Journal of Theoretical Biology 419 (2017) 193–200

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Theoretical structural characterization of lymphoguanylin: A potential candidate for the development of drugs to treat gastrointestinal disorders

Állan S. Pires[a,b,1], William F. Porto[a,b,d,1], Pryscilla O. Castro[a], 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
[d] Porto Reports, Brasília-DF, Brazil

ARTICLE INFO

Keywords:
Disordered protein
Guanylin peptides
Core flexibility
Molecular dynamics

ABSTRACT

Guanylin peptides (GPs) are small cysteine-rich peptide hormones involved in salt absorption, regulation of fluids and electrolyte homeostasis. This family presents four members: guanylin (GN), uroguanylin (UGN), lymphoguanylin (LGN) and renoguanylin (RGN). GPs have been used as templates for the development of drugs for the treatment of gastrointestinal disorders. Currently, LGN is the only GP with only one disulfide bridge, making it a remarkable member of this family and a potential drug template; however, there is no structural information about this peptide. In fact, LGN is predicted to be highly disordered and flexible, making it difficult to obtain structural information using *in vitro* methods. Therefore, this study applied a series of 1 μs molecular dynamics simulations in order to understand the structural behavior of LGN, comparing it to the C115Y variant of GN, which shows the same Cys to Tyr modification. LGN showed to be more flexible than GN C115Y. While the negatively charged N-terminal, despite its repellent behavior, seems to be involved mainly in pH-dependent activity, the hydrophobic core showed to be the determinant factor in LGN's flexibility, which could be essential in its activity. These findings may be determinant in the development of new medicines to help in the treatment of gastrointestinal disorders. Moreover, our investigation of LGN structure clarified some issues in the structure-activity relationship of this peptide, providing new knowledge of guanylin peptides and clarifying the differences between GN C115Y and LGN.

1. Introduction

Guanylin peptides (GPs) are small cysteine-rich peptide hormones involved in salt absorption, regulation of fluids and electrolyte homeostasis (Carvalho et al., 2008). They are commonly associated with the *Escherichia coli* heat-stable peptide (STa), a toxin which can cause diarrhea and is associated with gut infection (Currie et al., 1992). STa and GPs are involved in the same signaling pathway, activating guanylate cyclase receptor C, which in turn activates cyclic guanosine monophosphate (cGMP) as a second messenger. This activation leads to the inhibition of water and sodium absorption and to the induction of chloride and potassium secretion (Sindic, 2013; Sindic and Schlatter, 2006).

Until now, the GP family has four members, namely guanylin (GN) (de Sauvage et al., 1992), uroguanylin (UGN) (Hamra et al., 1993), lymphoguanylin (LGN) (Forte et al., 1999) and renoguanylin (RGN) (Yuge et al., 2003). Only GN and UGN have been found in humans so far; while LGN and RGN were found in the American opossum (*Didelphis virginiana*) and European eel (*Anguilla japonica*), respectively. GPs can be produced by several tissues such as gastrointestinal and reproductive systems, adrenal glands, lung and pancreas (Nakazato et al., 1998). They are also produced as preprohormones that are cleaved to release the mature active peptide (Hill et al., 1995). Mature GPs consist of 15 or 16 amino acid residues, stabilized by two disulfide bridges (Fig. 1). LGN is the exception, being stabilized by a unique disulfide bridge (between Cys^{98} and Cys^{106}), since there is a tyrosine located at position 109 in the C-Terminus of this peptide, where all GPs identified to date have a cysteine residue at this position (Lima and Fonteles, 2014). This substitution hinders the formation of the second disulfide bridge (Fig. 1), making LGN a remarkable member

* Corresponding author.
E-mail address: sergiodealencar@gmail.com (S.A. Alencar).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jtbi.2017.02.016
Received 18 November 2016; Received in revised form 18 January 2017; Accepted 13 February 2017
Available online 16 February 2017

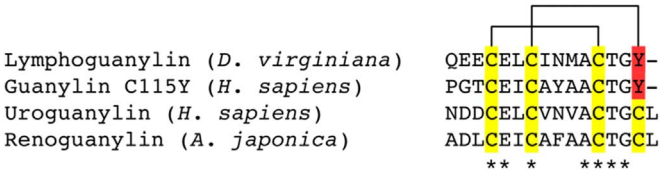

Fig. 1. Multiple sequence alignment of the four types of guanylin mature peptides. Conserved residues are highlighted with an asterisk; the cysteine residues are highlighted in yellow and the disulfide bonds are indicated above the alignment. LGN and GN C115Y C-Terminal Tyr residue is highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of this peptide family (Fonteles et al., 2001; Forte et al., 1999). Recently, Porto and colleagues (2015) identified a GN variant driven by a missense single nucleotide polymorphism (rs139468524:C > T), with the same modification of cysteine to tyrosine at position 115, similar to the sequence of LGN (Fig. 1). Molecular dynamics simulations of this GN variant showed that it caused a decrease in GN flexibility. As a result, it was predicted to have a deleterious effect on the protein function (Porto et al., 2015). Despite that, in contrast to this guanylin variant, LGN has been described as a bioactive GP (Forte et al., 1999).

Lymphoguanylin could be widely expressed in many lymphoid tissues of opossums (Forte et al., 1999). However, the great abundance of mRNA coding this peptide in kidney and heart indicates a paracrine or autocrine intrarenal cGMP signaling mechanism of LGN, influencing Na^+ and Cl^- transport in kidneys (Fonteles et al., 2001). However, LGN is less effective in opossum kidney and T84 intestinal cells compared to UGN and GN (Forte, 1999). The lower activity of LGN has been related to the absence of a second disulfide bond, despite the sequence identity to other GPs, as depicted in Fig. 1 (Forte et al., 1999). However, by a single amino acid modification (Cys7Ser), LGN could present a similar activity to UGN, capable of generating natriuretic responses in an isolated perfused rat kidney, highlighting LGN as a potential template for drug development (Fonteles et al., 2001).

Indeed, GPs have already been used as drug templates. Up to now, three drugs derived from GPs or guanylin-like peptides have been approved or are in the approval process at the US Food and Drug administration (FDA): Linaclotide (Ironwood Pharmaceuticals Inc, Boston, MA, USA and Forest Laboratories Inc, New York, NY, USA) is derived from *E. coli* STa, while Plecanatide and Dolcanatide (Synergy Pharmaceuticals Inc, New York, NY, USA) are both analogs of UGN used to treat gastrointestinal disorders, despite UGN renal function. These drugs have been described as excellent options for the treatment of gastrointestinal disorders such as irritable bowel syndrome with constipation (Busby et al., 2010; Camilleri, 2015; Shailubhai et al., 2013), chronic idiopathic constipation (Camilleri, 2015; Shailubhai et al., 2015) and ulcerative colitis (Shailubhai et al., 2015, 2013). For this reason, analogs of GPs could be considered as promising drugs for controlling gastrointestinal disorders. Despite the equivalence of LGN Cys7Ser and UGN (Fonteles et al., 2001), LGN has been poorly explored as a drug template. In addition, currently there is no structural information available for this peptide.

To understand the biomacromolecules' mechanisms of action, it is crucial to consider the dynamical information acquired by studying their internal motions in addition to the static structural information. The reliability of analyses of biomacromolecules from a dynamic point of view has been demonstrated by studies where the low-frequency vibrations were revealed to possess significant biological functions (Chou, 1988; Chou et al., 1989; Zhou, 1989). Moreover, proteins internal motion knowledge is the key to the understanding of biomacromolecules' mechanisms of action (Chou, 1988, 2004). Therefore, it is crucial also to consider the impact of internal motions on structure, which could be assessed by dynamics simulations.

Molecular dynamics simulations have been widely used to obtain structural information about proteins, as well as kinetic and thermodynamic information (Adcock and Mccammon, 2006; Jia et al., 2014; Kumar and Purohit, 2014; Yang et al., 2012). Besides, long-term molecular dynamics simulations can provide more detailed information about protein conformation, allowing the visualization of fold and unfold processes in a time scale inaccessible by *in vitro* methods (Adcock and Mccammon, 2006; Klepeis et al., 2009).

Since the structural knowledge of LGN is crucial for the design of new drugs, molecular modelling and dynamics could be used to shed some light on its structure and behavior. In addition, UGN (Marcolino et al., 2016) and GN (Porto et al., 2015) were recently studied by molecular dynamics simulations, demonstrating the suitability of this technique for peptide analysis. Therefore, we performed an *in silico* structural analysis of LGN, comparing it with the Cys115Tyr GN variant, in order to understand the structural behavior of LGN.

2. Material and methods

2.1. Evolutionary conservation analysis

The ConSurf server is a tool for estimating the evolutionary conservation of amino acid positions in a protein molecule based on the phylogenetic relations between homologous sequences (Celniker et al., 2013). Using the preprolymphoguanylin protein sequence, ConSurf, in ConSeq mode, carried out a search for close homologous sequences using CSI-BLAST (3 iterations and 0.0001 of e-value cutoff) against the UNIREF-90 protein database (Angermüller et al., 2012; Suzek et al., 2007). The maximum number of homologs to collect was set as 150, and the minimal and maximal percentages for ID between sequences were set as 35 and 95, respectively. The multiple sequence alignment and calculation methods were set as default (MAFFT-L-INS-i and Bayesian). The sequences were then clustered and highly similar sequences removed using CD-HIT (Li and Godzik, 2006). Position-specific conservation scores were computed using the empirical Bayesian algorithm (Mayrose et al., 2004).

2.2. Physicochemical sequence analysis

The sequences of LGN and GN C115Y were analyzed in relation to their disordered regions by means of PrDOS (Ishida and Kinoshita, 2007) using default parameters. PrDOS is a server that uses amino acid sequences to predict natively disordered regions in proteins (Ishida and Kinoshita, 2007). Two steps of prediction were performed, being (i) sequence-analyzed residue by residue by the Support Vector Machine algorithm; and (ii) a PSI-BLAST carried out to compare disordered regions with intrinsic disorder in protein families. PrDOS returns a disorder probability of each residue as a combination of two prediction results. In addition, we analyzed LGN and GN C115Y for their instability index and hydropathicity by ProtParam (Gasteiger et al., 2005). The instability index estimates the stability of proteins in a test tube, classifying it into stable and unstable (Guruprasad et al., 1990). Proteins with an instability index lower than 40 are predicted to be stable, while those higher than 40 are predicted to be unstable. The hydropathicity calculates the sum of hydropathy values by Kyte and Doolittle scale (Kyte and Doolittle, 1982). Besides, the propensity to aggregate was predicted by AGGRESCAN (Conchillo-Solé et al., 2007). Its algorithm is based on protein aggregation modulation by short and specific stretches, and also on an *in vivo* derived scale of aggregation-propensity for natural amino acids.

2.3. Molecular modelling

One hundred molecular models were generated for each peptide by means of MODELLER 9.16 (Fiser and Šali, 2003) using the human proguanylin structure as a template (PDB ID: 1O8R) (Lauber et al., 2003), which was indicated as the best template by HHPred server (Söding et al., 2005). The models were constructed using the default

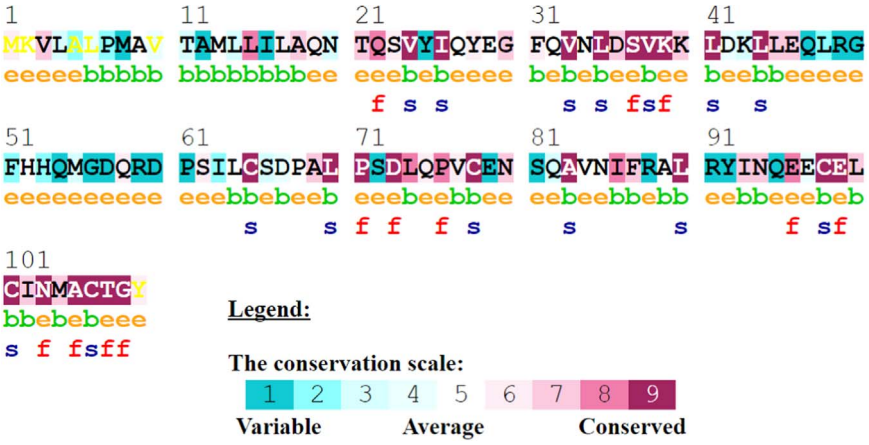

Fig. 2. Evolutionary conservation of LGN amino acid residues obtained from multiple sequence alignment by ConSurf. The amino acids' colors are based on their conservation grades and conservation levels. A grade of 1 indicates rapidly evolving (variable) sites, a grade of 5 indicates sites that are evolving at an average rate, and 9 indicates slowly evolving (evolutionarily conserved) sites. The colors are explained in the figure. e – represents 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).

methods of auto model and environ classes from MODELLER. The best models were selected according to DOPE (Discrete Optimized Protein Energy) score, which evaluates the models' energies and indicates the most probable structure. The best model was evaluated through PROSA II (Wiederstein and Sippl, 2007) and PROCHECK (Laskowski et al., 1993). PROCHECK was used in order to check the stereochemical quality of protein structure through the Ramachandran plot, where good quality models are expected to have more than 90% of amino acid residues in the most favored and additional allowed regions. PROSA II indicates the fold quality by means of the Z-score. Structural visualization was done in PyMOL (http://www.pymol.org).

2.4. Electrostatic surface calculation

The solvation potential energy was calculated by Adaptive Poisson-Boltzmann Solver (APBS) under default parameters (Baker et al., 2001). APBS evaluates electrostatic interactions in biomolecular assemblages by continuum solvation methods. The utility PDB2PQR using the AMBER force field (Dolinsky et al., 2004) was used for the conversion of pdb files into pqr files. The grid dimensions for APBS calculation were also determined by PDB2PQR. Electrostatic surface visualization was done in PyMOL, by means of the APBS plugin.

2.5. Molecular dynamics simulations

The molecular dynamics simulations of complexes were performed in GROMOS96 43A1 force field (Hess et al., 2008) using the GROMACS 4 package (Hess et al., 2008). To obtain neutralized charged Lymphoguanylin (LGN^0), all charged residues were neutralized by pdb2gmx function of the GROMACS package (Hess et al., 2008). Structures were simulated in a water environment using the Single Point Charge water model (Berendsen et al., 1981). The dynamics were done 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 systems with negative charge. The water molecule geometry was constrained by 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 1.4 nm threshold to minimize computational time. The same cut-off was 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. System temperature was normalized to 310 K for 100 ps, using the velocity rescaling thermostat (NVT ensemble), and system pressure was normalized to 1 bar for 100 ps, using the Parrinello-Rahman barostat (NPT ensemble). The complete systems' simulations were done for 1 μs using the leap-frog algorithm as the integrator.

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 (SASA) using the g_rms, g_rmsf, g_gyrate and g_sas built in functions of the GROMACS package (Hess et al., 2008), respectively. The essential dynamics was performed using the g_covar and g_anaeig utilities of the GROMACS package (Hess et al., 2008) to analyze and visualize motions of peptides during simulations. The covariance matrices of structures were constructed using the main chain atoms. Secondary structural conservation was evaluated by DSSP 2.0.4 (Kabsch and Sander, 1983; Touw et al., 2015).

3. Results

3.1. The majority of mature LGN residues are conserved and functional

Conservation analyses showed that most residues from the mature sequence (residues 95–109) have high conservation levels, with grades from 7 to 9 (Fig. 2). These values indicate that mature LGN presents an average to slow evolution rate. However, residues Gln^{95} and Glu^{97} showed a variable grade, being the only variable sites in the mature sequence. Residue Tyr^{109} presents an average grade, mainly because in this position other GPs present a cysteine (Fig. 2). Furthermore, ConSurf predicted six amino acids of the mature peptide (Glu^{96}, Glu^{99}, Asn^{103}, Ala^{105}, Thr^{107} and Gly^{108}) as functional, as well as extremely conserved. Thus, except for the N-terminal, LGN residues showed to be conserved, and the majority of them play a functional role in the peptide.

3.2. LGN presents disordered terminals

The sequences of LGN and GN C115Y were analyzed by PrDOS to verify whether peptides present disordered regions. The analysis showed that in both sequences, N-terminal (^{1}QEECE^5 and ^{1}PGTCE^5 to LGN and GN C115Y, respectively) and C-terminal (^{10}MACTGY^{15} and ^{10}AACTGY^{15} to LGN and GN C115Y, respectively) are disordered. Moreover, both peptides were considered unstable by ProtParam, with instability indices of 72.01 and 121.13, respectively, for GN C115Y and

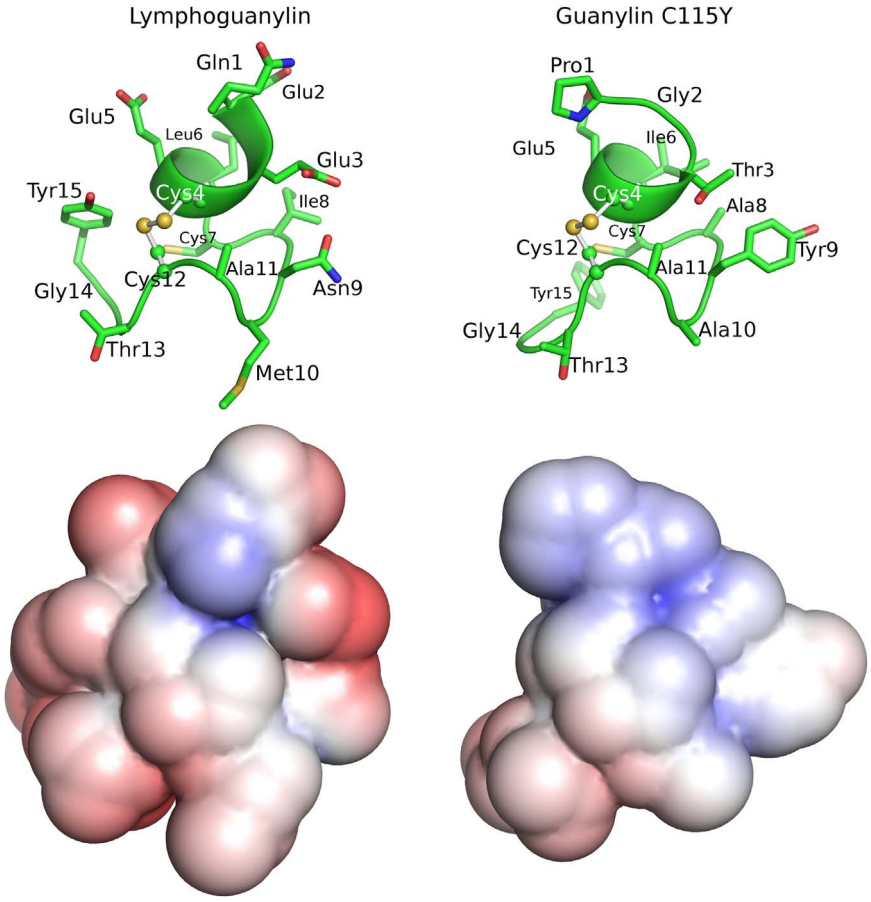

Fig. 3. Structures and electrostatic surfaces of Lymphoguanylin and Guanylin C115Y. (Top) The molecular models of LGN (left) and GN C115Y (right). Disulfide bridges are shown in ball and stick. Both structures present similar folding. The residues are labeled according to the mature peptide (Bottom) Electrostatic surface representation of the structure models, shown in the same position as the cartoon structures. LGN (left) possesses an anionic surface, with an acidic N-terminal, while C115Y GN (right) presents a more neutral surface. Surface potentials were set to ± 5 KTe^{-1} (133.56 mV). The electrostatic surface values are 9.283 E+2 and 4.921 E+2 KJ/mol for LGN and GN C115Y, respectively.

LGN; thus LGN is more unstable than GN C115Y. LGN also presented higher hydropathicity than GN C115Y (−0.027 and 0.5, respectively). These findings showed that LGN is more unstable and hydrophilic than GN C115Y. In addition, AGGRESCAN indicated that both peptides could form aggregates, where the aggregation area for LGN is 2.86, while for GN C115Y is 2.931.

3.3. LGN and GN C115Y show similar structures but have differences in molecular surface

The molecular models of LGN and GN C115Y were generated by MODELLER 9.16 to evaluate and compare LGN structural behavior. All peptides showed a similar structure, with a small α-helix in the N-terminal region (^{2}EECEL^{5} and ^{3}TCEI^{6} to LGN and GN C115Y, respectively) and the remaining residues in a loop structure (Fig. 3, top). The molecular modelling assessments are summarized in Table 1.

Due to the very similar secondary structure, we performed APBS calculations in order to verify the differences between peptides in terms of electrostatic surface. This showed that LGN is more acidic than GN C115Y, which, in turn, showed an amphoteric surface (Fig. 3, bottom).

3.4. LGN is more flexible than GN C115Y

In order to compare the motions of the structure, we performed molecular dynamics simulations of all peptides. The RMSD showed that LGN ranged from 3 to 6 Å during the simulation, while GN C115Y ranged from 4 to 5.5 Å (Fig. 4, top left). Therefore, LGN is more unstable for most of the time when compared with GN C115Y. Regarding RMSF, a deviation was observed for each residue during the simulation. Overall, the LGN residues presented more moves than GN C115Y, with fluctuations higher than 3 Å, which reflects the RMSD variation (Fig. 4, top right). Although PrDOS showed that the terminals tended to be disordered, the RMSF showed that this only occurred for LGN (Fig. 4, right top), indicating that there is some stabilization of the GN C115Y structure.

Due to this tendency of stabilization of GN C115Y, we analyzed the

Table 1
Summary of structural validation parameters.

Peptide	DOPE Score	Z-score (ProSA II)	Ramachandran Plot (%)		Supplementary File
			Most favored regions	Additional allowed regions	
LGN	−632.4	1.49	84.6	13.6	File S1
GN Cys115Tyr	−655.1	0.66	100	0.0	File S2
Chimera 1	−718.0	0.33	100	0.0	File S3
Chimera 2	−584.9	1.09	100	0.0	File S4

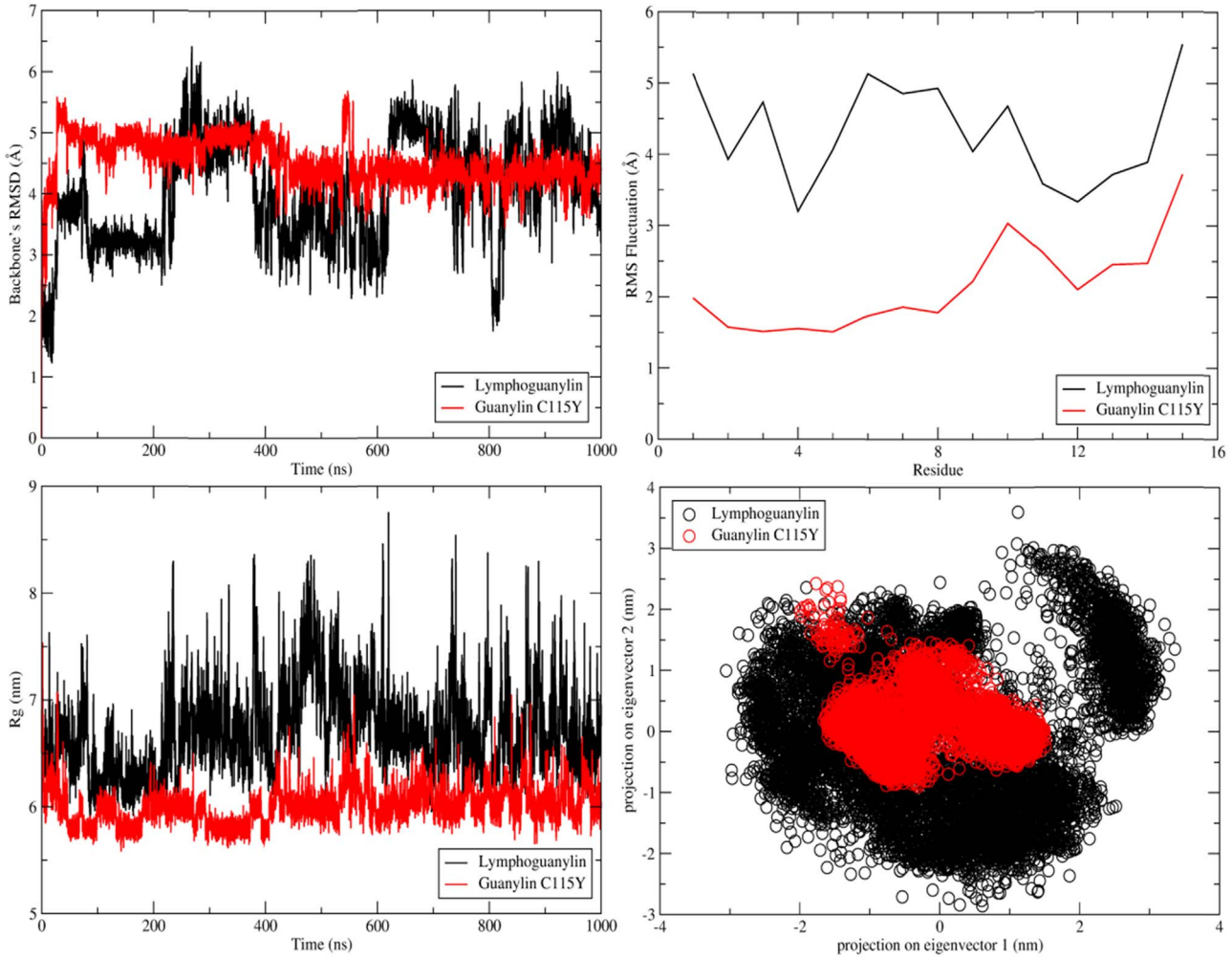

Fig. 4. Overall analyses of simulation trajectories. Results of molecular dynamics analysis of LGN (black) and GN C115Y (red) structures. Backbone RMSD variation during the simulations (Top left). RMS fluctuation of the residues (Top right). Radius of gyration (Bottom left). Projection of the motion of guanylin in phase space along the first two principal eigenvectors (Bottom right). The values of covariance matrix are: 7.647 for lymphoguanylin and 1.583 for guanylin C115Y.

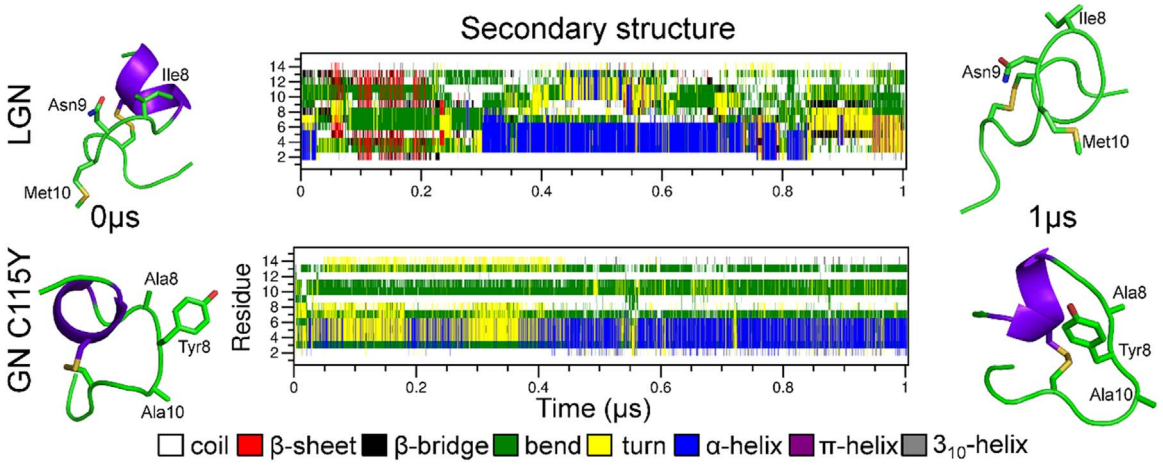

Fig. 5. DSSP analyses and structural movement of core residues after simulation. LGN presents a more unstable secondary structure than GN C115Y, with more variations in gain and loss of secondary structures during the simulation. Residues of LGN and GN C115Y core present a different pattern of motion, with Tyr[8] generating a more stable structure in GN C115Y than [8]INM[10] in LGN, which stays in motion.

radius of gyration to verify whether there are differences in structure compactness. It was observed that GN C115Y is more compact than LGN (Fig. 4, bottom left). LGN, in turn, has presented an expansion and compression behavior with periodic spikes in the radius of gyration (Fig. 4, bottom left). Following this line, the essential dynamics was done to assess the flexibility differences between LGN and GN C115Y. This analysis showed LGN as more flexible than GN C115Y, with covariance matrix values of 7.6472 and 1.5829 for LGN to GN C115Y, respectively (Fig. 4, bottom right). That behavior supports the higher LGN instability index. Corroborating the analyses, DSSP showed that

the secondary structure of LGN is more unstable, with many gains and losses of structures over time (Fig. 5). Moreover, core residues present a different pattern of motion with no stabilization of residues ^{8}INM^{10} in LGN, despite the more stable folding in ^{8}AYA^{10} in GN C115Y (Fig. 5).

3.5. The LGN hydrophobic core determines the structure flexibility

The N-terminal and the hydrophobic core are the main differences between LGN and GN C115Y. While the N-terminal of LGN has a predominance of acidic residues (^{1}QEE^{3}), GN C115Y has hydrophobic ones (^{1}PGT^{3}) (Fig. 1); and while the hydrophobic core of LGN is composed of ^{6}LCINM^{10}, GN C115Y's is composed of ^{6}ICAYA^{10} (Fig. 1). Since such regions could explain the difference in the moves observed in the simulations, we performed three additional simulations in order to verify the contribution of the acidic N-terminal and the hydrophobic core for the peptide flexibility. Two chimeric peptides were simulated, exchanging the N-terminals of GN C115Y and LGN: chimera 1 (QEECEICAYAACTGY) and chimera 2 (PGTCELCINMACTGY). Besides, we also simulated the LGN structure with all charged residues neutralized (LGN^{0}).

These simulations showed that despite the more constant RMSD variation of LGN^{0} compared to LGN, with no great deviation after 500 ns, the flexibility increased when compared to LGN (Fig. 6, left). The exchange of N-terminal sequences of peptides generated a slight increase in flexibility compared to native peptides; however, there was more variation in RMSD for the chimera 1 structure (Fig. 6, right).

Together with the simulation of LGN^{0}, chimera simulations showed that charged residues are responsible for most inconstancy in RMSD (Fig. 6, left). This occurs because negative charges cause electrostatic repulsion of the residues. Therefore, the exchange of these three residues (QEE) of LGN as well as neutralization of charges generated a more constant RMSD (Fig. 6, left).

However, despite the importance of N-terminal negatively charged residues in RMSD, these findings highlighted the importance of the hydrophobic core in the flexibility of LGN, being a determining factor of this feature. According to that, LGN^{0} and chimera 2, which have the same hydrophobic core (^{6}LCINM^{10}), presented higher flexibility than chimera 1, which has the GN C115Y core (^{6}ICAYA^{10}). Thus, the absence of charge in LGN^{0} leads to an increase in flexibility, enhancing the covariance matrix from 7.647 to 9.689 (Fig. 4, right bottom and Fig. 6, right). Moreover, the exchange of N-terminal sequences leads to a similar scenario. Such behavior indicates that, despite the RMSD pattern, the flexibility observed for LGN is mostly determined by hydrophobic core unlike the charged terminal (Fig. 4, bottom left).

4. Discussion

LGN is an active guanylin peptide isolated from opossum tissues. In contrast to other GPs, LGN presents a substitution of the terminal cysteine, generating a peptide with only one disulfide bridge, which in turn makes this peptide a singular member of the family (Fonteles et al., 2001; Forte et al., 1999). GPs and even the toxin STa have been used as templates for drug development, where Plecanatide and Linaclotide were developed based on UGN and STa, respectively (Camilleri, 2015). These medicines have been used in treatment of functional gastrointestinal disorders such as irritable bowel syndrome with constipation and chronic idiopathic constipation (Busby et al., 2010; Camilleri, 2015; Roque and Camilleri, 2011). These drugs may be also used to supply the absence of functional GPs, caused by deleterious mutations in GN (Porto et al., 2015) and/or UGN (Marcolino et al., 2016). However, to date, all drugs derived from GPs present two or three disulfide bonds, which represents a restraint in large-scale production. This occurs due to the possibility of errors in creating the link between the cysteines, which could generate an inactive isoform of the peptide (Fonteles et al., 2001). In this way, the use of LGN could solve this obstacle. Despite the lesser efficacy of LGN in relation to UGN, Fonteles et al. (2001) demonstrated that a single amino acid substitution (Cys7Ser) makes LGN equipotent to UGN in *in vitro* studies, showing that this peptide could be a promising template for drug development.

However, until now, there has been no structural information on LGN. Although X-ray crystallography and nuclear magnetic resonance are powerful tools in determining protein 3D structures, these techniques are time-consuming and expensive. In addition, the structure of some disordered proteins could not be determined by these techniques (Porto et al., 2014). Thus, given the data from PrDOS and the loss of the second disulfide bridge, LGN could be one of these cases. Consequently, structural bioinformatics could be helpful in lowering the costs and time to generate feasible data. Therefore, in order to evaluate the behavior of LGN we performed molecular dynamics simulations of the three-dimensional models of this peptide. LGN was shown to be much more flexible than GN C115Y, with a covariance matrix almost five times higher (Fig. 4, bottom right). Such flexibility was predicted by instability index in ProtParam, where LGN was shown to be more unstable than GN C115Y. According to Porto et al. (2015), the GN C115Y structure is less flexible than the native GN itself, so this mutation was considered deleterious. In this way, flexibility could be an important factor in activity. Given the fact that LGN presents activity in *in vitro* tests, it is possible that this flexibility could influence LGN action.

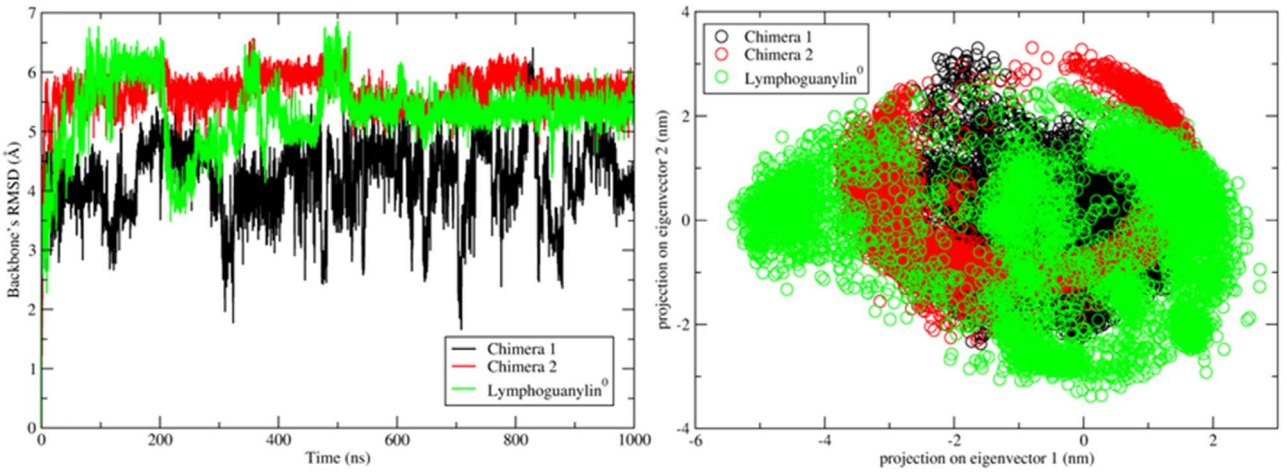

Fig. 6. Analyses of chimeras and LGN^{O} simulation trajectories. Results of molecular dynamics analysis of chimera 1 (black), chimera 2 (red) and LGN^{0} (green) structures. Backbone RMSD variation during the simulations (Left). Projection of the motion of structures in phase space along the first two principal eigenvectors (Right). The values of covariance matrices are: 9.689 for LGN^{0}, 5.472 for chimera 2 and 4.26 for chimera 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Indeed, LGN seems to be a better template than other GPs by the fact that GPs with two disulfide bridges often present two structurally distinct isoforms, and only isoform A is active (Marx et al., 1998; Skelton et al., 1994). In addition, LGN presents an evolutionarily conserved mature sequence, with only two variable sites (Fig. 2). Interestingly, these two residues (Gln^{95} and Glu^{97}) along with Glu^{96} compose a negatively charged N-terminal that could influence pH-dependent activity (Hamra et al., 1997). This is based on the fact that UGN and GN have a pH dependent activity that is highly influenced by N-terminal sequence, which is a marked difference between these peptides, being acidic in UGN and hydrophobic in GN (Fig. 1) (Hamra et al., 1997; Sindic, 2013).

In order to understand the critical factors in the move of LGN we made three additional simulations, exchanging N-terminal sequences of GN C115Y and LGN and neutralizing charges of LGN (Fig. 6). Based on these simulations we could determine the hydrophobic core as a determinant factor in LGN flexibility, which could be explained by the tyrosine between two alanine residues present in GN, which moves to the center of the structure, where it is stabilized by hydrophobic interactions (Fig. 5). Nevertheless, the stereochemical restrictions of the LGN core do not allow great twists, generating the variations of the hydrophobic core. Thus, LGN has more flexibility (Fig. 4).

In this way, given the fact that N-terminal sequences are not determinant in LGN flexibility, it could be possible to make a substitution in the N-terminal sequence without harming LGN activity. Further, the LGN core presents an Asn^{9} classified as a conserved and functional residue. Coincidentally, in most GPs this position presents amino acids that could interact by means of hydrogen bonds (Fig. 1). Considering this, the other core residues (I^{8} and M^{10}) that presented a less conserved grade could be replaced by other aliphatic residues, such as alanine or valine residues, or even the non-proteinogenic amino acids norleucine and norvaline. Besides, it could be used to generate analogs of LGN with activity in specific pHs since the N-terminal is closely linked to the pH dependent activity (Hamra et al., 1993). Furthermore, the substitution of L-amino acids by D-amino acids in the N-terminal could reduce the proteolytic degradation, increasing the peptide's half-life, without affecting the peptide function, since the functional residues are not in the N-terminal.

Moreover, we have to remember that due to the absence of the second disulfide bond, the Cys^{7} could be another useful modification. As demonstrated by Fonteles et al. (2001), substituting this residue by a serine allows easier disulfide bridge formation, since the isoforms with wrong cysteine pairing will be avoided. Also, the C-terminus Tyr^{15} could be another interesting target for modification. In GN C115Y, this residue moves to the core of the structure, as well as Tyr^{9}, providing hydrophobic interactions and a hydrogen bond made by the hydroxyl group in the side chain (Porto et al., 2015). This behavior grants a lower flexibility in GN C115Y despite the loss of cysteine (Porto et al., 2015). However, in LGN, this residue does not remain stabilized as in GN C115Y, due to differences in interaction with the LGN hydrophobic core (Fig. 5).

5. Conclusions

Indeed, flexibility seems to be intrinsically linked to LGN structure. Here, in the first in-depth report of structural characterization of LGN, we showed that LGN is more flexible than GN C115Y, and this flexibility is related to the hydrophobic core. In addition, the similarities between the LGN and GN C115Y structures could be exploited to design new agonists and/or antagonists of guanylate cyclase receptor C. Furthermore, the conservation analyses showed that most of LGN is evolutionarily conserved, but the non-conserved N-terminal region could be used to make modifications to control the pH activity. In the future, new medicines could be developed based on this peptide, helping in the treatment of gastrointestinal disorders. Our investigation of LGN structure clarified some issues in the structure-activity relationship of this peptide, providing new knowledge on guanylin peptides and clarifying the differences between the GN C115Y and LGN structures.

6. Accession number

The preprolymphoguanylin sequence was deposited at GenBank accession KY211743.

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 à Pesquisa do Distrito Federal); Fundect (Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul) 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.2017.02.016.

References

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

Angermüller, C., Biegert, A., Söding, J., 2012. Discriminative modelling of context-specific amino acid substitution probabilities. Bioinformatics 28, 3240–3247. http://dx.doi.org/10.1093/bioinformatics/bts622.

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. U.S.A. 98, 10037–10041. http://dx.doi.org/10.1073/pnas.181342398.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J., 1981. Interaction models for water in relation to protein hydration. Intermol. Forces 31, 331–338. http://dx.doi.org/10.1007/978-94-015-7658-1_21.

Busby, R.W., Bryant, A.P., Bartolini, W.P., Cordero, E.A., Hannig, G., Kessler, M.M., Mahajan-Miklos, S., Pierce, C.M., Solinga, R.M., Sun, L.J., Tobin, J.V., Kurtz, C.B., Currie, M.G., 2010. Linaclotide, through activation of guanylate cyclase C, acts locally in the gastrointestinal tract to elicit enhanced intestinal secretion and transit. Eur. J. Pharm. 649, 328–335. http://dx.doi.org/10.1016/j.ejphar.2010.09.019.

Camilleri, M., 2015. Guanylate cyclase C agonists: Emerging gastrointestinal therapies and actions. Gastroenterology 148, 483–487. http://dx.doi.org/10.1053/j.gastro.2015.01.003.

Carvalho, a.F., Santos-Neto, M.S., Monteiro, H.S. a., Freitas, S.M., Morhy, L., Nascimento, N.R.F., Fonteles, M.C., 2008. BTCI enhances guanylin-induced natriuresis and promotes renal glomerular and tubular effects. Braz. J. Biol. 68, 149–154. http://dx.doi.org/10.1590/S1519-69842008000100021.

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.. http://dx.doi.org/10.1002/ijch.201200096.

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

Chou, K.C., Maggiora, G.M., Mao, B., 1989. Quasi-continuum models of twist-like and accordion-like low-frequency motions in DNA. Biophys. J. 56, 295–305. http://dx.doi.org/10.1016/S0006-3495(89)82676-1.

Chou, K.-C., 2004. Structural bioinformatics and its impact to biomedical science. Curr. Med. Chem. 11, 2105–2134, (0929-8673/04 $45.00+.00).

Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X., Ventura, S., 2007. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinform. 8, 65. http://dx.doi.org/10.1186/1471-2105-8-65.

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. U.S.A. 89, 947–951. http://dx.doi.org/10.1073/pnas.89.3.947.

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.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A., 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res., 32. http://dx.doi.org/10.1093/nar/gkh381.

Fiser, A., Šali, A., 2003. MODELLER: generation and refinement of homology-based protein structure models. Methods Enzym. 374, 461–491. http://dx.doi.org/10.1016/S0076-6879(03)74020-8.

Fonteles, M.C., Carrithers, S.L., Monteiro, H.S., Carvalho, a.F., Coelho, G.R., Greenberg, R.N., Forte, L.R., 2001. Renal effects of serine-7 analog of lymphoguanylin in ex vivo

rat kidney. Am. J. Physiol. Ren. Physiol. 280, F207–F213.
Forte, L.R., 1999. Guanylin regulatory peptides: structures, biological activities mediated by cyclic GMP and pathobiology. Regul. Pept.. http://dx.doi.org/10.1016/S0167-0115(99)00033-6.
Forte, L.R., Eber, S.L., Fan, X., London, R.M., Wang, Y., Rowland, L.M., Chin, D.T., Freeman, R.H., Krause, W.J., 1999. Lymphoguanylin: cloning and characterization of a unique member of the guanylin peptide family. Endocrinology 140, 1800–1806. http://dx.doi.org/10.1210/en.140.4.1800.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server, in: The Proteomics Protocols Handbook, 571–607. http://dx.doi.org/10.1385/1592598900.
Guruprasad, K., Reddy, B.V.B., Pandit, M.W., 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. Des. Sel. 4, 155–161. http://dx.doi.org/10.1093/protein/4.2.155.
Hamra, F.K., Forte, L.R., Eber, S.L., Pidhorodeckyj, N.V., Krause, W.J., Freeman, R.H., Chin, D.T., Tompkins, J.A., Fok, K.F., Smith, C.E., et al., 1993. Uroguanylin: structure and activity of a second endogenous peptide that stimulates intestinal guanylate cyclase. Proc. Natl. Acad. Sci. U.S.A. 90, 10464–10468. http://dx.doi.org/10.1073/pnas.90.22.10464.
Hamra, F.K., Eber, S.L., Chin, D.T., Currie, M.G., Forte, L.R., 1997. Regulation of intestinal uroguanylin/guanylin receptor-mediated responses by mucosal acidity. Proc. Natl. Acad. Sci. U.S.A. 94, 2705–2710. http://dx.doi.org/10.1073/pnas.94.6.2705.
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. GRGMACS 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.
Hill, O., Kuhn, M., Zucht, H.D., Cetin, Y., Kulaksiz, H., Adermann, K., Klock, G., Rechkemmer, G., Forssmann, W.G., Magert, H.J., 1995. Analysis of the human guanylin gene and the processing and cellular localization of the peptide. Proc. Natl. Acad. Sci. U.S.A. 92, 2046–2050. http://dx.doi.org/10.1073/pnas.92.6.2046.
Ishida, T., Kinoshita, K., 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res., 35. http://dx.doi.org/10.1093/nar/gkm363.
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. PLoS One, 9. http://dx.doi.org/10.1371/journal.pone.0104311.
Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637. http://dx.doi.org/10.1002/bip.360221211.
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.. 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. http://dx.doi.org/10.1371/journal.pcbi.1003318.
Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132. http://dx.doi.org/10.1016/0022-2836(82)90515-0.
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.
Lauber, T., Neudecker, P., Rösch, P., Marx, U.C., 2003. Solution structure of human proguanylin: the role of a hormone prosequence. J. Biol. Chem. 278, 24118–24124. http://dx.doi.org/10.1074/jbc.M300370200.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. http://dx.doi.org/10.1093/bioinformatics/btl158.
Lima, A.A.M., Fonteles, M.C., 2014. From Escherichia coli heat-stable enterotoxin to mammalian endogenous guanylin hormones. Braz. J. Med. Biol. Res.. http://dx.doi.org/10.1590/1414-431×20133063.
Marcolino, A.C.S., Porto, W.F., Pires, Á.S., Franco, O.L., Alencar, S.A., 2016. Structural impact analysis of missense SNPs present in the uroguanylin gene by long-term molecular dynamics simulations. J. Theor. Biol. 410, 9–17. http://dx.doi.org/10.1016/j.jtbi.2016.09.008.
Marx, U.C., Klodt, J., Meyer, M., Gerlach, H., Rösch, P., Forssmann, W.G., Adermann, K., 1998. One peptide, two topologies: structure and interconversion dynamics of human uroguanylin isomers. J. Pept. Res. 52, 229–240. http://dx.doi.org/10.1111/j.1399-3011.1998.tb01480.x.
Mayrose, I., Graur, D., Ben-Tal, N., Pupko, T., 2004. Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. Mol. Biol. Evol. 21, 1781–1791. http://dx.doi.org/10.1093/molbev/msh194.
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.
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.
Porto, W.F., Nolasco, D.O., Franco, O.L., 2014. Native and recombinant Pg-AMP1 show different antibacterial activity spectrum but similar folding behavior. Peptides 55, 92–97. http://dx.doi.org/10.1016/j.peptides.2014.02.010.
Porto, W.F., Franco, O.L., Alencar, S. a., 2015. Computational analyses and prediction of guanylin deleterious SNPs. Peptides, 1–11. http://dx.doi.org/10.1016/j.peptides.2015.04.013.
Roque, M.V., Camilleri, M., 2011. Linaclotide , a synthetic guanylate cyclase C agonist , for the treatment of functional gastrointestinal disorders associated with constipation. Expert Rev. Gastroenterol. Hepatol. 5, 301–310.
de Sauvage, F.J., Keshav, S., Kuang, W.J., Gillett, N., Henzel, W., Goeddel, D.V., 1992. Precursor structure, expression, and tissue distribution of human guanylin. Proc. Natl. Acad. Sci. U.S.A. 89, 9089–9093.
Shailubhai, K., Comiskey, S., Foss, J.A., Feng, R., Barrow, L., Comer, G.M., Jacob, G.S., 2013. Plecanatide, an oral guanylate cyclase C agonist acting locally in the gastrointestinal tract, is safe and well-tolerated in single doses. Dig. Dis. Sci. 58, 2580–2586. http://dx.doi.org/10.1007/s10620-013-2684-z.
Shailubhai, K., Palejwala, V., Priya Arjunan, K., Saykhedkar, S., Nefsky, B., Foss, J.A., Comiskey, S., Jacob, G.S., Plevy, S.E., 2015. Plecanatide and dolcanatide, novel guanylate cyclase-C agonists, ameliorate gastrointestinal inflammation in experimental models of murine colitis Basic study. World J. Gastrointest. Pharm. Ther. Novemb. 6, 213–222. http://dx.doi.org/10.4292/wjgpt.v6.i4.213.
Sindic, A., 2013. Current understanding of guanylin peptides actions. ISRN Nephrol. 2013, 813648. http://dx.doi.org/10.5402/2013/813648.
Sindic, A., Schlatter, E., 2006. Cellular effects of guanylin and uroguanylin. J. Am. Soc. Nephrol. 17, 607–616. http://dx.doi.org/10.1681/asn.2005080818.
Skelton, N.J., Garcia, K.C., Goeddel, D.V., Quan, C., Burnier, J.P., 1994. Determination of the solution structure of the peptide hormone guanylin: observation of a novel form of topological stereoisomerism. Biochemistry 33, 13581–13592. http://dx.doi.org/10.1021/bi00250a010.
Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res., 33. http://dx.doi.org/10.1093/nar/gki408.
Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23, 1282–1288. http://dx.doi.org/10.1093/bioinformatics/btm098.
Touw, W.G., Baakman, C., Black, J., Te Beek, T.A.H., Krieger, E., Joosten, R.P., Vriend, G., 2015. A series of PDB-related databanks for everyday needs. Nucleic Acids Res. 43, D364–D368. http://dx.doi.org/10.1093/nar/gku1028.
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. http://dx.doi.org/10.1093/nar/gkm290.
Yang, Z., Kurpiewski, M.R., Ji, M., Townsend, J.E., Mehta, P., Jen-Jacobson, L., Saxena, S., 2012. PNAS Plus: ESR spectroscopy identifies inhibitory Cu2+ sites in a DNA-modifying enzyme to reveal determinants of catalytic specificity. Proc. Natl. Acad. Sci. 109, E993–E1000. http://dx.doi.org/10.1073/pnas.1200733109.
Yuge, S., Inoue, K., Hyodo, S., Takei, Y., 2003. A novel guanylin family (guanylin, uroguanylin, and renoguanylin) in eels: possible osmoregulatory hormones in intestine and kidney. J. Biol. Chem. 278, 22726–22733. http://dx.doi.org/10.1074/jbc.M303111200.
Zhou, G.-P., 1989. Biological functions of soliton and extra electron motion in DNA structure. Phys. Scr. 40, 698–701. http://dx.doi.org/10.1088/0031-8949/40/5/021.