Journal of Theoretical Biology 410 (2016) 9–17

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Structural impact analysis of missense SNPs present in the uroguanylin gene by long-term molecular dynamics simulations

Antonio C.S. Marcolino [a], William F. Porto [a,b], Állan S. Pires [a,b], Octavio 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] S-Inova Biotech, Pós-graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, MS, Brazil

HIGHLIGHTS

- The impact of missense SNPs on the GUCA2B protein was predicted by *in silico* tools.
- Seven convergent deleterious SNPs were further analyzed by long-term MD simulations.
- Four SNPs could cause major damage to the protein on a structural level.

ARTICLE INFO

Article history:
Received 15 June 2016
Received in revised form
19 August 2016
Accepted 8 September 2016
Available online 9 September 2016

Keywords:
Uroguanylin
Missense SNPs
Molecular dynamics
Structural impact

ABSTRACT

The guanylate cyclase activator 2B, also known as uroguanylin, is part of the guanylin peptide family, which includes peptides such as guanylin and lymphoguanylin. The guanylin peptides could be related to sodium absorption inhibition and water secretion induction and their dysfunction may be related to various pathologies such as chronic renal failure, congestive heart failure and nephrotic syndrome. Besides, uroguanylin point mutations have been associated with essential hypertension. However, currently there are no studies on the impact of missense SNPs on uroguanylin structure. This study applied *in silico* SNP impact prediction tools to evaluate the impact of uroguanylin missense SNPs and to filter those considered as convergent deleterious, which were then further analyzed through long-term molecular dynamics simulations of 1 μs of duration. The simulations suggested that all missense SNPs considered as convergent deleterious caused some kind of structural change to the uroguanylin peptide. Additionally, four of these SNPs were also shown to cause modifications in peptide flexibility, possibly resulting in functional changes.

1. Introduction

Guanylin peptides compose a family of peptide hormones involved in salt absorption processes. Their discovery is commonly related to the *Escherichia coli* heat-stable enterotoxin (ST), which can cause diarrhea and also gut infection (Currie et al., 1992). The guanylate cyclase activator 2B, also known as uroguanylin, is part of the guanylin peptide family, which also includes guanylin and lymphoguanylin.

Peptides from the guanylin family can be produced in the intestine, kidneys, adrenal glands, reproductive system, lungs and pancreas (Nakazato et al., 1998). While the role of lymphoguanylin is still unknown, guanylin is secreted into the intestinal lumen, after salt ingestion. Guanylin activates the guanylate cyclase C (GC-C) receptor with cyclic guanosine monophosphate (cGMP) as a second messenger. GC-C receptor activation leads to sodium absorption inhibition and induced water secretion (Sindic, 2006). Besides, in kidneys, uroguanylin induces natriuresis and diuresis, along with increased cGMP urinary levels.

The human uroguanylin coding gene, *GUCA2B*, is located on chromosome 1p33-p34 and has approximately 2.5 kb, including three exons and two introns (Miyazato et al., 1997). Uroguanylin peptide production can be divided into two stages. Firstly, the precursor pro-uroguanylin, which has 112 amino acid residues, is cleaved into pro-uroguanylin, containing 86 amino acid residues. Then, after another cleavage, the uroguanylin mature active form is released (Miyazato et al., 1996). Mature uroguanylin is a peptide

* Corresponding author.
E-mail address: sergiodealencar@gmail.com (S.A. Alencar).

http://dx.doi.org/10.1016/j.jtbi.2016.09.008
0022-5193/

composed of 16 amino acid residues and shows in its structure four cysteines on the mature peptide domain, which are conserved in the guanylin family. Disulfide bonds are made between Cys^{I} and Cys^{III}, and Cys^{II} and Cys^{IV}. There are two known uroguanylin isoforms: isoform A and isoform B. Isoform A resembles the structure of the *Escherichia coli* heat-stable enterotoxin.

Uroguanylin dysfunction has been correlated with several diseases: knockout mice lacking uroguanylin showed higher blood pressure (Lorenz et al., 2003); patients with chronic renal failure presented higher uroguanylin plasma levels (Nakazato et al., 1996); congestive heart failure was associated with an increase in uroguanylin excretion in the urine (Carrithers et al., 2000); and nephrotic syndrome was related to higher uroguanylin plasma levels and low urine excretion of uroguanylin (Kinoshita et al., 1999). However, there are no in-depth studies of single nucleotide polymorphisms (SNPs) in *GUCA2B* and their respective consequences at protein level.

SNPs are the most common kind of variation in the human genome. SNPs that occur in a gene's coding region and result in an amino acid substitution at the coded protein corresponding region are known as missense SNPs, also called non-synonymous SNPs (nsSNPs). In most cases, this variation is neutral or has little effect on the protein function. However, when this variation causes an alteration in the protein's structure, this change could also result in protein function alteration, which could lead to a disease (Ng and Henikoff, 2006; Yates and Sternberg, 2013). Several approaches have been used to study the impact of missense SNPs on a protein structure; however testing these impacts in a laboratory can be expensive, so analysis by computational tools has become a powerful and inexpensive approach for preliminary analyses (Shen et al., 2006). As a result, a number of *in silico* tools have been developed to predict the effect of missense SNPs, including methods based on sequence homology (Choi et al., 2012), supervised-learning (Zhao et al., 2014), protein-sequence and structure (Dorn et al., 2014), and consensus-based (González-Pérez and López-Bigas, 2011). Nevertheless, as there are limitations to these methods, a more refined approach such as molecular dynamics simulations helps to evaluate further the impact of these alterations in terms of protein flexibility, motion and secondary structure gain or loss.

Early investigations of low-frequency global modes of oscillation of atoms within proteins have stimulated a series of studies about biomacromolecules from a dynamic point of view (Chou and Chen, 1977; Martel, 1992). Later, as other studies clearly demonstrated this phenomenon (Chou, 1988), works with protein and DNA molecules showed that low-frequency vibrations possess significant biological functions (Zhou, 1989), and are key in order to gain insights into the mechanisms underlying the dynamic process occurring in biomacromolecules (Chou, 1988). Furthermore, many important biological functions and dynamic mechanisms of these molecules, such as switch between active and inactive states (Martel, 1992), cooperative effects (Chou, 1988), allosteric transition intercalation of drugs into DNA (Chou and Mao, 1988), and assembly of microtubules (Chou et al., 1994), can be revealed by studying the low-frequency internal motions (Chou, 1988). More recently this knowledge has been also applied for medical treatments (Madkan et al., 2009), including mitigation of inflammation and stimulation of classes of genes following onset of illness and injury (Gordon, 2007). Therefore, in order to really understand the biomacromolecules' mechanisms of action, it is crucial to consider not only the static structural information but also the dynamical information acquired by studying their internal motions. To achieve this, molecular dynamics simulations are one of the most adequate analysis. Among several studies, this computational method has been used to reveal the switch mechanism between active and inactive states of protein tyrosine phosphatases 1B (J.-F. Wang et al., 2009a), to investigate cytochrome P450 2C19 binding in pharmacogenetics studies (Wang et al., 2007), and to reveal the relationship between DNA sequence recognition and catalytic specificity (Yang et al., 2012).

Molecular dynamics simulations are also widely used to obtain information about protein conformational evolution, as well as kinetic and thermodynamic information. Also, simulations can provide details about particle movement given a determined time (Adcock and Mccammon, 2006). Therefore, this method can be very useful to gain insight into the impact of missense SNPs on protein structure. Most of the studies about the impact of missense SNPs on protein structure are based on short-term molecular dynamics simulations. The simulation time applied in such an analysis is variable, ranging from 1 ns (Jia et al., 2014) up to 200 ns (Kumar and Purohit, 2014). However, studies lasting longer, such as a 1 μs simulation, performed in 1998 (Duan and Kollman, 1998) and a 10 μs simulation (Freddolino et al., 2008) have not been used to study the impact of missense SNPs on protein structure. The advantage of using longer simulations is that it allows us to visualize how proteins fold and unfold during a given time. It also allows a better understanding of the proteins, generating more information about how each behaves in a controlled environment (Klepeis et al., 2009).

In the present study, using an *in silico* approach, the structural impact of all known uroguanylin missense SNPs was evaluated by a combination of SNP prediction tools and long-term molecular dynamics simulations throughout 1 μs.

2. Methodology

2.1. SNP data

The uroguanylin (*GUCA2B*) gene data was obtained from the National Center for Biotechnology Information (NCBI) website (NCBI access code: NP_009033.1) and its structure, the Isoform A, was obtained from the Protein Data Bank (PDB) (Berman et al., 2000) (PDB access code: 1UYA). The information about the missense SNPs was collected from NCBI's Variation Viewer (http://www.ncbi.nlm.nih.gov/variation/view/), including the SNP database (dbSNP) (Sherry et al., 2001), Single Nucleotide Variant and Missense Variant filters.

2.2. SNP impact prediction tools

All missense SNPs were submitted to twelve protein analysis programs. The programs were split into three categories, each containing four programs: sequence homology-based methods, supervised learning methods and consensus-based methods. Additionally, missense SNPs that occurred in regions present in a guanylin protein structure available in PDB also underwent analysis by four other programs based on protein sequence and structure methods. After this step, all missense SNPs resulting in "deleterious" classification by at least three programs within each category were selected for molecular dynamics simulations.

2.2.1. Sequence homology-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 (Miller and Kumar, 2001). Therefore, information contained in multiple sequence alignments (MSAs) of homologous protein sequences can help in understanding contemporary deleterious variations in humans. A missense SNP can lead to an amino acid with altered physicochemical properties compared to the original one and this change can in turn be used to predict functional consequence to the protein (Ng

and Henikoff, 2006). The following algorithms are based on these principles and combine MSAs, generated through a variety of methods, with scoring functions based on measures of amino acid similarity to produce functional predictions: Sorting Intolerant From Tolerant (SIFT) (Kumar et al., 2009), Provean (Choi et al., 2012), MutationAssessor (Reva et al., 2011) and Panther (Mi et al., 2005).

2.2.2. Supervised learning methods

The supervised learning methods include Neural Networks (SNAP) (Bromberg et al., 2008), Support Vector Machines (PhD-SNP) (Capriotti et al., 2006), SuSPect (Yates et al., 2014) and Random Forests (EFIN) (Zeng et al., 2014). In neural networks and support vector machine methods, two training sets are constructed: one containing variants associated with disease and another without disease association. The conservation patterns and physical-chemical properties of the variants on both sets are assessed and used to program the algorithm to "learn" the difference between the variants in the different sets. Also used for SNP impact prediction, Random Forests are an ensemble learning method used for classification, which constructs a multitude of decision trees and outputs the prediction as the majority vote from all individual trees (Zeng et al., 2014).

2.2.3. Consensus-based methods

There are currently many computational tools widely employed for the prediction of the effects of mutations on protein function. The following softwares combine a variety of methods into a consensus classifier, resulting into significantly improved prediction performance: Condel (González-Pérez and López-Bigas, 2011), Meta-SNP (Capriotti et al., 2013), PON-P2 (Niroula et al., 2015) and PredictSNP (Bendl et al., 2014).

2.2.4. Protein sequence and structure-based methods

The functional consequences of amino acid residue changes will depend on the individual amino acids involved and their degree of physical-chemical similarities. Also, other structural implications of missense variants include physical disruption of ligand-binding sites, overpacking, backbone strain, loss of eletrostatic interactions, and regions crucial for maintaining stability and flexibility (Yue et al., 2005). The following methods either combine information from protein sequence and structure or use protein structural information alone to analyze missense variants: PolyPhen (Adzhubei et al., 2010), Site Directed Mutator (SDM) (Worth et al., 2011), SNPs3D (Yue et al., 2005) and PoPMuSiC (Dehouck et al., 2011).

2.3. Evolutionary conservation analysis

Consurf (Celniker et al., 2013) is a tool used to estimate the evolutionary conservation of amino acid or nucleic acid positions in their respective molecules based on phylogenetic relationships between homologous sequences. It was used to assess the conservation of amino acid residues in the GUCA2B sequence. The following parameters were set: CSI-BLAST algorithm for the homolog search, using 3 iterations with an E-value cutoff of 0.0001, against the UNIREF-90 protein database. The Venn diagram (Oliveros, 2007) was used to show the relationship between the missense SNPs classified as deleterious and the four prediction categories.

2.4. Molecular modeling

One hundred molecular models were generated for each variant using MODELLER 9.10 (Eswar et al., 2007) adopting the uroguanylin isoform A structure (PDB access code: 1UYA). The models were constructed using the default methods of auto model and environ classes from MODELLER. The best models were selected according to discrete optimized protein energy score (DOPE score), which evaluates the model's energy and indicates the best probable structure. The best models were evaluated according to ProSA II (Wiederstein and Sippl, 2007) and PROCHECK (Laskowski et al., 1993). PROCHECK is used to assess the stereochemical quality of a given protein structure, as compared with stereochemical parameters derived from well-refined high-resolution structures, while PROSA determines the protein's native fold. Structural visualization was done in PyMOL (http://www.pymol.org).

2.5. Molecular dynamics simulations

The molecular dynamics simulations of the native protein and its variants were performed in an aqueous environment using Single Point Charge water model (Berendsen et al., 1969). The analysis used the GROMOS96 43A1 force fields, and the computational package GROMACS 4 (Hess et al., 2008). The dynamics used three-dimensional models as starting structures, immersed in water, in cubic boxes with 0.8 nm minimum distances between the proteins and the limits of the box. Sodium ions were added to neutralize the system. The water molecules were constrained using the SETTLE algorithm (Miyamoto and Kollman, 1992). All atomic bonds were made using LINCS algorithm (Hess et al., 1997). Electrostatic corrections were made by Particle Mesh Ewald algorithm (Darden et al., 1993) with a 1.4 nm cut-off radius, to minimize computational time. The same radius was also used for the van der Waals interactions. Each atom neighborhood was updated every 20 steps of a 2 fs simulation. The system underwent an energy minimization using 50,000 steps of the Steepest Descent algorithm. Thereafter, the system temperature was normalized at 310 K for 100 ps using the velocity rescaling thermostat (NVT ensemble). Next, the system pressure was normalized at 1 bar for 100 ps using the Parrinello-Rahman barostat (NPT ensemble). Systems with minimized energy, temperature and balanced pressure were simulated for 1 μs using the leap-frog algorithm.

2.6. Analysis of molecular dynamics simulations

Molecular dynamics simulation trajectories 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, 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 (Hess et al., 2008). The solvation potential energy was calculated by APBS under default parameters and using the AMBER force field (Baker et al., 2001). A one-sided Student's *t*-test was applied for verifying the differences between the solvation potential energies of the wild type and variant uroguanylin structures with a critical value of 0.05.

3. Results

A total of 269 entries were retrieved from the "Variation Viewer" website when searching for the *GUCA2B* gene, being 261 SNPs, 6 deletions, 5 insertions and 1 indel. Out of the 261 SNPs found using the Single Nucleotide Variant filter, 50 (19.2%) were

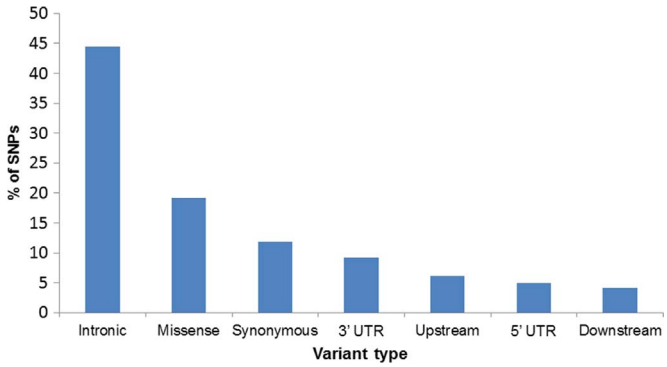

Fig. 1. Distribution of SNPs present in the *GUCA2B* gene. The distribution was based on coding regions (Missense and synonyms variants) and noncoding (upstream, downstream, 5′UTR, 3′UTR and intronic variants). From 261 SNPs, 50 (19.2%) were missense SNPs, 31 (11.9%) synonymous SNPs, 116 (44.4%) were in introns, 13 (5.0%) in 5′ UTR, 24 (9.2%) in 3′ UTR, 11 (4.2%) in downstream and 16 (6.1%) in upstream regions.

classified as missense SNPs (Fig. 1).

3.1. SNP impact prediction analysis

The 50 missense SNPs found using the "non-synonymous variant" filter were analyzed using the prediction programs divided into four different approaches (Supplementary Table 1). After this analysis, 7 out of 50 missense SNPs were considered to be convergent deleterious: Cys80Tyr, Asp98His, Cys100Gly, Ala107Glu, Cys108Tyr, Gly110Ser and Gly110Asp (Table 1).

All variations are located in the mature region, and were therefore selected for the molecular dynamics simulations. Due to the lack of a three-dimensional structure representing the whole structure of the GUCA2B protein, we were unable to assess the impact of the Cys80Tyr variant since this variant could not be analyzed by the protein sequence and structure-based methods.

The results of the Consurf analysis showed that seven convergent deleterious missense SNPs are located in highly conserved regions, with conservation values ranging between 7 and 9, which suggests that these positions are important for the peptide functioning (Fig. 2A).

For structural analysis, the isoform A of mature uroguanylin structure was selected. This structure is 16 residues long, makes a loop with itself and is stabilized by two disulfide bonds (Fig. 2B). The validation parameters for molecular modeling are summarized in Supplementary Table 2.

3.2. Molecular dynamics simulations

In the molecular dynamics simulations, the RMSD evolution showed that the native protein ranged from 2 to 5 Å, but it remained stable for most of the time. The RMSD analysis of Asp98His variant showed a greater variation in the beginning of the simulation, when compared to the native protein (Fig. 3A). For the Cys100Gly variant, the RMSD showed greater variation, ranging from 3 Å up to 6.5 Å in the first 350 ns and, after that, it was more stable than the native protein and with almost the same flexibility (Fig. 3B). The RMSD for the Ala107Glu variant was similar to the native protein (Fig. 3C). However, RMSD for the Cys108Tyr variant showed a great variation in the beginning, but after the first 100 ns it became much more stable than the native protein, although it varied greatly during the last 150 ns of the simulation (Fig. 3D). Similarly, for the Gly110Asp variant, the RMSD showed higher variation than the native protein (Fig. 3E). Regarding the Gly110Ser variant, the RMSD showed a great difference between the native protein and the variant; while the native protein ranged from 2 Å

Table 1
Prediction results of *GUCA2B* converging deleterious missense SNPs analyzed by 16 prediction tools classified in four different groups. "D" corresponds to the deleterious classification and "N" to neutral.

SNP rs #	Amino acid change	Sequence-based				SLM-based			Consensus-based			Structure-based			
		SIFT	Provean	Mutation Accessor	Panther	SNAP	PhD-SNP	SuSPect	EFIN	Meta-SNP	PON-P2	PredictSNP	CONDEL	PolyPhen	SDM	PopMusic	SNPs3D
rs746926535: G > A	Cys80Tyr	D	D	D	D	D	D	D	N	D	D	D	D	–	–	–	–
rs377375076: G > C	Asp98His	D	D	D	D	D	D	D	N	D	N	D	D	D	N	D	D
rs781213638: T > G	Cys100Gly	D	D	D	D	D	D	D	N	D	N	D	D	D	D	D	D
rs777178660: C > A	Ala107Glu	D	D	D	D	D	N	D	D	D	N	D	D	D	N	D	D
rs773317305: G > A	Cys108Tyr	D	D	D	D	D	D	D	N	D	D	D	D	D	D	D	D
rs148141677: G > A	Gly110Ser	D	D	D	D	D	N	D	D	N	D	D	D	D	D	N	D
rs531643305: G > A	Gly110Asp	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D

*The Cys80Tyr variant does not have a tridimensional structure and, although it was considered as convergent deleterious, this variant cannot be studied in further detail as there are no uroguanylin precursor solved structures covering the signal peptide.

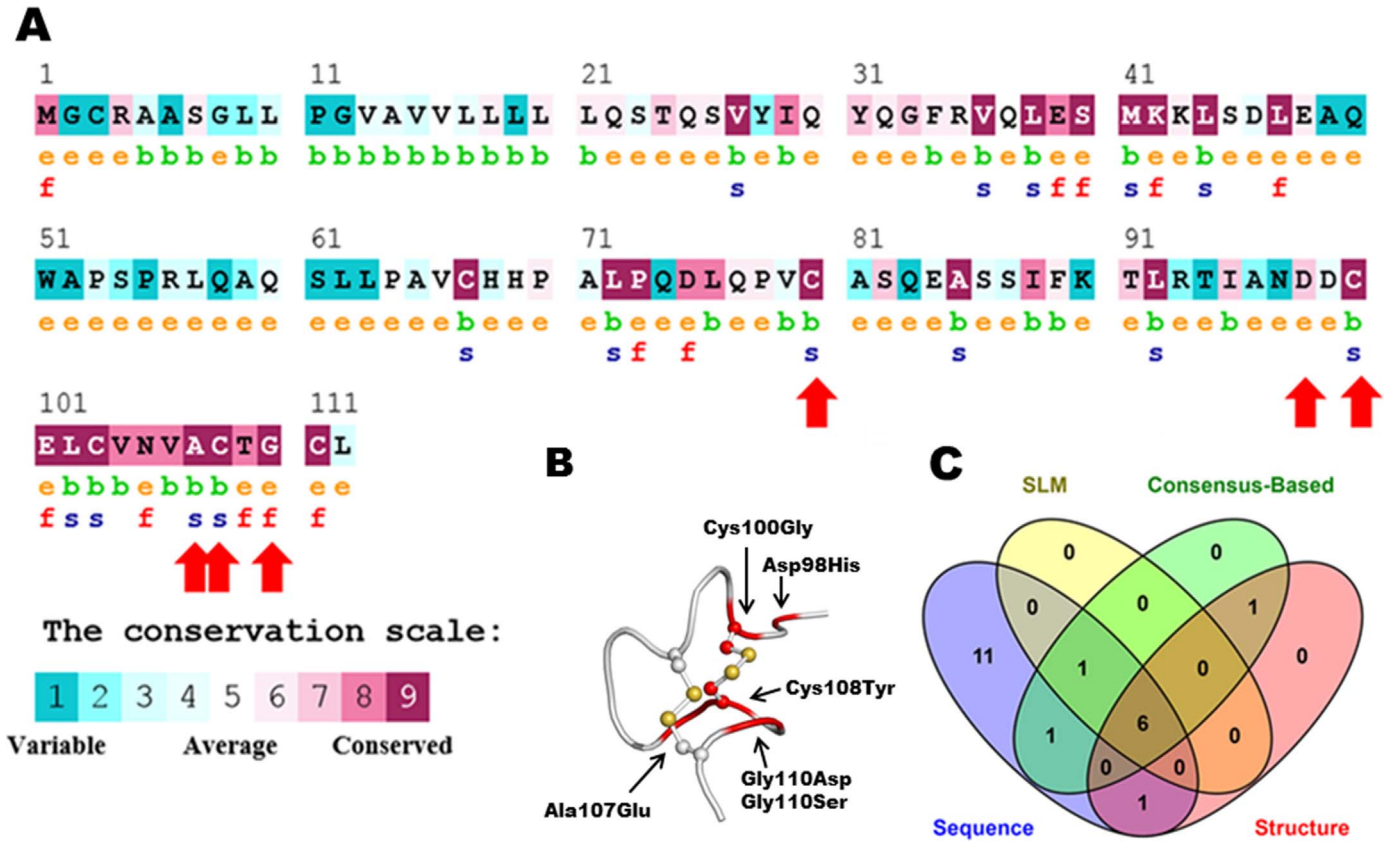

Fig. 2. Consurf profile of uroguanylin precursor (A). Value 1 indicates a high variability region. The value increases as the region becomes more conserved, up to value 9. The position values of the seven convergent deleterious missense SNPs are marked by red arrows (there are only six arrows because two missense SNPs involve mutation of the glycine at position 110). e – denotes an exposed residue according to the neural-network algorithm; b – indicates a buried residue according to the neural-network algorithm; f – indicates a predicted functional residue (highly conserved and exposed); s – indicates a predicted structural residue (highly conserved and buried); The isoform A of uroguanylin used in this study (B). Disulfide bonds are represented in ball and stick and the variants considered as convergent deleterious are pointed; The Venn diagram (C) shows the relationship between the missense SNPs considered deleterious and the four prediction program groups (Consensus-based methods, Protein sequence and structure-based methods, Supervised learning methods (SLM), and Sequence homology-based methods) used for the analysis.

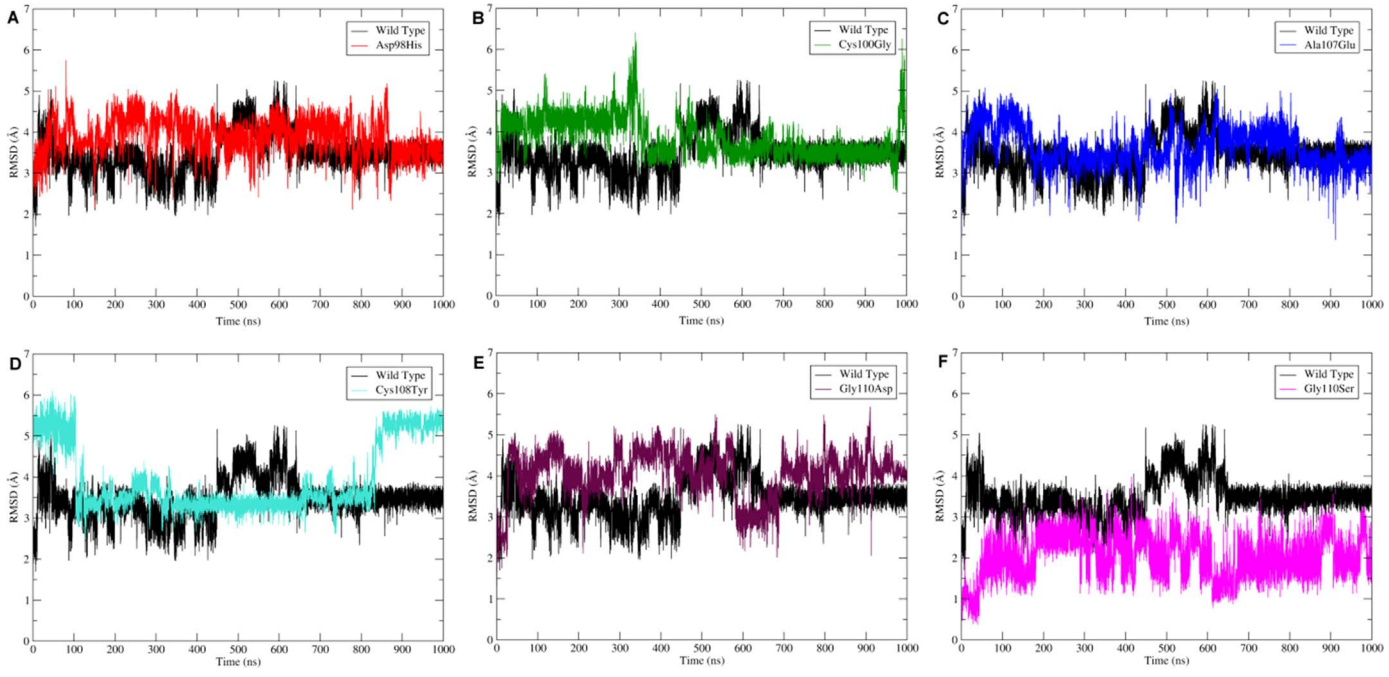

Fig. 3. The RMSD variation during the simulations. The black lines represent the native wild type and each variant as follows: Asp98His (A); Cys100Gly (B); Ala107Glu (C); Cys108Tyr (D), Gly110Asp (E) and Gly110Ser (F).

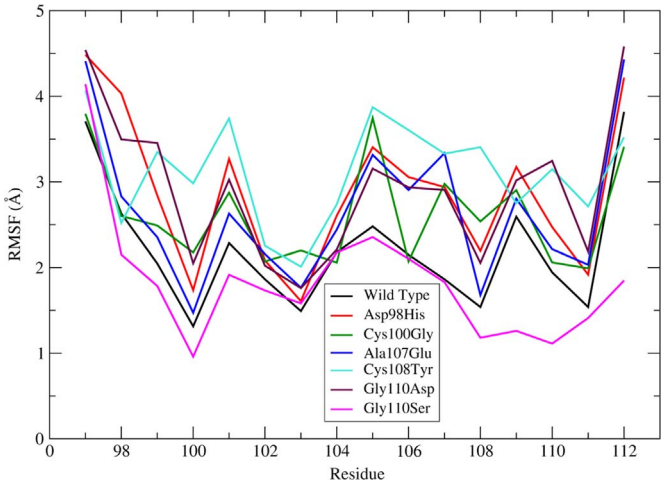

Fig. 4. The RMSF analysis for the residues of both the native wild type protein and its variants.

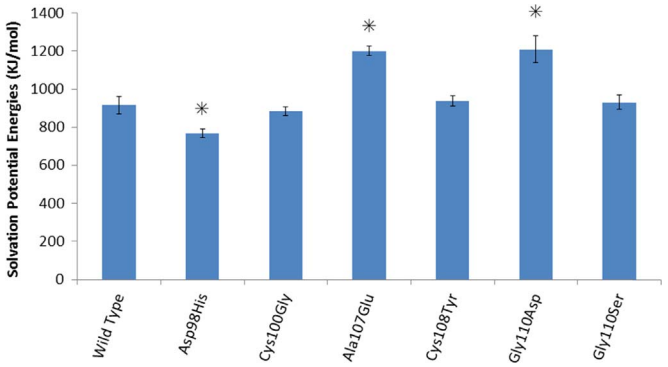

Fig. 6. Solvation Potential Energies of wild type uroguanylin and its variants. The star indicates a statistical difference of solvation potential energy between the wild type and the variant ($p < 0.05$).

to 5 Å, this variant ranged from 0.5 Å to 3 Å (Fig. 3F).

The RMSF analysis showed a variation, for the native protein, of approximately 1 Å, ranging from about 1.5 Å to about 2.5 Å, with peaks above 3.5 Å in the first and last residues. The Asp98His variant, for the RMSF analysis, showed a change of 0.5 Å above the native protein for all residues (Fig. 4). Both cysteine variants (Cys100Gly and Cys108Tyr) had greater RMS fluctuations compared to the wild type analysis, and a huge variation at the position of residue 105 (Fig. 4). RMSF of the Ala107Glu variant was similar to the native protein, although it showed a greater fluctuation at the 107 residue position (Fig. 4). RMSF of the Gly110Asp variant showed a greater fluctuation on the 110 residue position (Fig. 4). And the Gly110Ser variant RMSF varied less than the native protein, showing lower variation at the 110 residue position (Fig. 4).

The essential dynamics analysis was very similar when compared to the native protein for the Asp98His and Ala107Glu variants (Fig. 5A and C). The essential dynamics analysis showed the cysteine variants (Cys100Gly and Cys108Tyr) to be more flexible when compared to the native protein, this gain of flexibility can be explained by the disulfide bond disruption, which may result in alterations to the structure (Fig. 5B and D). Meanwhile, the essential dynamics for the Gly110Asp variant was shown the variant to be slightly more flexible than the native protein (Fig. 5E) and the Gly110Ser variant shows a loss of flexibility when compared to the native protein (Fig. 5F).

Also, the solvation potential energy analysis showed a statistically significant difference for the Asp98His, Ala107Glu and Gly110Asp variants (Fig. 6).

4. Discussion

In order to understand the behavior of proteins it is vitally important to know their three-dimensional structures. Although X-ray crystallography is a powerful tool in determining protein 3D structures, it is time-consuming and expensive, and not all proteins can be successfully crystallized. For instance, membrane

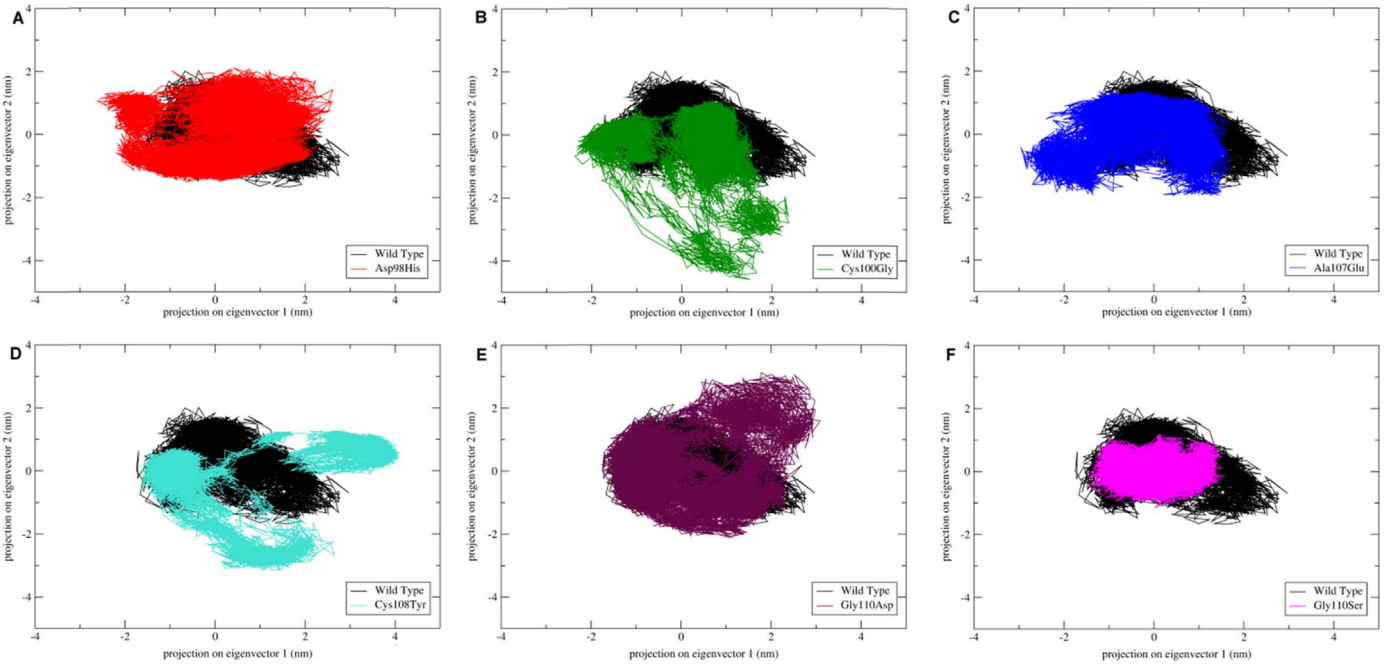

Fig. 5. Essential dynamics variation during the molecular dynamics simulations. The black lines represent the native wild type and each variant as follows: Asp98His (A); Cys100Gly (B); Ala107Glu (C); Cys108Tyr (D), Gly110Asp (E) and Gly110Ser (F). The covariance matrix is a numeric representation of the essential dynamics; higher numbers mean a more flexible structure. The values are: Wild type: 2.27188; Asp98His: 3.91503; Cys100Gly: 3.02795; Ala107Glu: 3.31169; Cys108Tyr: 4.70406; Gly110Asp: 4.05913; Gly110Ser: 1.00362.

proteins are difficult to crystallize and most of them will not dissolve in normal solvents, resulting in very few membrane protein structures currently determined. Recent studies have shown that Nuclear Magnetic Resonance can be very powerful in determining their structures (Dev et al., 2016; Fu et al., 2016; Oxenoid et al., 2016). Nevertheless, the number of protein sequences has been increasing at a much higher rate than the number of protein structures elucidated (Chou, 2004). Therefore, the development of structural bioinformatics tools was a huge advance, as it allowed the prediction of protein structural information in a feasible manner, being very useful for drug development (Wang et al., 2007; S.-Q. Wang et al., 2009b). In view of this, structural bioinformatics tools, such as sequence homology-based methods, supervised learning methods, consensus-based methods, protein sequence and structure-based methods, evolutionary conservation analysis, and molecular dynamics simulation, were also utilized in this study to analyze missense SNPs present in the uroguanylin gene in hope to provide useful information about the impact of the variations on the protein structure.

The programs were divided into three categories, each one consisting of four programs. Additionally, missense SNPs that occurred in regions present in a guanylin protein structure available in PDB also underwent analysis by four other programs based on protein sequence and structure methods. Filtering missense SNPs considered to be deleterious by at least three of the four programs in each category (convergent deleterious missense SNPs) allowed us to select a number of variations for evaluation by a more refined analysis.

Initially, the Consurf analysis showed that all convergent deleterious missense SNPs alter highly conserved amino acids; thus, all these SNPs could cause some impact to the protein (Fig. 2A). Based on this, we classified the convergent deleterious SNPs in three categories, involving (i) disulfide bridge disruption, (ii) charged residues, and (iii) glycine replacement.

In the first group, the Cys100Gly and Cys108Tyr variants may affect the protein structure, as they are involved in the formation of the first disulfide bridge and could have a similar effect made by the break of the disulfide bond in the guanylin peptide (Porto et al., 2015). Besides, residues considered to be buried by Consurf are linked to extremely conserved amino acids; hence mutations on these amino acids often destabilize the protein structure, resulting in loss of function (Armon et al., 2001).

Since the uroguanylin activity is pH dependent (Joo et al., 1998), the charged residues play a crucial role in its function; and the Asp98His, Ala107Glu and Gly110Asp variants could lead to loss of function.

In the case of glycine replacement, (Gly110Asp and Gly110Ser variants), it could cause loss of flexibility, as observed for guanylin (Porto et al., 2015). Therefore, molecular dynamics simulations were performed to investigate the possible effects of the convergent deleterious SNPs in the uroguanylin structure.

Molecular dynamics simulations have been applied to evaluate the impact of point mutations on protein structures. However, most of them use simulations ranging from a few nanoseconds to approximately 100 ns. These short duration molecular dynamics simulations are done to avoid the need of computers with high computational power in order to carry out longer simulations. If, for instance, a 1μs simulation was performed on a computer with low computing power, this simulation could last for a few months (Durrant and McCammon, 2011). The majority of the molecular dynamics simulations currently carried out are in the range of nanoseconds, as longer simulations demand a system with better processing capacity. To date, the longest duration of a simulation by molecular dynamics is in the millisecond (ms) range, but this required the development of a computer suitable for simulations, capable of generating 100 μs (μs) per day simulation time (Shaw et al., 2014). As for uroguanylin, its small size makes it a great target for longer simulations, as it does not require as much computational time as other longer proteins. During the 1 μs simulation time, we could observe the process of folding and unfolding of the peptide. It was possible to perceive many variations on the native wild type as well as for its variants, the same variations that would occur within a living cell.

Thus, taking into account the RMSD evolution in the simulations (Fig. 3), we could observe some periods of stabilization in wild type and some variants (e.g. Cys100Gly, Cys108Tyr and Gly110Ser), while other variants seem to not stabilize. In fact, we have to take into account that the N-terminal of uroguanylin is populated by negatively charged residues (Asp^{98}, Asp^{99} and Glu^{101}) and such residues constantly repel each other. This scenario is more evident in the Gly110Asp variant, where no stabilization was observed, mainly for the addition of another negatively charged residue near to the C-terminal (Fig. 2B). Regarding the radius of gyration and solvent accessible surface area, all variants showed similar results to those seen in the native protein (Supplementary Figs. 1 and 2).

We performed essential dynamics analysis in order to verify whether any variants could increase or decrease the flexibility of uroguanylin. It was observed that all variants, except the Gly110Ser variant, were more flexible than the native protein. Interestingly, the Gly110Ser and Gly110Asp variants showed opposite behaviors, mainly for their physicochemical properties. It is important to highlight that the variants involving glycine residues were expected to be less flexible. Nevertheless, for the Gly110Asp variant, it was observed more flexibility, which could be a reflection of the repulsion between the aspartic acid and the C-terminal portion of the Leu^{112}, while Gly110Ser does not have this repulsive interaction, once the serine residue is not charged. The same reasoning could explain the increase in flexibility of Asp98His and Ala107Glu variants. While Asp98His alters the net charge from -3 to -1, the His residue could repel the N-terminal portion of Gln^{97}, increasing the peptide motion; and Ala107Glu alters the charge to -4, increasing the repulsive interactions. In the case of the cysteine variants (Cys100Gly and Cys108Tyr), both disrupt the same disulfide bond and were therefore more flexible than the native protein. A similar result was observed in the guanylin peptide (Porto et al., 2015).

As three out of six convergent deleterious SNPs involve some charged residues and the uroguanylin activity is pH dependent (Joo et al., 1998), the solvation potential energy was calculated to verify whether such variation could increase or decrease this property. The variants with charge of -4 (Ala107Glu and Gly110Asp) displayed statistically significant higher values than the uroguanylin native wild type, while the variant with charge -1 (Asp98His) displayed a statistically significant lower value than the native wild type (Fig. 6). Unlike what was reported in the guanylin study (Porto et al., 2015), none of the cysteine variants present in the uroguanylin gene showed significant statistical variation.

Taking into account the solvation potential energy and the essential dynamics, the convergent deleterious SNPs involving charged residues (Asp98His, Ala107Glu and Gly110Asp) would generate the most damaging variants, since they present modifications in such properties, while the Gly110Ser variant seems to have minor effects. The cysteine variants would be more damaging than Gly110Ser and less damaging than the SNPs involving charged residues.

5. Conclusions

In this study, using missense SNP impact prediction programs and molecular dynamics simulations, we observed that some SNPs

have a harmful effect on the GUCA2B protein. For this, 16 *in silico* SNP impact prediction programs were used to investigate the protein coded by the *GUCA2B* gene. Results showed that from the 50 missense SNPs initially found, seven (Cys80Tyr, Asp98His, Cys100Gly, Ala107Glu, Cys108Tyr, Gly110Asp and Gly110Ser) were considered as convergent deleterious by the four categories. The analysis of the structural impact caused by these seven missense SNPs was then carried out by long-term molecular dynamics simulations. The variants described here can alter the flexibility of the protein and can compromise their relationship with the GC-C receptor. Among the convergent deleterious variants, four (Cys100Gly, Cys108Tyr, Gly110Asp and Gly110Ser) could cause major structural differences to the protein; they could result in a loss of protein function and may be related to the development of diseases. To date, long-term molecular dynamics simulations of 1μs duration have not yet been used to investigate the impact of missense SNPs on proteins. The 1 μs molecular dynamics simulation allowed us to evaluate the relationship of the protein with its variants for a longer time, making it possible to observe changes that could not be observed in simulations with a shorter duration.

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 Apoio a Pesquisa do Distrito Federal); and UCB (Universidade Católica de Brasília).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jtbi.2016.09.008.

References

Adcock, S. a, Mccammon, J.A., 2006. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem. Rev. 106, 1589–1615. http://dx.doi.org/10.1021/cr040426m.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249. http://dx.doi.org/10.1038/nmeth0410-248.

Armon, A., Graur, D., Ben-Tal, N., 2001. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J. Mol. Biol. 307, 447–463. http://dx.doi.org/10.1006/jmbi.2000.4474.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA 98, 10037–10041. http://dx.doi.org/10.1073/pnas.181342398.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J., Damborsky, J., 2014. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput. Biol. 10, e1003440. http://dx.doi.org/10.1371/journal.pcbi.1003440.

Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F. Van, Hermans, J., 1969. Interaction Models for Water in Relation to Protein Hydration, in: Intermolecular Forces. pp. 191–218 ⟨http://dx.doi.org/10.1146/annurev.pc.20.100169.001203⟩.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, P.E., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242. http://dx.doi.org/10.1093/nar/28.1.235.

Bromberg, Y., Yachdav, G., Rost, B., 2008. SNAP predicts effect of mutations on protein function. Bioinformatics 24, 2397–2398. http://dx.doi.org/10.1093/bioinformatics/btn435.

Capriotti, E., Altman, R.B., Bromberg, Y., 2013. Collective judgment predicts disease-associated single nucleotide variants. BMC Genom. 14 (Suppl 3), S2. http://dx.doi.org/10.1186/1471-2164-14-S3-S2.

Capriotti, E., Calabrese, R., Casadio, R., 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22, 2729–2734. http://dx.doi.org/10.1093/bioinformatics/btl423.

Carrithers, S.L., Eber, S.L., Forte, L.R., Greenberg, R.N., 2000. Increased urinary excretion of uroguanylin in patients with congestive heart failure. Am. J. Physiol. Heart Circ. Physiol. 278, H538–H547.

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., Ben-Tal, N., 2013. ConSurf: using evolutionary data to raise testable hypotheses about protein function. Isr. J. Chem. 53, 199–206. http://dx.doi.org/10.1002/ijch.201200096.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P., 2012. Predicting the functional effect of amino acid substitutions and indels. PLoS One 7, e46688. http://dx.doi.org/10.1371/journal.pone.0046688.

Chou, K., Chen, N., 1977. The biological functions of low-frequency phonons. Sci. Sin. 20, 447–457.

Chou, K.-C., 2004. Structural bioinformatics and its impact to biomedical science. Curr. Med. Chem. 11, 2105–2134.

Chou, K.-C., 1988. Low-frequency collective motion in biomacromolecules and its biological functions. Biophys. Chem. 30, 3–48. http://dx.doi.org/10.1016/0301-4622(88)85002-6.

Chou, K.C., Mao, B., 1988. Collective motion in DNA and its role in drug intercalation. Biopolymers 27, 1795–1815. http://dx.doi.org/10.1002/bip.360271109.

Chou, K.C., Zhang, C.T., Maggiora, G.M., 1994. Solitary wave dynamics as a mechanism for explaining the internal motion during microtubule growth. Biopolymers 34, 143–153. http://dx.doi.org/10.1002/bip.360340114.

Currie, M.G., Fok, K.F., Kato, J., Moore, R.J., Hamra, F.K., Duffin, K.L., Smith, C.E., 1992. Guanylin: an endogenous activator of intestinal guanylate cyclase. Proc. Natl. Acad. Sci. USA 89, 947–951.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089. http://dx.doi.org/10.1063/1.464397.

Dehouck, Y., Kwasigroch, J.M., Gilis, D., Rooman, M., 2011. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinform. 12, 151. http://dx.doi.org/10.1186/1471-2105-12-151.

Dev, J., Park, D., Fu, Q., Chen, J., Ha, H.J., Ghantous, F., Herrmann, T., Chang, W., Liu, Z., Frey, G., Seaman, M.S., Chen, B., Chou, J.J., 2016. Structural basis for membrane anchoring of HIV-1 envelope spike. Science 353, 172–175. http://dx.doi.org/10.1126/science.aaf7066.

Dorn, M., Silva, M.B.E., Buriol, L.S., Lamb, L.C., 2014. Three-dimensional protein structure prediction: methods and computational strategies. Comput. Biol. Chem. . http://dx.doi.org/10.1016/j.compbiolchem.2014.10.001

Duan, Y., Kollman, P.A., 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 282, 740–744. http://dx.doi.org/10.1126/science.282.5389.740.

Durrant, J.D., McCammon, J.A., 2011. Molecular dynamics simulations and drug discovery. BMC Biol. 9 (71). http://dx.doi.org/10.1186/1741-7007-9-71.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., Sali, A., 2007. Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein Sci. . http://dx.doi.org/10.1002/0471140864.ps0209s50

Freddolino, P.L., Liu, F., Gruebele, M., Schulten, K., 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys. J. 94, L75–L77. http://dx.doi.org/10.1529/biophysj.108.131565.

Fu, Q., Fu, T.-M., Cruz, A.C., Sengupta, P., Thomas, S.K., Wang, S., Siegel, R.M., Wu, H., Chou, J.J., 2016. Structural basis and functional role of intramembrane trimerization of the Fas/CD95 death receptor. Mol. Cell 61, 602–613. http://dx.doi.org/10.1016/j.molcel.2016.01.009.

González-Pérez, A., López-Bigas, N., 2011. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am. J. Hum. Genet. 88, 440–449. http://dx.doi.org/10.1016/j.ajhg.2011.03.004.

Gordon, G.A., 2007. Designed electromagnetic pulsed therapy: clinical applications. J. Cell. Physiol. 212, 579–582. http://dx.doi.org/10.1002/jcp.21025.

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem. 18, 1463–1472. http://dx.doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. 4, 435–447. http://dx.doi.org/10.1021/ct700301q.

Jia, M., Yang, B., Li, Z., Shen, H., Song, X., Gu, W., 2014. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2. Gene, 9. http://dx.doi.org/10.1371/journal.pone.0104311.

Joo, N.S., London, R.M., Kim, H.D., Forte, L.R., Clarke, L.L., 1998. Regulation of intestinal Cl- and HCO3-secretion by uroguanylin. Am. J. Physiol. - Gastrointest. Liver Physiol., 247.

Kinoshita, H., Fujimoto, S., Fukae, H., NaotoYokota, Hisanaga, S., Nakazato, M., Eto, T., 1999. Plasma and urine levels of uroguanylin, a new natriuretic peptide, in nephrotic. Nephron 81, 160–164.

Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O., Shaw, D.E., 2009. Long-timescale molecular dynamics simulations of protein structure and function. Curr. Opin. Struct. Biol. 19, 120–127. http://dx.doi.org/10.1016/j.sbi.2009.03.004.

Kumar, A., Purohit, R., 2014. Use of long term molecular dynamics simulation in predicting cancer associated SNPs. PLoS Comput. Biol. 10, e1003318. http://dx.doi.org/10.1371/journal.pcbi.1003318.

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081. http://dx.doi.org/10.1038/nprot.2009.86.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a

program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291. http://dx.doi.org/10.1107/S0021889892009944.
Lorenz, J.N., Nieman, M., Sabo, J., Sanford, L.P., Hawkins, J.A., Elitsur, N., Gawenis, L. R., Clarke, L.L., Cohen, M.B., 2003. Uroguanylin knockout mice have increased blood pressure and impaired natriuretic response to enteral NaCl load. J. Clin. Investig. 112, 1244–1254. http://dx.doi.org/10.1172/JCI18743.
Madkan, A., Blank, M., Elson, E., Chou, K.-C., Geddis, S., Goodman, R, M., 2009. Steps to the clinic with ELF EMF. Nat. Sci. 01, 157–165. http://dx.doi.org/10.4236/ns.2009.13020.
Martel, P., 1992. Biophysical aspects of neutron scattering from vibrational modes of proteins. Prog. Biophys. Mol. Biol. 57, 129–179. http://dx.doi.org/10.1016/0079-6107(92)90023-Y.
Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., Kitano, H., Thomas, P.D., 2005. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res. 33, D284–D288. http://dx.doi.org/10.1093/nar/gki078.
Miller, M.P., Kumar, S., 2001. Understanding human disease mutations through the use of interspecific genetic variation. Hum. Mol. Genet. 10, 2319–2328.
Miyamoto, S., Kollman, P.A., 1992. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. 13, 952–962. http://dx.doi.org/10.1002/jcc.540130805.
Miyazato, M., Nakazato, M., Matsukura, S., Kangawa, K., Matsuo, H., 1997. Genomic structure and chromosomal localization of human uroguanylin. Genomics 43, 359–365. http://dx.doi.org/10.1006/geno.1997.4808.
Miyazato, M., Nakazato, M., Yamaguchi, H., Date, Y., Kojima, M., Kangawa, K., Matsuo, H., Matsukura, S., 1996. Cloning and characterization of a cDNA encoding a precursor for human uroguanylin. Biochem. Biophys. Res. Commun. 219, 644–648. http://dx.doi.org/10.1006/bbrc.1996.0287.
Nakazato, M., Yamaguchi, H., Date, Y., Miyazato, M., Kangawa, K., Goy, M.F., Chino, N., Matsukura, S., 1998. Tissue distribution, cellular source, and structural analysis of rat immunoreactive uroguanylin. Endocrinology 139, 5247–5254.
Nakazato, M., Yamaguchi, H., Kinoshita, H., Kangawa, K., Matsuo, H., Chino, N., Matsukura, S., 1996. Identification of biologically active and inactive human uroguanylins in plasma and urine and their increases in renal insufficiency. Biochem Biophys. Res. Commun. 220, 586–593. http://dx.doi.org/10.1006/bbrc.1996.0447.
Ng, P.C., Henikoff, S., 2006. Predicting the effects of amino acid substitutions on protein function. Annu. Rev. Genom. Hum. Genet. 7, 61–80. http://dx.doi.org/10.1146/annurev.genom.7.080505.115630.
Niroula, A., Urolagin, S., Vihinen, M., 2015. PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One 10, e0117380. http://dx.doi.org/10.1371/journal.pone.0117380.
Oliveros, J.C., 2007. VENNY. An interactive tool for comparing lists with Venn Diagrams. [WWW Document]. BioinfoGP of CNB-CSIC.
Oxenoid, K., Dong, Y., Cao, C., Cui, T., Sancak, Y., Markhard, A.L., Grabarek, Z., Kong, L., Liu, Z., Ouyang, B., Cong, Y., Mootha, V.K., Chou, J.J., 2016. Architecture of the mitochondrial calcium uniporter. Nature 533, 269–273. http://dx.doi.org/10.1038/nature17656.
Porto, W.F., Franco, O.L., Alencar, S.A., 2015. Computational analyses and prediction of guanylin deleterious SNPs. Peptides 69, 92–102. http://dx.doi.org/10.1016/j.peptides.2015.04.013.
Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 39, e118. http://dx.doi.org/10.1093/nar/gkr407.
Shaw, D.E., Grossman, J.P., Bank, J.A., Batson, B., Butts, J.A., Chao, J.C., Deneroff, M.M., Dror, R.O., Even, A., Fenton, C.H., Forte, A., Gagliardo, J., Gill, G., Greskamp, B., Ho, C.R., Ierardi, D.J., Iserovich, L., Kuskin, J.S., Larson, R.H., Layman, T., Lee, L.-S., Lerer, A.K., Li, C., Killebrew, D., Mackenzie, K.M., Mok, S.Y.-H., Moraes, M.A., Mueller, R., Nociolo, L.J., Peticolas, J.L., Quan, T., Ramot, D., Salmon, J.K., Scarpazza, D.P., Schafer, U. Ben, Siddique, N., Snyder, C.W., Spengler, J., Tang, P.T.P., Theobald, M., Toma, H., Towles, B., Vitale, B., Wang, S.C., Young, C., 2014. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer, in: SC14: International Conference for High Performance Computing, Networking, Storage and Analysis. pp. 41–53. ⟨http://dx.doi.org/10.1109/SC.2014.9⟩.
Shen, J., Deininger, P.L., Zhao, H., 2006. Applications of computational algorithm tools to identify functional SNPs in cytokine genes. Cytokine 35, 62–66. http://dx.doi.org/10.1016/j.cyto.2006.07.008.
Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311.
Sindic, A., 2006. Cellular effects of guanylin and uroguanylin. J. Am. Soc. Nephrol. 17, 607–616. http://dx.doi.org/10.1681/ASN.2005080818.
Wang, J.-F., Gong, K., Wei, D.-Q., Li, Y.-X., Chou, K.-C., 2009a. Molecular dynamics studies on the interactions of PTP1B with inhibitors: from the first phosphate-binding site to the second one. Protein Eng. Des. Sel. 22, 349–355. http://dx.doi.org/10.1093/protein/gzp012.
Wang, J.-F., Wei, D.-Q., Li, L., Zheng, S.-Y., Li, Y.-X., Chou, K.-C., 2007. 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design. Biochem. Biophys. Res. Commun. 355, 513–519. http://dx.doi.org/10.1016/j.bbrc.2007.01.185.
Wang, S.-Q., Du, Q.-S., Huang, R.-B., Zhang, D.-W., Chou, K.-C., 2009b. Insights from investigating the interaction of oseltamivir (Tamiflu) with neuraminidase of the 2009 H1N1 swine flu virus. Biochem. Biophys. Res. Commun. 386, 432–436. http://dx.doi.org/10.1016/j.bbrc.2009.06.016.
Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407–W410. http://dx.doi.org/10.1093/nar/gkm290.
Worth, C.L., Preissner, R., Blundell, T.L., 2011. SDM–a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res. 39, W215–W222. http://dx.doi.org/10.1093/nar/gkr363.
Yang, Z., Kurpiewski, M.R., Ji, M., Townsend, J.E., Mehta, P., Jen-Jacobson, L., Saxena, S., 2012. ESR spectroscopy identifies inhibitory Cu2+ sites in a DNA-modifying enzyme to reveal determinants of catalytic specificity. Proc. Natl. Acad. Sci. USA 109. http://dx.doi.org/10.1073/pnas.1200733109.
Yates, C.M., Filippis, I., Kelley, L.A., Sternberg, M.J.E., 2014. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. J. Mol. Biol. 426, 2692–2701. http://dx.doi.org/10.1016/j.jmb.2014.04.026.
Yates, C.M., Sternberg, M.J.E., 2013. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. J. Mol. Biol. 425, 3949–3963. http://dx.doi.org/10.1016/j.jmb.2013.07.012.
Yue, P., Li, Z., Moult, J., 2005. Loss of protein structure stability as a major causative factor in monogenic disease. J. Mol. Biol. 353, 459–473. http://dx.doi.org/10.1016/j.jmb.2005.08.020.
Zeng, S., Yang, J., Chung, B.H.-Y., Lau, Y.L., Yang, W., 2014. EFIN: predicting the functional impact of nonsynonymous single nucleotide polymorphisms in human genome. BMC Genom. 15, 455. http://dx.doi.org/10.1186/1471-2164-15-455.
Zhao, N., Han, J.G., Shyu, C.-R., Korkin, D., 2014. Determining effects of non-synonymous SNPs on protein-protein interactions using supervised and semi-supervised learning. PLoS Comput. Biol. 10, e1003592. http://dx.doi.org/10.1371/journal.pcbi.1003592.
Zhou G.-P., Biological functions of soliton and extra electron motion in DNA structure, Phys. Scr. 40, 1989, 698–701, http://iopscience.iop.org/article/10.1088/0031-8949/40/5/021/meta.