Peptides 69 (2015) 92–102

Contents lists available at ScienceDirect

Peptides

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

Computational analyses and prediction of guanylin deleterious SNPs

William F. Porto[a,b], Octávio L. Franco[a,b,c,*], Sérgio A. Alencar[a,*]

[a] Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília-DF, Brazil
[b] Centro de Análises Proteômicas e Bioquímicas, Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília-DF, Brazil
[c] C S-Inova, Pos-Graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, MS, Brazil

ARTICLE INFO

Article history:
Received 3 March 2015
Received in revised form 10 April 2015
Accepted 12 April 2015
Available online 18 April 2015

Keywords:
Guanylin
GUCA2A
Missense SNPs
Functional impact
Molecular dynamics

ABSTRACT

Human guanylin, coded by the *GUCA2A* gene, is a member of a peptide family that activates intestinal membrane guanylate cyclase, regulating electrolyte and water transport in intestinal and renal epithelia. Deregulation of guanylin peptide activity has been associated with colon adenocarcinoma, adenoma and intestinal polyps. Besides, it is known that mutations on guanylin receptors could be involved in meconium ileus. However, there are no previous works regarding the alterations driven by single nucleotide polymorphisms in guanylin peptides. A comprehensive *in silico* analysis of missense SNPs present in the *GUCA2A* gene was performed taking into account 16 prediction tools in order to select the deleterious variations for further evaluation by molecular dynamics simulations (50 ns). Molecular dynamics data suggest that the three out of five variants (Cys104Arg, Cys112Ser and Cys115Tyr) have undergone structural modifications in terms of flexibility, volume and/or solvation. In addition, two nonsense SNPs were identified, both preventing the formation of disulfide bonds and resulting in the synthesis of truncated proteins. In summary the structural analysis of missense SNPs is important to decrease the number of potential mutations to be *in vitro* evaluated for associating them with some genetic diseases. In addition, data reported here could lead to a better understanding of structural and functional aspects of guanylin peptides.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Guanylin peptides (GPs) are a family of peptide hormones involved in salt absorption. The main physiological function assigned to GPs is related to postprandial hypernatremia prevention, inhibiting the sodium absorption in the intestine and kidneys [72]. There are currently three known mammalian GPs, including guanylin [18], uroguanylin and lymphoguanylin [24,32]. Guanylin and uroguanylin evolved distinctly different amino acid sequences, which enable these peptides to function in a pH-dependent manner. Uroguanylin is more potent than guanylin at an acidic pH (~5.0), whereas guanylin is more potent than uroguanylin at an alkaline pH (~8.0) [31]. Lymphoguanylin has been isolated from opossum and found to be related to guanylin (40% identity) and uroguanylin (80% identity), although it is not known if it has activities similar to these two peptides [24].

Human guanylin is also known as guanylate cyclase activating peptide-1 and is encoded by the *GUCA2A* gene. Its precursor protein contains 115 amino acid residues, whereas the mature protein is composed of the 15 C-terminal residues (101 to 115) of the precursor. Historically, guanylin is compared to the heat-stable enterotoxin (STa) produced by *Escherichia coli*, since both bind to the same receptor [66,74]. However, STa has stronger activity than guanylin, stimulating chlorine secretion, which results in fluid accumulation in gastrointestinal lumen and secretory diarrhea [11,12,18,23,35,52,64]. Although guanylin has a high degree of identity with STa, it possesses two disulfide bonds whereas STa shows three disulfide bonds [18,66]. Also, guanylin can fold into two different isoforms, named A and B, while STa has only one fold, which is structurally similar to isoform A of guanylin, indicating that this isoform could be more active than isoform B [74].

Guanylin acts as an agonist of guanylate cyclase C (GC-C), which is a multidomain membrane-associated receptor located on intestinal epithelial cells within the gastrointestinal tract [9,73]. Upon binding of peptide agonists to the extracellular domain, the intracellular catalytic domain produces cyclic guanosine monophosphate (cGMP), which regulates fluid homeostasis and chloride secretion, preventing dehydration and intestinal obstruction [19,25,68].

The importance of GC-C in normal cell function has been described in a wide number of articles showing associations

* Corresponding authors. Tel.: +55 6192354942.
E-mail addresses: ocfranco@gmail.com (O.L. Franco), sergiodealencar@gmail.com (S.A. Alencar).

http://dx.doi.org/10.1016/j.peptides.2015.04.013
0196-9781/© 2015 Elsevier Inc. All rights reserved.

between GC-C dysfunctions and several diseases. In most cases this is caused by the loss of GC-C activation, either through the loss of the endogenous peptide agonists, or by altered function of GC-C and its downstream mediators [9]. There are reports showing that GPs have a lower expression in patients with colon adenocarcinoma, adenoma and intestine polyps compared to healthy individuals [17,46,72,76]. Such down-regulation of gene expression and peptide loss directly interferes with cell cycle regulation, affecting the balance between epithelial proliferation and differentiation in normal intestinal physiology [60,70]. Also, there is evidence that guanylin is highly expressed in pleomorphic adenoma and Warthin tumors [39].

Recently, Romi et al. have identified in a Bedouin family a homozygous autosomal-recessive mutation in the GC-C encoding gene (*GUCY2C*) which is associated with meconium ileus [65]. This mutation resulted in an amino acid residue substitution (Asp387Gly) in a highly conserved position of the protein located within one of the two essential regions of the ligand-binding domain [65], highlighting that proper binding between GPs and the GC-C receptor is crucial for normal cell function. However, there are no previous works regarding the alterations driven by single nucleotide polymorphisms in the guanylin peptides.

Nowadays, the effects of missense SNPs have been substantially studied using computational tools. Initially, all missense SNPs are classified as deleterious or not by a number of prediction tools and then, the deleterious ones are evaluated by molecular dynamics simulations. There are a wide variety of computational tools used for predicting the effects of missense SNPs on protein function. In general, depending on the strategy used to develop the algorithm, these tools can be classified into four different groups: sequence homology, supervised-learning, protein-sequence and structure, and consensus-based methods.

Disease-causing missense SNPs tend to occur at evolutionarily conserved positions that have an essential role in the structure and/or function of the encoded protein [51]. Therefore, information contained in multiple sequence alignments (MSAs) of homologous protein sequences can help in understanding contemporary deleterious variations in humans [54].

In supervised learning methods such as neural networks and support vector machines, two training sets are constructed: one containing missense SNPs associated with disease and another without disease association. The conservation patterns and physical-chemical properties of the variants in both sets are assessed and used to program the algorithm to “learn” the difference between the variants in the different sets [10,13].

The functional consequences of amino acid residue modifications resulting from missense SNPs depend on the individual amino acids involved and their degree of physical-chemical similarities. Also, other structural implications of missense SNPs include physical disruption of ligand-binding sites, overpacking, backbone strain, loss of electrostatic interactions, and regions crucial for maintaining stability and flexibility [84]. Protein-sequence and structure-based methods account for the impact of missense SNPs on these protein structure modifications.

Using computational tools these approaches could significantly aid in targeting the disease associated mutations. However, once there are many computational tools and different strategies for the prediction of the effects of mutations on protein function, the combination of a variety of these strategies into a consensus classifier could improve significantly the prediction performance [12]. Different aspects of protein sequence and structure are considered by these different methods, such as protein stability, sequence conservation and physical-chemical properties of amino acids [69,80]. However, these methods, even in combination, give only the tip of the iceberg. Modifications in amino-acid composition could affect the native conformation of protein structure and to evaluate the conformational alterations, molecular dynamics simulations are used for an in depth analysis [40]. Molecular dynamics simulations allow evaluating protein flexibility, motion and secondary structure gain or loss over the simulation time. Although there are limitations in computational power (only tens or hundreds of nanoseconds are sampled), molecular dynamics simulations have been extensively used for determining the effect of point mutations. Such approaches were applied for several human proteins, including aurora-A kinase [40,41], ras-related C3 botulinum toxin substrate 1 [42], aldosterone synthase [36], p53 [15], angiogenin [58], protein tyrosine phosphatase 1B [48], receptor tyrosine kinase KIT [61], lamin A/C protein [62] and P protein [38].

Our hypothesis is that missense SNPs could cause significant alterations in the structure of guanylin. To our knowledge, no work has been done before on the structural impact of polymorphisms on any of the known GPs. In this study, we aim to analyze, through the use of an *in silico* approach, the impact of all currently known missense SNPs, which result in amino acid residue substitutions in the coded protein, present in the gene encoding guanylin (*GUCA2A*).

Material and methods

Datasets

The dbSNP database contains SNPs and multiple small-scale variations that include insertions/deletions, microsatellites, and non-polymorphic variants [71]. Using the Variation Viewer navigator from the NCBI [67] (http://www.ncbi.nlm.nih.gov/variation/view), only *GUCA2A* SNPs and non-polymorphic single nucleotide variants (SNVs) deposited in dbSNP were selected. The proguanylin protein sequence in the FASTA format (NCBI Accession: NP_291031.2) was retrieved from the NCBI Protein database (http://www.ncbi.nlm.nih.gov/protein), and the protein structure files of human proguanylin (PDB ID: 1O8R) and the mature peptide (PDB ID: 1GNA) were obtained from the RCSB Protein Data Bank [8,45,74].

The frequency data of missense SNPs found in the *GUCA2A* gene were obtained from the publicly available 1000 Genomes Project (phase I) (http://www.1000genomes.org) [1]. The variant format file (phase 1 release v3.20101123) corresponding to chromosome 1 contained the frequencies of all SNPs identified in the genomes of 1,092 individuals from 14 populations obtained through a combination of low-coverage (2–6×) whole-genome sequence data, targeted deep (50–100×) exome sequencing and dense SNP genotype data. The 14 populations studied were grouped by the predominant component of ancestry into four super-populations: African (AFR) (246 samples), East Asian (ASN) (286 samples), European (EUR) (379 samples) and Ad Mixed American (AMR) (181 samples).

In silico *functional analyses of* GUCA2A *variants*

In order to evaluate the potential functional impact of the obtained *GUCA2A* missense SNPs, a total of 16 computational tools were used, divided into 4 different groups, as described above. We filtered all SNPs that were classified as deleterious by at least three tools in each of the four groups, and denominated these as convergent deleterious predicted SNPs.

Sequence homology-based methods

The following methods based on sequence homology principles were used to produce missense SNP functional predictions: Sorting Intolerant From Tolerant (SIFT) [43], Provean [16], Mutation Assessor [63] and Panther [50].

Download English Version:

https://daneshyari.com/en/article/2005935

Download Persian Version:

https://daneshyari.com/article/2005935

Daneshyari.com