failure [28], we investigated haplotype patterns that can be relevant for the binding of transcription factors. Details on sample handling, DNA extraction, amplification, sequencing and SNP calling are given in the original work [28].

2.2. Statistical analyses

Haplotypes were inferred through the Expectation Maximization algorithm (EM) implemented in Arlequin v.3.0 [29]. Only haplotypes with frequencies higher than 1% (HWF >1%) were aligned manually and considered for further analyses.

3. Results and discussion

Thirteen HWF >1% were observed in the sample, with frequency summing up to roughly 88% of all haplotypes inferred (Table 1). Based on visual inspection, two discrete conserved regions were identified and named *H-Dist* and *H-Prox*. As shown in Table 1, they coincide with two important transcriptional regions, the Locus Control Region (LCR), which includes SNPs located at −1306 to −1121 positions plus one SNP at −964 position (*H-Dist* region); and one region that comprises three different transcriptional elements: ISRE, GAS (between SNPs at −762 and −725) and HSE (comprehending SNPs at −486 and −477) (*H-Prox* region).

Table 1
Composition and frequencies of haplotypes composed by 16 SNPs ($n = 176$)

Haplotype	Freq.	H-Dist							H-Prox								
		LCR							ISRE	GAS					HSE		
	Freq. 1	Freq. 2	−1306	−1179	−1155	−1140	−1138	−1121	−964	−762	−725	−716	−689	−666	−633	−486	−477	−369
1	0.3075	0.3479	G	A	G	A	A	C	G	C	C	T	A	G	G	A	C	C
7	0.0259	0.0293	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.
4	0.0478	0.0541	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
10	0.0156	0.0176	.	.	.	.	.	.	.	.	T	.	.	.	.	.	G	A
11	0.0114	0.0129	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A
12	0.0114	0.0129	.	.	.	.	.	T	.	.	T	.	.	.	.	.	G	A
13	0.0104	0.0118	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
8	0.0249	0.0281	.	.	.	.	.	T	.	T	.	G	G	T	A	C	G	A
5	0.0454	0.0514	.	.	.	.	.	.	.	T	.	G	G	T	A	C	G	A
2	0.2792	0.3159	A	G	.	T	.	.	A	T	.	G	G	T	A	C	G	A
3	0.0568	0.0643	A	G	A	.	.	.	A	T	.	G	G	T	A	C	G	A
6	0.0282	0.0319	A	G	.	.	.	.	A	T	.	G	G	T	A	C	G	A
9	0.0194	0.0219	A	G	.	T	.	.	A	.	.	.	.	.	.	.	.	.
Total	0.8839	1.0000																

The haplotype number corresponds to the order of the most frequent to less (1–13). Freq 1: relative frequency of HWF >1%. Freq 2: normalized frequencies based on the sum of HWF >1%. The gap after −964 SNP marks the limit between H-Dist and H-Prox. White background represents H1-Dist and H1-Prox while grey background is used for the H2-Dist and H2-Prox groups.