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.

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/

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].

Supervised learning methods

Supervised learning algorithms used for missense SNP impact prediction included neural networks (SNAP) [10], support vector machines (PhD-SNP [13] and SuSPect [83]) and random forests (EFIN) [85].

Protein sequence and structure-based methods

The following methods either combine information from protein sequence and structure or use protein structural information alone to analyze missense variants: PolyPhen [2], Site Directed Mutator (SDM) [82], Fold-X [30] and PoPMuSiC [21].

Consensus-based methods

In order to obtain a consensus score based on many different SNP impact prediction strategies, the following types of consensus software were used: Condel [28], Meta-SNP [12], PON-P2 [55] and PredictSNP [5].

Evolutionary conservation analysis

The ConSurf server is a Bioinformatics tool for estimating the evolutionary conservation of amino acid positions in a protein molecule based on the phylogenetic relations between homologous sequences [14]. Using the GUCA2A proguanylin 3D structure obtained from the Protein Data Bank (PDB ID: 1O8R) [8,45], ConSurf carried out a search for close homologous sequences using CSI-BLAST (3 iterations and 0.0001 e-value cutoff) against the UNIREF-90 protein database [3,79]. The sequences were then clustered and highly similar sequences removed using CD-HIT [47].

Signal peptide prediction

In order to verify the impact of convergent deleterious SNPs on the signal peptide, Phobius [37] and SignalP 4.0 [59] were used for signal peptide topology prediction.

Molecular modeling

One hundred molecular models for each variant were constructed by comparative molecular modeling through MODELLER 9.10 [22], using the structures of 13-guanylin (isoform A) and 5-β-Mercaptopropionate STa (PDB IDs: 1GNA [74] and 1ETN [57], respectively). The models were constructed using the default methods of automodel and environ classes from MODELLER. The final models were selected according to the discrete optimized protein energy score (DOPE score). This score assesses the energy of the models and indicates the best probable structures. The best models were evaluated through PROSA II [81] and PROCHECK [44]. PROCHECK checks the stereochemical quality of a protein structure through the Ramachandran plot, where good quality models are expected to have more than 90% of amino acid residues in most favored and additional allowed regions, while PROSA II indicates the fold quality. Structural visualization was done in PyMOL (http://www.pymol.org).

Molecular dynamics simulation

The molecular dynamics simulations of the ensembles (native wild type guanylin, variants and also STa from *E. coli*) were carried out in water environment, using the Single Point Charge water model [6]. The analyses were performed by using the GROMOS96 43A1 force field and computational package GROMACS 4 [34]. The dynamics used the guanylin and STa three-dimensional models as initial structures, immersed in water, in cubic boxes with a minimum distance of 0.7 nm between the proteins and the edges of the boxes. Chlorine ions were also inserted into the complexes with positive charges in order to neutralize the system charge. Geometry of water molecules was constrained by using the SETTLE algorithm [52]. All atom bond lengths were linked by using the LINCS algorithm [33]. Electrostatic corrections were made by Particle Mesh Ewald algorithm [20], with a cut-off radius of 1.4 nm in order to minimize the computational time. The same cut-off radius was also used for van der Waals interactions. The list of neighbors of each atom was updated every 20 simulation steps of 2 fs. The system underwent an energy minimization using 50,000 steps of the steepest descent algorithm. After that, the system temperature was normalized to 310 K for 100 ps, using the velocity rescaling thermostat (NVT ensemble). Then the system pressure was normalized to 1 bar for 100 ps, using the Parrinello-Rahman barostat (NPT ensemble). The systems with minimized energy, balanced temperature and pressure were simulated for 50 ns by using the leap-frog algorithm.

Analyses of molecular dynamics trajectories

Molecular dynamics simulations were analyzed by means of the backbone root mean square deviation (RMSD), residue root mean square fluctuation (RMSF), radius of gyration and solvent accessible surface area using the `g_rms`, `g_rmsf`, `g_gyrate` and `g_sas` built-in functions of the GROMACS package [34], respectively. The essential dynamics was used to analyze and visualize the overall motions of simulation. The covariance matrices of wild type and variant peptides were constructed using the main chain atoms. The essential dynamics was performed using the `g_covar` and `g_anaeig` utilities of the GROMACS package. Three snapshots were taken from the trajectories in 30, 40 and 50 ns for solvation potential energy calculation. The snapshots were taken using the utility `trjconv` from GROMACS. The solvation potential energy was calculated by APBS under default parameters and using the AMBER force field [4]. A one-sided Student's *t*-test was applied for verifying the differences between the solvation potential energies of wild type and variant guanylin structures with a critical value of 0.05. The Student's *t*-test was calculated using the R package for statistical computing (http://www.r-project.org).

Results

The majority of SNPs occur in non-coding regions of GUCA2A

Out of 135 *GUCA2A* SNPs, 28 are missense, 9 are synonymous, and 2 are nonsense variants. There are also 61 intronic, 1 5′ UTR, 17 3′ UTR, 11 downstream and 6 upstream variants. A graphical representation of the distribution of SNPs in the coding and non-coding regions of the gene represented in terms of percentage is shown in Fig. 1. It can be seen that the majority of the SNPs occur in non-coding regions: 45.2% in introns, 12.6% in 3′ UTR, 8.1% downstream, 4.4% upstream and 0.7% 5′ UTR. In the coding regions, the majority of the SNPs are missense (20.7%), followed by synonymous (6.7%) and nonsense (1.5%) variants.

Frequency information was obtained from the 1000 Genomes Project for ten *GUCA2A* missense SNPs (Table S1). Most of them are rare SNPs with Global Allele Frequency (GAF) values below 1% and occurring only in one of the four populations studied, with the exception of Leu6Pro and Ser7Phe, which have GAF values of 1% and 46%, respectively.

Most convergent deleterious predicted SNPs occur in the mature peptide

Only five SNPs were classified as convergent deleterious predicted SNPs (Leu6Pro, Cys104Arg, Cys112Ser, Gly114Arg and Cys115Tyr) (Table 1). All of them are located within the mature

Table 1
Prediction results of *GUCA2A* missense SNPs analyzed by 16 Bioinformatics tools classified in four different groups.

SNP rs #	Amino acid change[a]	Validation method[b]	Sequence-based[c]				SLM-based[c]				Consensus-based[c]				Structure-based[c]			
			SIFT	Provean	Mutation Assessor	Panther	PhD-SNP	EFIN	SNAP	SuSPect	Condel	MetaSNP	PON-P2	Predict SNP	PolyPhen	SDM	Fold-X	PoP MuSiC
rs117123518:A > G	**Leu6Pro**	1000G, freq.	D	D	D	U	D	D	D	N	D	D	N	D	U	U	U	U
rs2071499:C > T	Ser7Phe	1000G, cluster, freq.	N	N	D	N	N	N	D	N	N	N	N	N	U	U	U	U
rs145299984:A > G	Trp15Arg	1000G, freq.	N	D	N	N	D	D	D	D	N	N	N	D	U	U	U	U
rs201660564:C > G	Val22Leu	1000G, freq.	D	D	D	U	N	N	D	D	D	D	N	D	D	ST	ST	N
rs200135624:A > G	Pro48Ser	1000G	N	N	N	N	N	N	N	N	N	N	N	N	D	DT	DT	N
rs368412585:C > T	Val50Ile	–	N	N	N	N	N	N	N	N	N	N	N	N	N	N	ST	N
rs2232221:C > A	Asn55Lys	1000G, cluster	N	N	N	N	N	N	N	N	N	N	N	N	N	N	ST	DT
rs201386677:C > G	Pro58Arg	1000G, freq.	N	N	N	U	N	N	N	N	N	N	N	N	N	DT	DT	DT
rs111699037:C > T	Gly61Ser	1000G, cluster, freq.	N	N	N	U	N	N	N	N	N	N	N	N	N	DT	DT	N
rs377095864:A > T	Pro66Leu	–	N	D	D	N	N	N	N	N	N	N	N	N	D	DT	DT	DT
rs201815416:A > T	Ser70Cys	cluster	N	D	D	D	N	N	D	N	N	N	N	D	D	ST	ST	N
rs145850777:A > G	Leu81Phe	–	N	D	D	N	N	N	D	D	N	N	N	D	D	DT	DT	DT
rs112720589:C > T	Asn86Ser	cluster	N	N	N	N	N	N	N	N	N	N	N	N	N	N	DT	N
rs35502103:A > T	Asn86Lys	1000G, cluster, freq.	N	N	D	N	N	N	D	N	N	N	N	N	N	DT	DT	DT
rs182438227:A > G	Pro101Leu	1000G	D	D	D	D	N	N	N	N	N	N	N	N	D	ST	DT	DT
rs201516999:A > G	**Cys104Arg**	cluster	D	D	D	U	D	D	D	D	D	D	N	D	D	N	DT	DT
rs139198692:A > G	Ala110Val	1000G, freq.	D	N	N	U	N	N	N	N	N	N	N	N	D	N	ST	N
rs142157365:A > T	**Cys112Ser**	1000G, cluster, freq.	D	D	D	U	D	D	D	D	D	D	U	D	D	DT	DT	DT
rs144271570:C > T	**Gly114Arg**	1000G, cluster, freq.	D	D	D	U	D	D	D	D	D	D	N	D	D	DT	DT	DT
rs139468524:C > T	**Cys115Tyr**	cluster, freq	D	D	D	U	D	D	D	D	D	D	N	D	D	DT	DT	DT

Convergent deleterious SNPs are in bold face.

[a] GUCA2A amino acid positions is relative to GenBank Accession number NP_291031.2.

[b] 1000G: SNP has been sequenced in the 1000 Genomes Project; freq.: Validated by frequency or genotype data: minor alleles observed in at least two chromosomes; cluster: Validated by multiple, independent submissions to the refSNP cluster.

[c] N: Neutral; D: Deleterious; ST: stabilizing; DT: Destabilizing; U: Unknown.

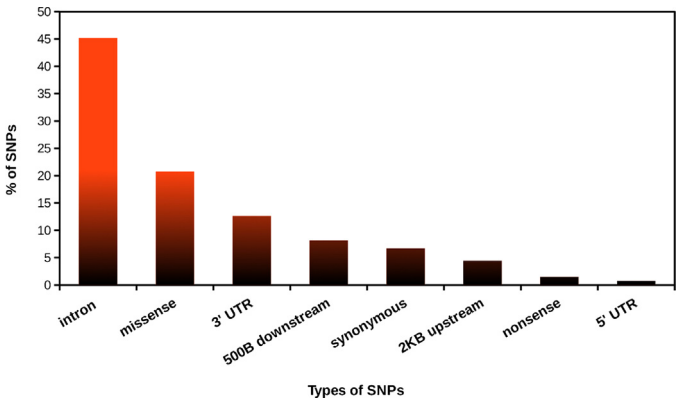

Fig. 1. Distribution of SNPs within the *GUCA2A* gene. The distribution was based on amino acid coding regions (missense, synonymous and nonsense) and on non-coding regions (Up or Downstream, intronic, 5' UTR and 3' UTR).

peptide, with the exception of Leu6Pro, located within the signal peptide. It was also considered as a convergent deleterious predicted SNP, although it cannot be studied in further detail as there are no guanylin precursor solved structures covering the signal peptide.

In addition, ConSurf [7,14] was used to investigate the conservation of the preproguanylin primary sequence. ConSurf exploits evolutionary variation in multiple sequence alignments, in order to determine the degrees of conservation. The results from the ConSurf analysis showed that the majority of the variations (73.7%) occurred in sites classified as "variable" or "average" in terms of conservation, including the location of the Leu6Pro SNP within the signal peptide region (Fig. 2). However, a high degree of conservation was identified in the four convergent deleterious predicted SNPs located in the mature peptide region (residues 101–115), which indicates the evolutionary significance of these alterations. In fact, the mature peptide region showed a higher percentage (75%) of amino acid residues having high levels of conservation than in the rest of the protein (residues 1–114).

The Leu6Pro variant seems not to alter the signal peptide

In order to evaluate the impact of Leu6Pro on the signal peptide, two prediction servers were used (Phobius and SignalP 4.0); however, neither of them indicated any changes in the signal peptide topology (data not shown).

Convergent Deleterious Predicted SNPs Alter the Guanylin Structure, with the exception of Gly114Arg

Regarding the convergent deleterious predicted SNPs in the mature peptide (Fig. 3), according to ConSurf predictions, two of them (Gly114Arg and Cys115Tyr) were predicted as functional residues in exposed regions and the other two were predicted as structural residues (Cys104Arg and Cys112Ser) in buried regions of the protein (Fig. 3B). In fact, all variations are non-conservative and could cause modifications in guanylin properties, especially the three variations involving cysteine residues, which disrupt a disulfide bond.

The structure of isoform A of mature guanylin was used as a template to construct the variant structures and also to reconstruct the native wild type peptide, since the solved structure lacks the residues Pro^{101} and Gly^{102}. In addition, the structure of 5-β-mercaptopropionate STa was used to construct the complete structure of STa from *E. coli*, which was used as an additional control. The validation parameters for molecular modeling are summarized in Table 2. Guanylin shows a very compact structure, stabilized by two cross disulfide bonds (Fig. 3A and B). Similarly, STa is also very compact and has an additional disulfide bond (Fig. 3A).

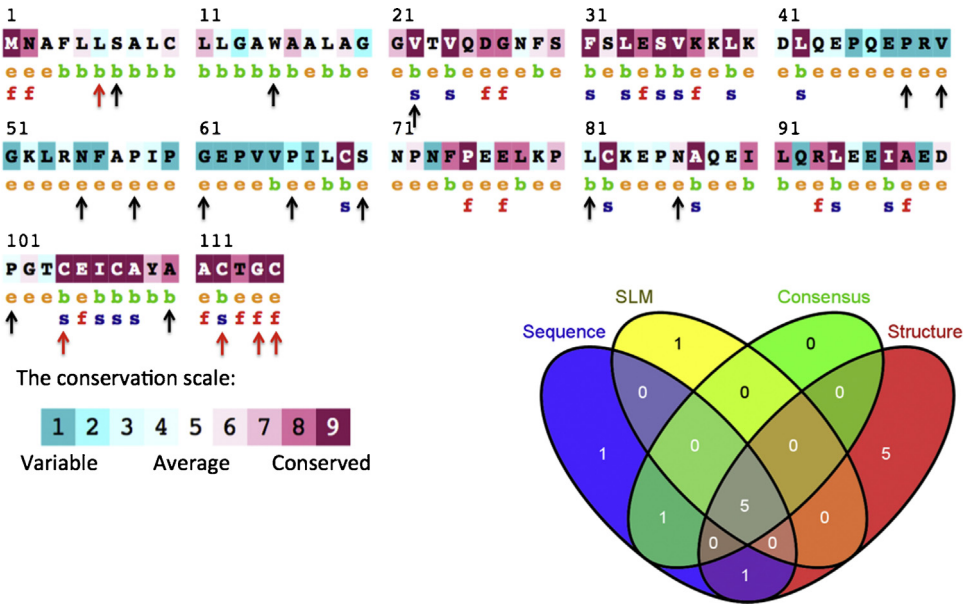

Fig. 2. Conservation pattern of amino acid residues in GUCA2A obtained from multiple sequence alignment using ConSurf. Color intensity increases with degree of conservation. The amino acids are colored based on their conservation grades and conservation levels. A grade of 1 indicates rapidly evolving (variable) sites, which are color-coded in turquoise; 5 indicates sites that are evolving at an average rate, which are colored white; and 9 indicates slowly evolving (evolutionarily conserved) sites, which are color-coded in maroon. The Venn diagram on the bottom right shows the relationships between SNPs predicted as deleterious by the four different groups (Sequence homology, supervised-learning (SLM), protein-sequence and structure, and consensus-based methods). The four convergent deleterious predicted SNPs (classified as deleterious by at least three tools in each of the four different groups) are marked below the peptide sequence as red arrows, and the remaining SNPs are marked as black arrows. e - an exposed residue according to the neural-network algorithm; b - a buried residue according to the neural-network algorithm; f - a predicted functional residue (highly conserved and exposed); s - a predicted structural residue (highly conserved and buried). (For interpretation of the color information in this figure legend, the reader is referred to the web version of the article.)

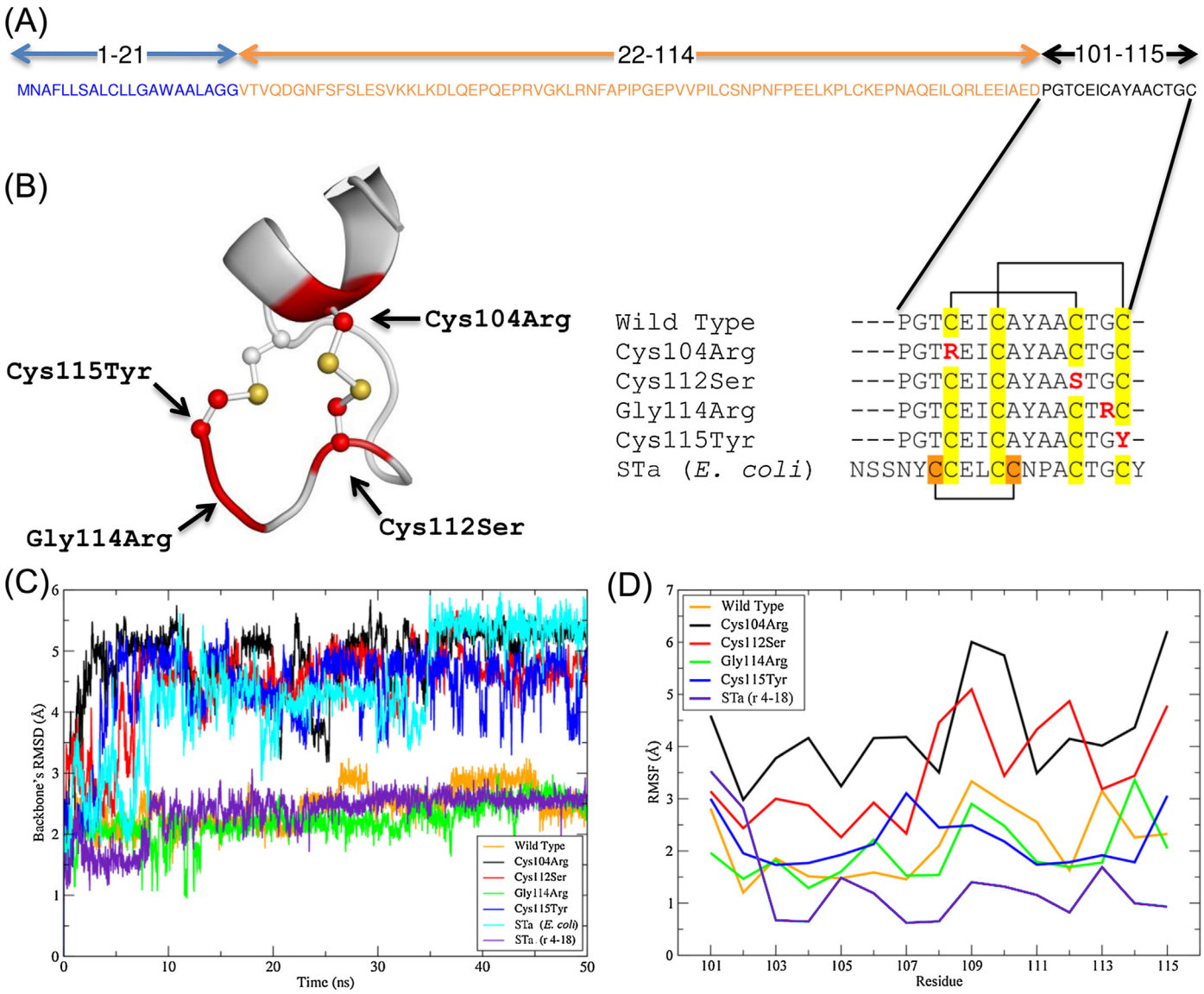

Fig. 3. Guanylin precursor and its variants. (A) The positions of the signal peptide sequence (blue), pro-peptide (brown + black) and mature guanylin peptide (black) are shown above the primary sequence. The two disulfide bonds (shown as brackets) are formed at cysteines 104-112 and 107-115 (highlighted in yellow). For comparisons, the STa from *E. coli* was also included in the sequence alignment (the additional disulfide bonds are highlighted in orange). Each of the amino acid residue changes caused by missense SNPs mapped to these regions are highlighted in red in their respective positions. The amino acid positions are relative to GenBank Accession number NP_291031.2. (B) The isoform A of guanylin used in this study and the amino acid residue changes caused by missense SNPs are mapped into the structure. Disulfide bonds are represented in ball and stick. (C) The Backbone RMSD variation during the simulations. (D) The residues RMS fluctuation. For STa, the RMS fluctuation was measured only for residues 4 to 18, which are the equivalents of mature guanylin, according to the alignment. (For interpretation of the color information in this figure legend, the reader is referred to the web version of the article.)

The molecular dynamics simulation of wild type guanylin indicated that this peptide has very little variation in its structure during the simulation time, showing a stable RMSD variation between 2 and 3 Å (Fig. 3C). On the other hand, the structure of STa showed a huge RMSD variation between 3 and 6 Å (Fig. 3C). However, regarding only the 15 guanylin equivalent residues, it showed a very similar behavior, with a stable RMSD variation between 2 and 3 Å (Fig. 3C). The stable RMSD variation of wild type guanylin and the 15 residues of STa is reflected in the residue RMS fluctuations. Overall, each residue floated less than 3.5 Å (Fig. 3D), with STa

Table 2
Summary of structural validation parameters.

Peptide	DOPE	*Z*-score (ProSA II)	Ramachandran Plot (%)	
			Most favored regions	Additional allowed regions
Wild Type	−829.82526	0.69	90.9	9.1
Cys104Arg	−758.18372	0.96	90.9	9.1
Cys112Ser	−637.07684	1.67	90.9	9.1
Gly114Arg	−789.69537	0.64	75	25
Cys115Tyr	−755.17853	0.78	90.9	9.1
STa (*E. coli*)	−1083.34998	−1.44	93.3	6.7

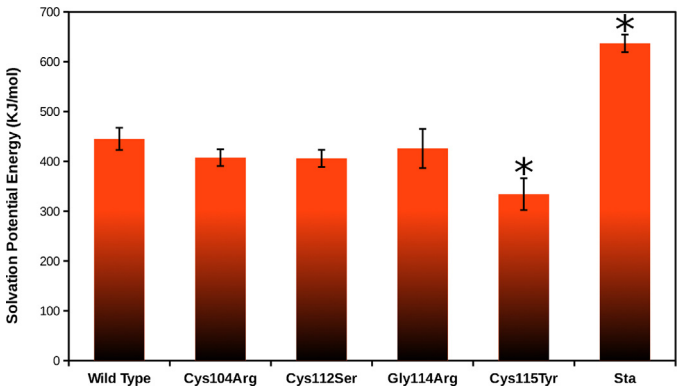

Fig. 4. Solvation Potential Energies of wild type guanylin, its variants and STa. The star indicates a statistical difference between solvation potential energy from the wild type and the variant ($p < 0.05$).

being more stable than wild type guanylin. In addition, the solvation potential energy of STa is higher than the guanylin wild type structure (Fig. 4).

The Gly114Arg variant showed similar behavior to native wild type guanylin. Since it keeps the two disulfide bonds, it also shows a stable RMSD variation between 2 and 3 Å and the residues RMS fluctuation of less than 3.5 Å (Fig. 3C and D). In addition, the radius of gyration and solvent accessible surface area show variations in the same intervals of wild type guanylin (Figs. 5 and 6). Despite the change in net charge, from -1 to 0, due to the variation of a glycine to an arginine residue, this did not cause significant changes in the solvation potential energy (Fig. 4). On the other hand, this variant seems to be less flexible than the wild type structure. The bulk side chain of arginine confers a motion restriction on the peptide structure, when compared with the glycine side chain, as can be observed by the overlap of essential dynamics of Gly114Arg and the wild type peptide (Fig. 7).

In contrast to Gly114Arg, the other three convergent deleterious predicted SNPs seem to alter guanylin structure, since they are involved in the disruption of one disulfide bond. The Cys115Tyr variant showed a backbone RMSD variation higher than 3.5 Å (Fig. 3C), indicating that it adopts a different conformation. However, the residues RMS fluctuation of less than 3.5 Å indicated that the disruption of the disulfide bond between Cys^{107} and Cys^{115} is not sufficient to improve the flexibility. It could also be observed in essential dynamics that for most of the simulation time it is in a less flexible conformation (Fig. 7) than the wild type structure. On the other hand, it could be observed differences in radius of gyration and solvent accessible surface area, where the variant Cys115Tyr has higher values than the wild type (Figs. 5 and 6). In addition, this variant showed a statistically significant difference in solvation potential energy (Fig. 4).

The residues Cys^{104} and Cys^{112} are involved in a disulfide bond and, as indicated by ConSurf, they have structural functions in guanylin structure. The last two variants, Cys104Arg and Cys112Ser, are involved in the disruption of this disulfide bond. Backbone's RMSD showed a variation between 4 and 6 Å (Fig. 3C), while the RMS fluctuation indicated that the structures of Cys104Arg and Cys112Ser are very flexible, especially

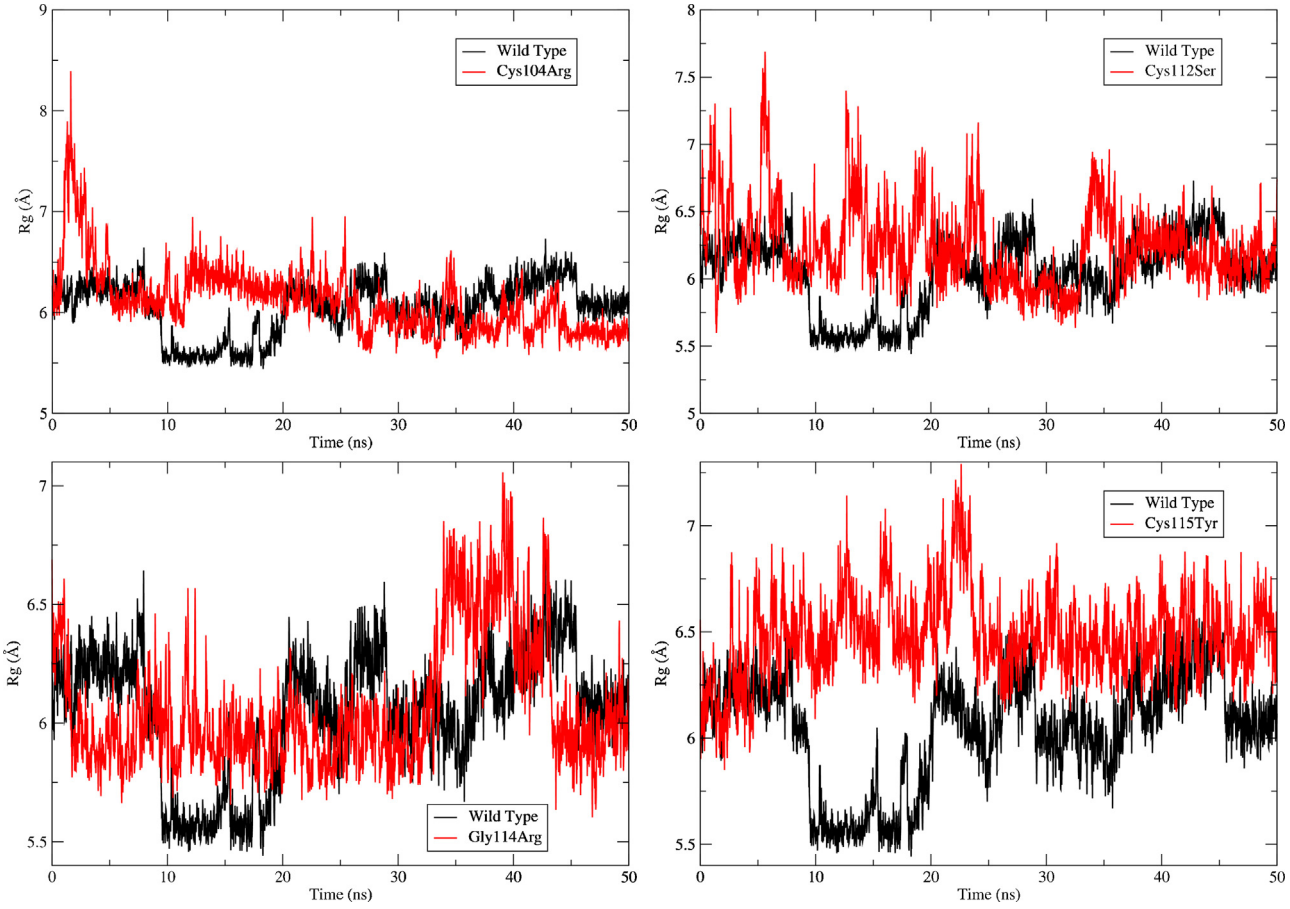

Fig. 5. The radius of gyration variation during the simulations. The black lines represent the native wild type peptide and the red lines the variants as follows: Cys104Arg (top left), Cys112Ser (top right), Gly114Arg (bottom left) and Cys115Tyr (bottom right). (For interpretation of the color information in this figure legend, the reader is referred to the web version of the article.)

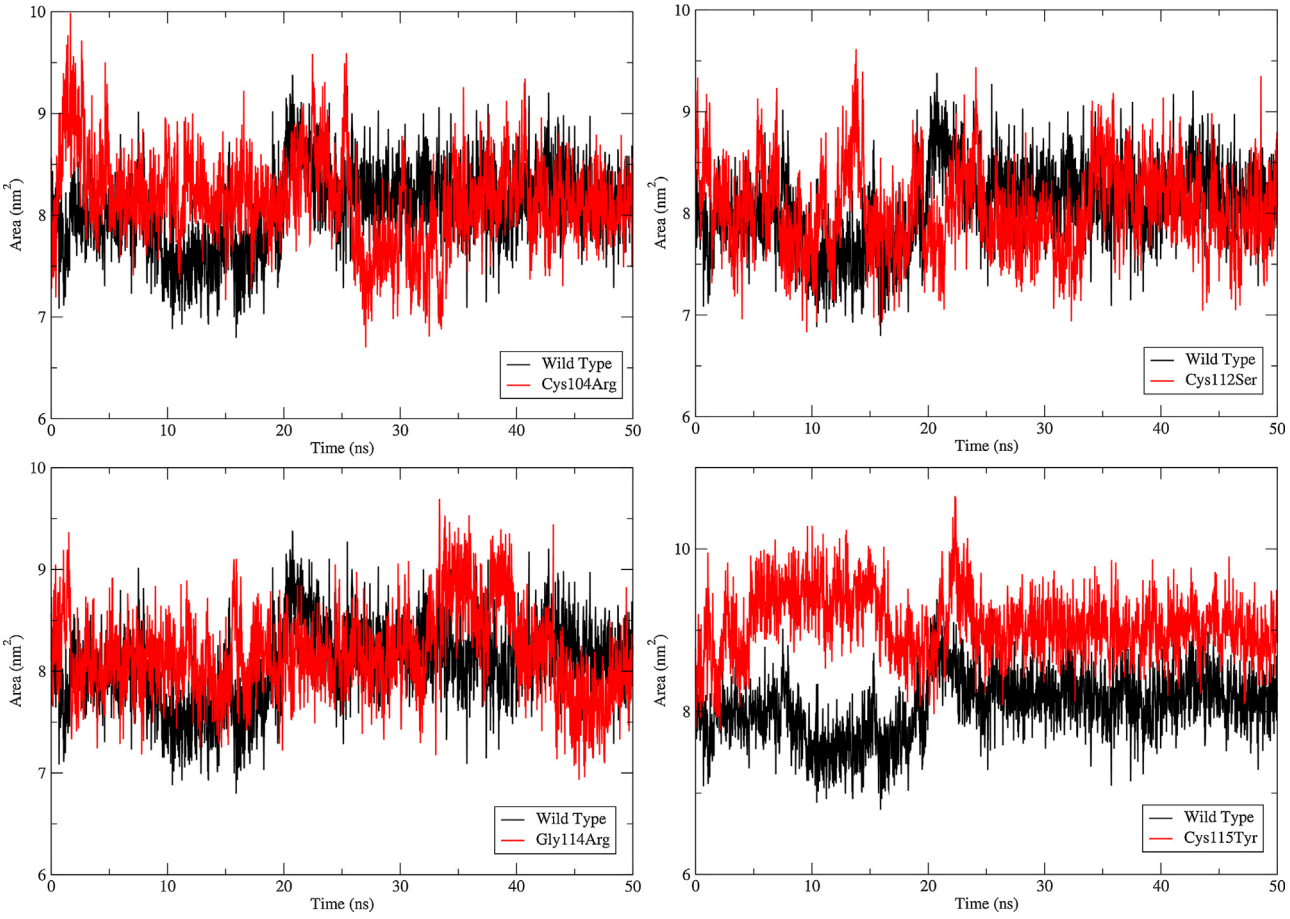

Fig. 6. The solvent accessible surface area variation during the simulations. The black lines represent the native wild type peptide and the red lines the variants as follows: Cys104Arg (top left), Cys112Ser (top right), Gly114Arg (bottom left) and Cys115Tyr (bottom right). (For interpretation of the color information in this figure legend, the reader is referred to the web version of the article.)

the Cys104Arg, which shows all residues floating more than 3 Å (Fig. 3D). The radius of gyration and solvent accessible surface area show variations in the same intervals of wild type guanylin (Figs. 5 and 6). Besides, no statistically significant differences in solvation potential energy were observed for either of these variants (Fig. 4). However, in the essential dynamics analysis, it was observed that these two variants have more flexible structures than the wild type (Fig. 7).

Potentially deleterious nonsense SNPs

Two nonsense SNPs were identified within the *GUCA2A* coding sequence. One of them (Glu89*) is located within the premature peptide sequence, and the other (Tyr109*) occurs within the mature peptide sequence. The Glu89* variant results in the complete absence of the mature peptide, whereas the Tyr109* variant results in a truncated mature peptide containing only 9 residues, none of which are capable of forming disulfide bonds. There is currently frequency information only for Glu89*, which is a rare variation that has a worldwide frequency of 0.05 (Table S1).

Discussion

The inclusion of rare SNPs in genotyping platforms used for association studies has been suggested as being important for the identification of deleterious SNPs [29]. Mutations in evolutionary conserved sites tend to disappear from the population when they impair protein function. The greater the degree of physicochemical constraint violation caused by an amino acid residue change in the protein, the higher is the probability that this replacement could affect the protein stability required for proper functioning, or other biochemical functions, such as ligand binding, catalysis and protein transport [78].

The low frequency of missense SNPs in *GUCA2A* may be a reflection of negative outcomes for protein function caused by these SNPs. With the exception of variant Ser7Phe, which has a notably high worldwide frequency, being present in all populations studied (Table S1), all identified SNPs in *GUCA2A* are rare.

In this report, 16 prediction tools were used to predict the impact caused by missense SNPs present in *GUCA2A*. The basis for the selection of the 16 tools was their individual performance, as described in the literature [26,27]. Consensus prediction seems to represent a more accurate and robust alternative to the predictions delivered by individual tools [5]. In our strategy, the 16 prediction tools were divided in four groups, with four tools in each group. Then, in order to maximize the probability of identifying a functional deleterious SNP, we filtered all SNPs that were classified as deleterious by at least three tools in each of the four analyzed groups.

According to the ConSurf results, all convergent deleterious SNPs located in the mature peptide region alter highly conserved amino acids of guanylin, which indicates that these SNPs could be deleterious. However, in the molecular dynamics simulation, the variant Gly114Arg virtually showed the same behavior as wild

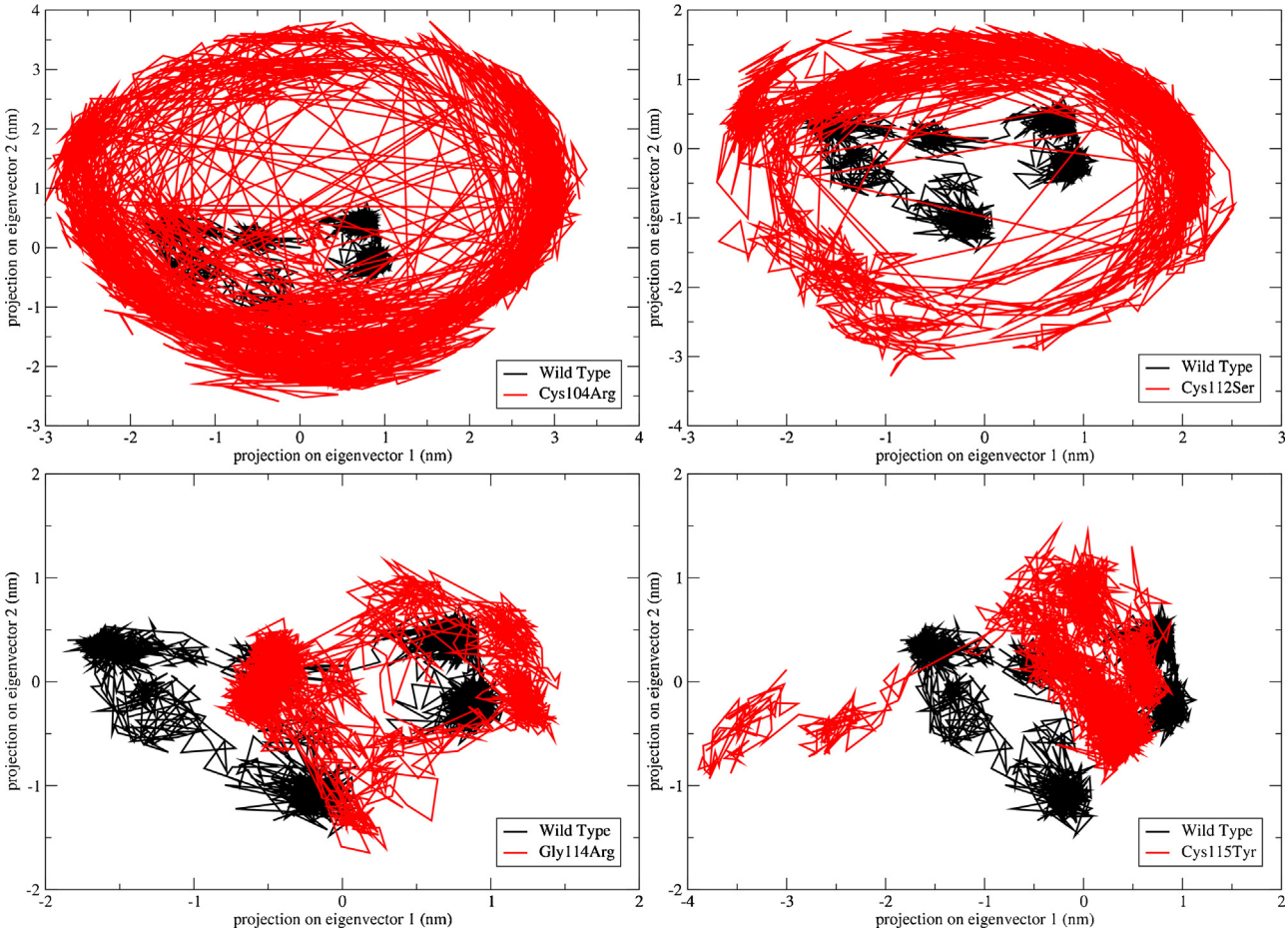

Fig. 7. Projection of the motion of guanylin in phase space along the first two principal eigenvectors. The black lines represent the native wild type peptide and the red lines the variants as follows: Cys104Arg (top left), Cys112Ser (top right), Gly114Arg (bottom left) and Cys115Tyr (bottom right). Cys104Arg and Cys112Ser are more flexible than the native wild type (traces of diagonalized covariance matrix of 8.14963 and 6.4322, respectively), while Gly114Arg and Cys115Tyr are less flexible than wild type (traces of diagonalized covariance matrix of 1.2033 and 1.68232, respectively). The wild type's trace of diagonalized covariance matrix is 1.79422.

type guanylin (Figs. 3–7). Actually, the greatest impact on guanylin structure was observed in variants that led to the disruption of the disulfide bonds. It has been shown previously that the disruption of the disulfide bonds in guanylin and in STa destroy their biological activity [18,75].

The disulfide bond disruption between Cys^{107} and Cys^{115} present in the Cys115Tyr variant leads to a different conformation compared to wild type guanylin. This variant has a less flexible structure (Figs. 3 and 7) and it has a higher radius of gyration (Fig. 5) and also a higher solvent accessible area (Fig. 6), which could result in the lower solvation potential energy (Fig. 4), which in turn, could hinder interaction with guanylin receptors. It is important to emphasize that the solvation potential energy of STa is higher than that of guanylin (Fig. 4), and it could be related to the production of cGMP, since 0.1 μM of STa produces twice as much cGMP than 10 μM of guanylin [66].

In contrast to the Cys115Tyr variant, the variants Cys104Arg and Cys112Ser show the disruption of the disulfide bond formed by Cys^{104} and Cys^{112} and generate very flexible structures. Despite the maintenance of radius of gyration (Fig. 5), solvent accessible area (Fig. 6) and solvation potential energy (Fig. 4), these variants could not be functional due to the loss of the active fold. Nokihara et al. [56] demonstrated that the exchange in cysteine pairs involved in disulfide bond formation alters guanylin activity, since the isoform with disulfide bonds formed by Cys^{104}-Cys^{115} and Cys^{107}-Cys^{112} was not able to bind in rat membranes or stimulate cGMP production.

Regarding the nonsense SNPs (Glu89* and Tyr109*) present in the *GUCA2A* gene, both abnormal proteins which result from these variations are likely to be deleterious. The truncated proteins are usually eliminated by a process called nonsense-mediated mRNA decay, which eliminates the production of the protein [49,53].

The relationship is not clear between the loss of guanylin function, either by convergent deleterious missense or nonsense SNPs, and the development of disease. Guanylin could be involved in defense mechanisms against potentially invasive microorganisms, since this peptide is expressed by Paneth cells of the small intestine, which in turn are increased in number in response to inflammation and bacterial colonization [66]. In knockout mice, the loss of guanylin is reflected in the proliferation of colonic epithelium, which could be related to susceptibility to intestinal adenoma formation and/or progression [77]. However, the same guanylin knockout mice did not show signs of salt malabsorption [77].

Conclusions

Indeed, point mutations driven by missense SNPs are widely studied in a number of human proteins. Here, using an extensive *in silico* evaluation of the potential impact of missense SNPs present in the *GUCA2A* gene, five out of twenty missense SNPs were classified as convergent deleterious predicted SNPs. The molecular dynamics simulations provided a clear insight of conformational changes in guanylin variants, indicating that Cys104Arg, Cys112Ser and Cys115Tyr might be associated with diseases, since they adopt

different structural characteristics, mainly by the disruption of one disulfide bond. In addition, two nonsense SNPs were identified in the *GUCA2A* gene, both preventing the formation of disulfide bonds in the mature peptide. In conclusion, an in depth knowledge of the functional and structural impact of these point mutations may be essential to reduce the number of mutations to be screened in genetic association studies, and data here reported could lead to a better understanding of structural and functional aspects of guanylin peptides.

Acknowledgements

This work was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico); CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior); FAPDF (Fundação de Amparo a Pesquisa do Distrito Federal); and UCB (Universidade Católica de Brasília).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.peptides.2015.04.013

References

[1] Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
[2] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.
[3] Angermüller C, Biegert A, Söding J. Discriminative modelling of context-specific amino acid substitution probabilities. Bioinformatics 2012;28:3240–7.
[4] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98:10037–41.
[5] Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 2014;10:e1003440.
[6] Berendsen H, Postma J, van Gunsteren W, Hermans J. Interaction models for water in relation to protein hydration. Intermol Forces 1981:331–42.
[7] Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, et al. ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics 2004;20:1322–4.
[8] Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–42.
[9] Brierley SM. Guanylate cyclase-C receptor activation: unexpected biology. Curr Opin Pharmacol 2012;12:632–40.
[10] Bromberg Y, Yachdav G, Rost B. SNAP predicts effect of mutations on protein function. Bioinformatics 2008;24:2397–8.
[11] Camici M. Guanylin peptides and colorectal cancer (CRC). Biomed Pharmacother 2008;62:70–6.
[12] Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. BMC Genom 2013;14(Suppl 3):S2.
[13] Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006;22:2729–34.
[14] Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: using evolutionary data to raise testable hypotheses about protein function. Isr J Chem 2013;53:199–206.
[15] Chitrala KN, Yeguvapalli S. Computational screening and molecular dynamic simulation of breast cancer associated deleterious non-synonymous single nucleotide polymorphisms in TP53 gene. PLoS One 2014;9:e104242.
[16] Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One 2012;7:e46688.
[17] Cohen MB, Hawkins JA, Witte DP. Guanylin mRNA expression in human intestine and colorectal adenocarcinoma. Lab Invest 1998;78:101–8.
[18] Currie MG, Fok KF, Kato J, Moore RJ, Hamra FK, Duffin KL, et al. Guanylin: an endogenous activator of intestinal guanylate cyclase. Proc Natl Acad Sci USA 1992;89:947–51.
[19] Cuthbert AW, Hickman ME, MacVinish LJ, Evans MJ, Colledge WH, Ratcliff R, et al. Chloride secretion in response to guanylin in colonic epithelial from normal and transgenic cystic fibrosis mice. Br J Pharmacol 1994;112:31–6.
[20] Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089.
[21] Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinform 2011;12:151.
[22] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci 2007. Chapter 2, Unit 2.9.
[23] Field M, Graf LH, Laird WJ, Smith PL. Heat-stable enterotoxin of Escherichia coli: in vitro effects on guanylate cyclase activity, cyclic GMP concentration, and ion transport in small intestine. Proc Natl Acad Sci USA 1978;75:2800–4.
[24] Forte LR. Guanylin regulatory peptides: structures, biological activities mediated by cyclic GMP and pathobiology. Regul Pept 1999;81:25–39.
[25] Forte LR, Thorne PK, Eber SL, Krause WJ, Freeman RH, Francis SH, et al. Stimulation of intestinal Cl- transport by heat-stable enterotoxin: activation of cAMP-dependent protein kinase by cGMP. Am J Physiol 1992;263:C607–15.
[26] Frousios K, Iliopoulos CS, Schlitt T, Simpson MA. Predicting the functional consequences of non-synonymous DNA sequence variants—evaluation of bioinformatics tools and development of a consensus strategy. Genomics 2013;102:223–8.
[27] Gnad F, Baucom A, Mukhyala K, Manning G, Zhang Z. Assessment of computational methods for predicting the effects of missense mutations in human cancers. BMC Genom 2013;14(Suppl 3):S7.
[28] Gonzalez A, Lo N. Improving the assessment of the outcome of nonsynonymous snvs with a consensus. Am J Hum Genet 2011;53:440–9.
[29] Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. Am J Hum Genet 2008;82:100–12.
[30] Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. J Mol Biol 2002;320:369–87.
[31] Hamra FK, Eber SL, Chin DT, Currie MG, Forte LR. Regulation of intestinal uroguanylin/guanylin receptor-mediated responses by mucosal acidity. Proc Natl Acad Sci USA 1997;94:2705–10.
[32] Hamra FK, Forte LR, Eber SL, Pidhorodeckyj NV, Krause WJ, Freeman RH, et al. Uroguanylin: structure and activity of a second endogenous peptide that stimulates intestinal guanylate cyclase. Proc Natl Acad Sci USA 1993;90:10464–8.
[33] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–72.
[34] Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–47.
[35] Hughes JM, Murad F, Chang B, Guerrant RL. Role of cyclic GMP in the action of heat-stable enterotoxin of Escherichia coli. Nature 1978;271:755–6.
[36] Jia M, Yang B, Li Z, Shen H, Song X, Gu W. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2 Gene. PLoS One 2014;9.
[37] Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res 2007;35:W429–32.
[38] Kamaraj B, Purohit R. Computational screening of disease-associated mutations in OCA2 Gene. Cell Biochem Biophys 2013;68:97–109.
[39] Kulaksiz H, Rehberg E, Stremmel W, Cetin Y. Guanylin and functional coupling proteins in the human salivary glands and gland tumors: expression, cellular localization, and target membrane domains. Am J Pathol 2002;161:655–64.
[40] Kumar A, Purohit R. Use of long term molecular dynamics simulation in predicting cancer associated SNPs. PLoS Comput Biol 2014;10:e1003318.
[41] Kumar A, Rajendran V, Sethumadhavan R, Purohit R. Computational investigation of cancer-associated molecular mechanism in Aurora A (S155R) mutation. Cell Biochem Biophys 2013;66:787–96.
[42] Kumar A, Rajendran V, Sethumadhavan R, Purohit R. Molecular dynamic simulation reveals damaging impact of RAC1 F28L mutation in the Switch I Region. PLoS One 2013;8:e77453.
[43] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–81.
[44] Laskowski R, Macarthur M, Moss D, Thornton J. {PROCHECK}: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–91.
[45] Lauber T, Neudecker P, Rösch P, Marx UC. Solution structure of human proguanylin: the role of a hormone prosequence. J Biol Chem 2003;278:24118–24.
[46] Li P, Lin JE, Schulz S, Pitari GM, Waldman SA. Can colorectal cancer be prevented or treated by oral hormone replacement therapy? Curr Mol Pharmacol 2009;2:285–92.
[47] Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658–9.
[48] Liu M, Wang L, Sun X, Zhao X. Investigating the impact of Asp181 point mutations on interactions between PTP1B and phosphotyrosine substrate. Sci Rep 2014;4:5095.
[49] Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 2004;5:89–99.
[50] Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, et al. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res 2005;33:D284–8.
[51] Miller MP, Kumar S. Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 2001;10:2319–28.

[52] Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 1992;13:952–62.
[53] Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci 1998;23:198–9.
[54] Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 2006;7:61–80.
[55] Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One 2015;10:e0117380.
[56] Nokihara K, Wray V, Ando E, Naruse S, Hayakawa T. Synthesis, solution structure, binding activity, and cGMP activation of human guanylin and its disulfide isomer. Regul Pept 1997;70:111–20.
[57] Ozaki H, Sato T, Kubota H, Hata Y, Katsube Y, Shimonishi Y. Molecular structure of the toxin domain of heat-stable enterotoxin produced by a pathogenic strain of Escherichia coli. A putative binding site for a binding protein on rat intestinal epithelial cell membranes. J Biol Chem 1991;266:5934–41.
[58] Padhi AK, Jayaram B, Gomes J. Prediction of functional loss of human angiogenin mutants associated with ALS by molecular dynamics simulations. Sci Rep 2013;3:1225.
[59] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8:785–6.
[60] Pitari GM, Di Guglielmo MD, Park J, Schulz S, Waldman SA. Guanylyl cyclase C agonists regulate progression through the cell cycle of human colon carcinoma cells. Proc Natl Acad Sci USA 2001;98:7846–51.
[61] Purohit R. Role of ELA region in auto-activation of mutant KIT receptor: a molecular dynamics simulation insight. J Biomol Struct Dyn 2014;32:1033–46.
[62] Rajendran V, Purohit R, Sethumadhavan R. In silico investigation of molecular mechanism of laminopathy caused by a point mutation (R482W) in lamin A/C protein. Amino Acids 2012;43:603–15.
[63] Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011;39:37–43.
[64] Robins-Browne RM. Traditional enteropathogenic Escherichia coli of infantile diarrhea. Rev Infect Dis 1987;9:28–53.
[65] Romi H, Cohen I, Landau D, Alkrinawi S, Yerushalmi B, Hershkovitz R, et al. Meconium ileus caused by mutations in GUCY2C, encoding the CFTR-activating guanylate cyclase 2C. Am J Hum Genet 2012;90:893–9.
[66] de Sauvage FJ, Keshav S, Kuang WJ, Gillett N, Henzel W, Goeddel DV. Precursor structure, expression, and tissue distribution of human guanylin. Proc Natl Acad Sci 1992;89:9089–93.
[67] Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2011;39:D38–51.
[68] Schulz S, Green CK, Yuen PST, Garbers DL. Guanylyl cyclase is a heat-stable enterotoxin receptor. Cell 1990;63:941–8.
[69] Schwartz CE, Chen C-F. Progress in detecting genetic alterations and their association with human disease. J Mol Biol 2013;425:3914–8.
[70] Shailubhai K, Yu HH, Karunanandaa K, Wang JY, Eber SL, Wang Y, et al. Uroguanylin treatment suppresses polyp formation in the Apc(Min/+) mouse and induces apoptosis in human colon adenocarcinoma cells via cyclic GMP. Cancer Res 2000;60:5151–7.
[71] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.
[72] Sindic A. Current understanding of guanylin peptides actions. ISRN Nephrol 2013:813648.
[73] Singh S, Lowe DG, Thorpe DS, Rodriguez H, Kuang WJ, Dangott LJ, et al. Membrane guanylate cyclase is a cell-surface receptor with homology to protein kinases. Nature 1988;334:708–12.
[74] Skelton NJ, Garcia KC, Goeddel DV, Quan C, Burnier JP. Determination of the solution structure of the peptide hormone guanylin: observation of a novel form of topological stereoisomerism. Biochemistry 1994;33:13581–92.
[75] Staples SJ, Asher SE, Giannella RA. Purification and characterization of heat-stable enterotoxin produced by a strain of E. coli pathogenic for man. J Biol Chem 1980;255:4716–21.
[76] Steinbrecher KA, Tuohy TM, Heppner Goss K, Scott MC, Witte DP, Groden J, et al. Expression of guanylin is downregulated in mouse and human intestinal adenomas. Biochem Biophys Res Commun 2000;273:225–30.
[77] Steinbrecher KA, Wowk SA, Rudolph JA, Witte DP, Cohen MB. Targeted inactivation of the mouse guanylin gene results in altered dynamics of colonic epithelial proliferation. Am J Pathol 2002;161:2169–78.
[78] Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res 2005;15:978–86.
[79] Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 2007;23:1282–8.
[80] Teng S, Michonova-Alexova E, Alexov E. Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions. Curr Pharm Biotechnol 2008;9:123–33.
[81] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35:W407–10.
[82] Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 2011;39:W215–22.
[83] Yates CM, Filippis I, Kelley LA, Sternberg MJE. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. J Mol Biol 2014;426:2692–701.
[84] Yue P, Li Z, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol 2005;353:459–73.
[85] Zeng S, Yang J, Chung BH-Y, Lau YL, Yang W. EFIN: predicting the functional impact of nonsynonymous single nucleotide polymorphisms in human genome. BMC Genom 2014;15:455.