Biochemical and Biophysical Research Communications 446 (2014) 504–518

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Revising angiotensinogen from phylogenetic and genetic variants perspectives

Abhishek Kumar [a,*], Sandeep J. Sarde [b], Anita Bhandari [c]

[a] Department of Genetics & Molecular Biology in Botany, Institute of Botany, Christian-Albrechts-University at Kiel, Kiel, Germany
[b] Master Program Agrigenomics, Christian-Albrechts-University at Kiel, Kiel, Germany
[c] Department of Zoophysiology II, Christian-Albrechts-University at Kiel, Kiel, Germany

ARTICLE INFO

Article history:
Received 7 February 2014
Available online 12 March 2014

Keywords:
Angiotensinogen
Serpin A8
Group V2
Synteny
Phylogenetic analysis
AGT variants

ABSTRACT

Angiotensinogen (AGT) belongs to the serpin superfamily. It acts as the unique substrate of all angiotensin peptides, which generates a spectrum of angiotensin peptides in the renin-angiotensin system and regulates hypertension. This serpin belongs to the multiple member group V2 of the intron encoded vertebrate serpin classification. Despite huge advancements in the understanding of angiotensinogen based on biochemical properties and its roles in the RAS, phylogenetic history of AGT remains forgotten. To date, there is no comprehensive study illustrating the phylogenetic history of AGT. Herein, we investigated phylogenetic traits of AGT gene across vertebrates. Gene structures of AGT gene from selected ray-finned fishes varied in exon I and II with insertions of two novel introns in the core domain for ray-finned fishes at the position 77c and 233c. We that found AGT loci is conserved from lampreys to human and estimated to be older than 500 MY. By comparing AGT protein in 57 vertebrate genomes, we illustrated that the reactive center loop (RCL) of AGT protein became from inhibitory (in lampreys, GTEAKAETVVGIMPI†SMPPT) to non-inhibitory (in human, EREPTESTQQLNKPE†VLEVT) during period of 500 MY. We identified 690 AGT variants by analysis of 1092 human genomes with top three variation classes belongs to SNPs (89.7%), somatic SNVs (5.2%) and deletion (2.9%). There are 32 key residues out of 121 missense variants, which are deleterious for AGT protein, computed by combination of SIFT and PolyPhen V2 methods. These results may have clinical implications for understanding hypertension.

1. Introduction

The renin–angiotensin system (RAS) is an enzyme-linked hormonal cascade that plays an important role in body fluid and cardiovascular regulation. The system is initiated by the action of renin on the precursor protein, angiotensinogen (AGT), which yields the active hormone, angiotensin II. AGT belongs to serine protease inhibitor (serpin) is a single chain glycoprotein with a size of 49.761 kDa, which is the only known precursor of all the peptides generated in the renin–angiotensin system. This gene belongs to clade A in the clade based serpin classification and also known as Serpin A8 [1] and the group V2 in the intron encoded group-wise vertebrate serpin classification system [2]. The protein structure of serpin domain consists of three β-sheets, sA-sC and 8–9 α-helices, hA–hI [3]. The hallmark of serpin biology is the exposed flexible loop (~17–20 residues) known as reactive center loop (RCL), which serves as a bait mimicking a protease substrate that is cleaved between the active sites P1 and P1′ [3]. Some of the serpins have mutations in the RCL region leading into non-inhibitory serpins with specialized functions other than inhibition and human AGT belongs to this category of serpins. R.F. Doolittle identified the AGT gene for the first time in 1983 [4]. In last three decades, AGT has been extensively characterized by biochemical and biophysical methods to demonstrate its roles in the RAS and ultimately controlling blood pressure. However, it lacks a comprehensive molecular phylogenetic analysis, primarily due to fact that notorious problems are countered during the reconstruction of phylogenetic relationships among animal serpins, as several paralogs are found in various animals, particularly for groups V1 and V2 [5]. This suggests that there is a requirement of an investigation on molecular phylogenetic perspectives. Herein, our data disclosed that the detailed molecular phylogeny of AGT genes by combining sequence, genetic variants, gene structures and genomic organization from 57 vertebrates. Furthermore, we have identified 690 genetic variants of human AGT with 121 missense mutations from which 32 are deleterious for AGT protein, which serves an excellent platform for understanding hypertension regulations.

* Corresponding author.
E-mail address: akumar@bot.uni-kiel.de (A. Kumar).

http://dx.doi.org/10.1016/j.bbrc.2014.02.139
0006-291X/

2. Materials and methods

2.1. Sequence collection

We extracted genomic DNA and protein sequences from different vertebrate genomes via Ensembl release 73 (September 2013) [6] using BLAST suite for AGT are provided in Table S1.

2.2. Predicting gene structures and mapping intron positions

We predicted gene structures using AUGUSTUS suite [7] and we combined with gene structure prediction within the Ensembl [6], which ensured accuracy. We used mature human α_1-antitrypsin as the standard sequence for intron position mapping and numbering of intron positions, followed by suffixes a–c for their location as reported previously [5].

2.3. Mining genetic variants of human AGT

We computed genetic variants of human AGT using 1092 human genomes from 14 different populations available in 1000 genomes project [8]. Sorting Intolerant From Tolerant (SIFT) is a software tool, which predicts whether an amino acid substitution affects protein function and it helps in prioritizing substitutions for further study [9]. The SIFT value $\leqslant$0.05 indicates the deleterious effects of missense variants on protein function [9]. Polymorphism Phenotyping V2 (PolyPhen-V2) is a tool that predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations [10]. PolyPhen-V2 score (close to 1) indicates the damaging effects of missense variants on protein function. We used these two methods to predict the impact of these variants on function of AGT.

2.4. Synteny analysis

We scanned chromosomal locus for AGT gene for each species using Ensembl genome browser [6] and mapviewer from the NCBI (website: http://www.ncbi.nlm.nih.gov/mapview/). We constructed synteny maps from fishes to mammalian genomes.

2.5. Sequence and structural analysis

We created protein alignment of AGT using the MUSCLE [11] and edited and visualized in GENEDOC [12] as shown in Fig. S1. We constructed sequence logos of conserved motifs in AGT proteins by Weblogo 3.3 [13].

2.6. Phylogenetic analysis

We constructed two phylogenetic trees aided by Bayesian (2 runs, until average standard deviation of split frequencies was lower than 0.0098, 25% burn-in-period) and Neighbor-Joining methods (1000 bootstraps) using Mrbayes 3.2.1 [14] and MEGA 5.2.2 [15] as following vertebrate serpins (259 serpins) and AGT proteins (57 sequences), respectively. These two trees are based on WAG [5 categories (+*G*, parameter = 4.6121)] and JTT [5 categories (+G, parameter = 1.3836)] protein substitution models, respectively, as their best models were computed in MEGA 5.2.2 [15].

2.7. Protein modelling of AGT protein from Petromyzon marinus

We created structural model of AGT protein from *Petromyzon marinus* using the I-TASSER [16] and we visualized the resulting model using YASARA [17].

3. Results

3.1. Variations in gene structures of AGT in fishes

Human AGT possesses four exons I–IV separated by three introns at the canonical positions 192a, 282b, and 331c in the conserved domain (Fig. 1A). These three introns are maintained in AGT gene from all investigated genomes, which assured membership in the group V2. However, there are changes observed in ray-finned fishes. The canonical exon I is divided into two pieces in ray-finned fishes (except in zebrafish, cave fish and spotted gar) as exons Ia and IIb by an intron at the position 77c with its length ranged from 75 bp (in Fugu) to 267 bp (in Atlantic cod).

Interestingly, exon II is also invaded by a novel intron at the position 233c in selected ray-finned fishes, forming two exons IIa and IIb with sizes 125 bp and 149 bp, respectively. This novel intron has length ranged from 80 bp (in Fugu and medaka) to 137 bp (in Atlantic cod). This insertion of intron within exon II is not observed in lamprey, cave fish, spotted gar and zebrafish. This aspect corroborates that it is a specific feature of selected ray-finned fishes evolved after separation from zebrafish. Exon III is conserved in all vertebrates with size 151–157 bp. Exon IV is ranged from 138 bp (in lamprey) to 192 bp (in chicken). The intron at the position 192a is ranged from 83 bp (in stickleback) to 7002 bp (in Chinese softshell turtle), where as the intron at the position 282b is ranged from 79 bp (in zebrafish) to 2387 bp (in *Petromyzon*) and the intron at the position 331c is ranged from 77 bp (in *Fugu*) to 1693 bp (in zebrafish). These two novel introns are localized in the helix C and the sheet s1B (Fig. S1), respectively. During this study, complete gene structures of AGT gene from spotted gar and *Tetraodon* was not possible to construct due to incomplete assembly. This problem also exists for AGT from lamprey in the Ensembl (December 2013). But, when we used previous version, PMAR3.0 assembly (accession id – GENSCAN00000089208), we are able to construct full-length gene. Noteworthy, AGT genes from *Petromyzon* possess all introns, which are > 1kb. In the nutshell, we report change of the gene structure patterns in selected ray-finned fishes in comparison to that of lampreys, cave fish, spotted gar, zebrafish and tetrapods.

3.2. AGT is conserved on the same genomic fragment from ~500 MY during vertebrate evolution

AGT gene is localized in the human chromosome 1, flanking a heptad of genes, ABCB10, TAF5L, URB2, GALNT2, PGBD5 and COG2 (details in Table S2) on the one side, whereas a tetrad of genes (CAPN9, ARV1, FAM89A, and TRIM67) is present (Fig. 1B). This micro-synteny is maintained in several other mammals including mouse (chromosome 8), rat (chromosome 19) horse (chromosome 1), opossum (chromosome 2) and sheep (chromosome 25). This genomic fragment is also conserved in several birds such as chicken (chromosome 3), duck (scaffold KB743153.1), flycatcher (scaffold JH603366.1), turkey (chromosome 2) and zebra finch (chromosome 3). Reptiles also possess AGT gene flanking same marker genes as of mammals and birds as shown for anole lizard (chromosome 1) and Chinese softshell turtle (scaffold JH208515.1). Amphibians also have these loci but only a heptad of genes are conserved on the one side and other side has genes which are not localized flanking AGT in any other vertebrates analyzed and hence considered as variable region, left blank in Fig. 1B. All fishes have single copy of AGT gene on same syntenic segment as of tetrapods but this segment has some variations. AGT gene is only flanked by COG2 gene on the scaffold JH126749.1 in coelacanth, living fossil lobe-finned fish. Ray-finned fishes with compact genomes possess AGT gene flanking a conserved heptad of genes as