Prediction of the Impact of Coding Missense and Nonsense Single Nucleotide Polymorphisms on HD5 and HBD1 Antibacterial Activity Against *Escherichia coli*

William F. Porto,[1,2] Diego O. Nolasco,[1,3] Állan S. Pires,[1,2] Rinaldo W. Pereira,[1] Octávio L. Franco,[1,2,4] Sérgio A. Alencar[1]

[1]*Programa De Pós-Graduação Em Ciências Genômicas E Biotecnologia, Universidade Católica De Brasília, Brasília, DF, Brazil*

[2]*Centro De Análises Proteômicas E Bioquímicas, Pós-Graduação Em Ciências Genômicas E Biotecnologia, Universidade Católica De Brasília, Brasília, DF, Brazil*

[3]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA*

[4]*S-Inova Biotech, Pos-Graduação Em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, MS, Brazil*

Received 23 December 2015; revised 14 March 2016; accepted 26 April 2016
Published online 10 May 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22866

ABSTRACT:

Defensins confer host defense against microorganisms and are important for human health. Single nucleotide polymorphisms (SNPs) in defensin gene-coding regions could lead to less active variants. Using SNP data available at the dbSNP database and frequency information from the 1000 Genomes Project, two DEFA5 *(L26I and R13H) and eight* DEFB1 *(C35S, K31T, K33R, R29G, V06I, C12Y, Y28* and C05*) missense and nonsense SNPs that are located within mature regions of the coded defensins were retrieved. Such SNPs are rare and population restricted. In order to assess their antibacterial activity against* Escherichia coli, *two linear regression models were used from a previous work, which models the antibacterial activity as a function of solvation potential energy, using molecular dynamics data. Regarding only the antibacterial predictions, for HD5, no biological differences between wild-type and its variants were observed; while for HBD1, the results suggest that the R29G, K31T, Y28* and C05* variants could be less active than the wild-type one. The data here reported could lead to a substantial improvement in knowledge about the impact of missense SNPs in human defensins and their world distribution. © 2016 Wiley Periodicals, Inc. Biopolymers (Pept Sci) 106: 633–644, 2016.*

Keywords: DEFA5; DEFB1; 1000 genomes project; protein solvation potential energy; molecular dynamics

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of any preprints from the past two calendar years by emailing the Biopolymers editorial office at biopolymers@wiley.com.

INTRODUCTION

Defensins are a family of small antimicrobial peptides that have been found in animals,[1,2] fungi,[3] and plants.[4] These molecules present low molecular masses (3.5–6 kDa) and positive net charge.[5] Currently, defensins are divided into four groups, including CS-$\alpha\beta$ present in invertebrates, fungi, and plants, and α-, β- and θ-defensins, present in vertebrates.[2]

Additional Supporting Information may be found in the online version of this article.

Correspondence to: Octávio L. Franco; e-mail: ocfranco@gmail.com

Contract grant sponsor: CNPq, CAPES, FUNDECT, and FAPDF

In humans, the most prominent subfamilies are the α- and β-defensins, with several reported molecules.[1,6] The α- and β-defensins of vertebrates could be involved in several defense processes, acting directly or indirectly against a great number of pathogenic microorganisms, such as bacteria, fungi, and viruses.[1] They could act in the chemotaxis of macrophages and lymphocytes, or even killing microorganisms by directly making pores in cell membranes, inhibiting formation of bacterial cell wall and inhibiting DNA's transcription and replication.[7–10]

The α-defensins are peptides that are present in azurophilic granules of neutrophils, epithelial, and mucosal tissues. Such molecules are of great immunological importance, acting on several pathways of immune response.[5] The β-defensins, on the other hand, are mainly present on mucous membranes, including respiratory and urinary tracts, as well as keratinocytes in skin.[11] Among the human defensins, the α-defensin 5 (HD5) and the β-defensin 1 (HBD1) are involved in several diseases, including inflammatory bowel disease,[12] asthma,[13,14] systemic lupus erythematosus,[15] atopy,[14] atopic dermatitis,[16,17] psoriasis,[18] Crohn disease,[19,20] AIDS,[21,22] and infertility.[23]

HD5 is encoded by the *DEFA5* gene in form of a prepropeptide, while the mature peptide is composed of 32 amino acid residues (positions 63 to 94). HD5 is synthesized and secreted by Paneth cells,[24] and it is also found in epithelial tissues of the female reproductive[25] and nasal tracts.[26]

On the other hand, HBD1 is constitutively produced in epithelial cells[5] and is encoded by the *DEFB1* gene. HBD1 is also produced in the form of a prepropeptide, in a complete sequence of 68 amino acid residues, whereas the mature peptide comprises 36 residues (33 to 68). HBD1 operates mainly in the skin, respiratory, and urogenital tracts.[27]

Although both of these peptides show activity against Gram-negative bacteria,[11,28] only HD5 is active against Gram-positive strains.[28,29] In previous works, it was demonstrated that such activity could be affected by point mutations in amino acid residues in the protein sequence.[11,30] However, there are no extensive studies about the effect of single nucleotide polymorphisms (SNPs) in HD5 and HBD1 antimicrobial activity, neither studies about the world SNPs distribution.

Large-scale sequencing projects, such as the 1000 Genomes Project,[31] represents a valuable resource of SNP frequency information from several world populations, and can be used to obtain information about the worldwide distribution of potentially deleterious defensin missense SNPs. Additionally, the low allele frequency of an amino acid variant can, by itself, serve as a predictor of its functional significance.[32]

Besides, over the past few years, molecular dynamics has been applied in a series of studies to evaluate the consequences of amino acid residue changes in proteins, including studies with guanylin,[33] P protein,[34] Ras-related C3 botulinum toxin substrate 1,[35] protein tyrosine phosphatase 1B,[36] aldosterone synthase,[37] and angiogenin.[38] However, recently our group demonstrated that, for HD5 and HBD1, structural parameters such as solvent accessible surface area, radius of gyration, and flexibility do not reflect the antibacterial activity against *Escherichia coli*.[39] The activity is actually correlated to their solvation potential energy.[39] Such correlation could be used to study the effects of all known SNPs that change amino acid residues (missense SNPs) in the mature HD5 and HBD1 proteins.

To date, the functional and structural impact of missense SNPs on defensins has not been predicted using computational tools. Therefore, in this study, we predicted the effects of missense SNPs on the HD5 and HBD1 antibacterial activity against *E. coli* through the evaluation of solvation potential energy. We further examined the native and variant protein structures for alterations which could have an impact on function. Our *in silico* analyses suggest the existence of several defensin variants that could be less active than the wild-type defensin.

MATERIAL AND METHODS

Defensin Gene Selection

From the NCBI Gene database (http://www.ncbi.nlm.nih.gov/gene), a set of 68 human defensin genes was collected. From these, only genes coding for products which have three-dimensional structures available in the Protein Data Bank (PDB - http://www.rcsb.org) were selected. Then, in order to select only validated SNPs, the genes that contained missense SNPs available in dbSNP (http://www.ncbi.nlm.nih.gov/snp) and frequency information available at the "1000 Genomes Project"[31] were filtered using a custom script. Finally, the SNPs that alter amino acids present in the resolved part of the protein available in PDB were selected for further analysis. As rare SNPs occur at very low frequencies (< 1%), there is great concern to avoid confounding putative SNPs with sequencing errors common in next-generation sequencing technologies. However, there are some recommendations which can help to minimize this problem, such as to aim for coverage with high-quality bases to reach at least 20x in the positions of the called SNPs.[40] Therefore, SNPs obtained through targeted deep (50-100X) exome sequencing were selected as well as a combination of this and low-coverage (2-6X) whole-genome sequencing (Figure 1). The only exception was variant allele rs142540429:G > T, which was, however, validated in dbSNP by multiple, independent submissions to the refSNP cluster. The flowchart of this initial screening is showed in Figure 1.

Frequency Data

The SNP frequency data of *DEFA5* and *DEFB1* genes were obtained from the 1000 Genomes Project (phase 3) browser (http://browser.1000genomes.org).[31] The browser provides the frequencies of all SNPs identified in the genomes of 2,504 individuals from 26 populations obtained through a combination of low-coverage (2-6x) whole-genome sequence data, targeted deep (50-100x) exome sequencing and dense SNP genotype data. The 26 populations studied were grouped by the predominant component of ancestry into five super-populations:

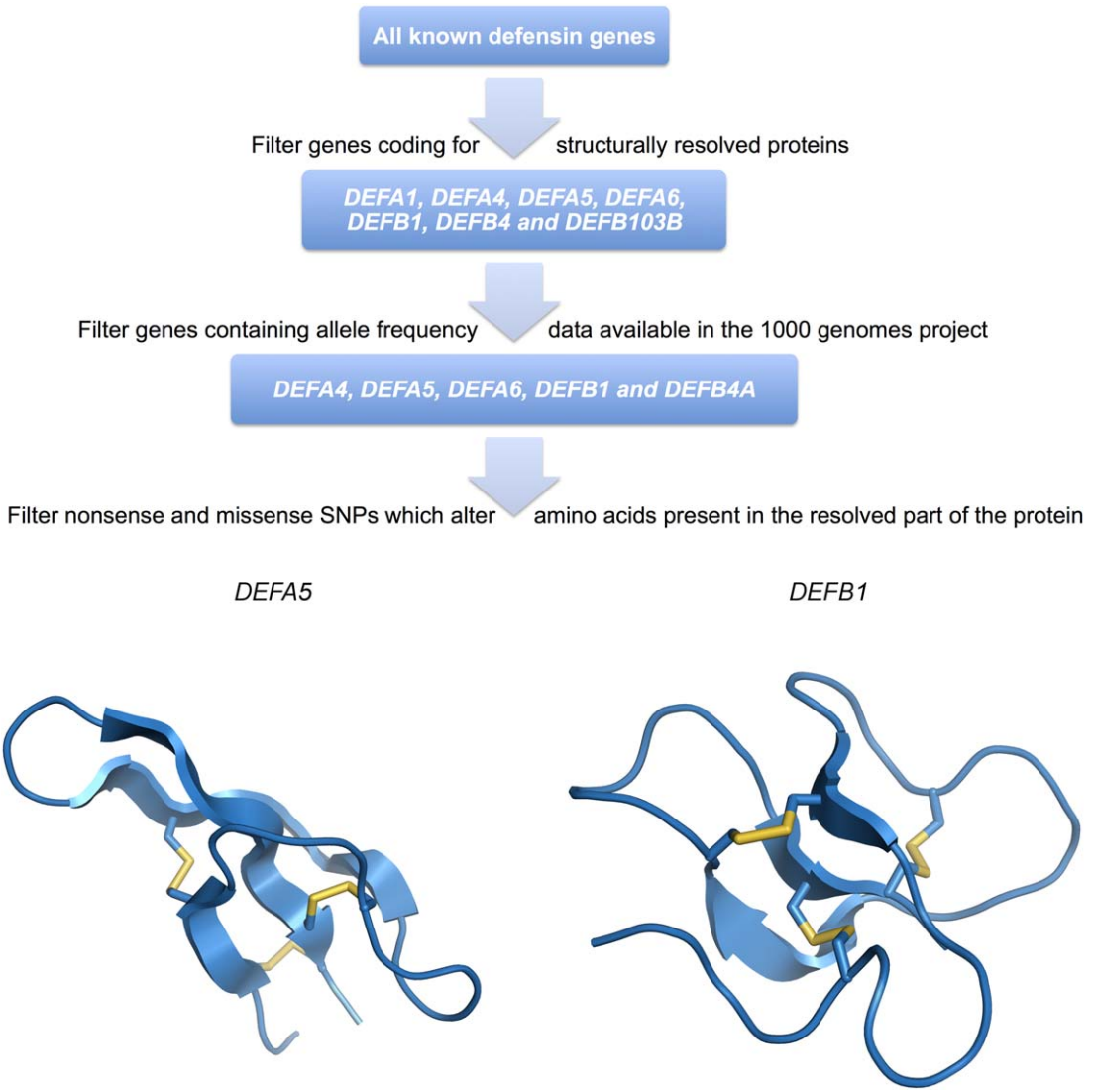

FIGURE 1 Selection of missense and nonsense SNPs in defensin genes. Only two genes (*DEFA5* and *DEFB1*) out of 68 human defensin genes fulfilled our established criteria for further analysis.

African (AFR) (661 individuals), Ad Mixed American (AMR) (347 individuals), East Asian (EAS) (504 individuals), South Asian (489 individuals), and European (EUR) (503 individuals). The fixation index (Fst) was measured using the "weir-fst-pop" option of vcftools,[41] which is based on the model developed by Weir and Cockerham.[42]

Nucleotide and Protein Accessions

DEFA5 and *DEFB1* nucleotide sequences were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank) and are both relative to the open reading frame from GenBank Accession number NT_023736.17. The amino acid positions of precursors HD5 and HBD1 are relative to GenBank Accession numbers NP_066290.1 and NP_005209.1, respectively. The amino acid positions of mature HD5 and HBD1 are relative to PDB IDs 1ZMP[43] and 1KJ5,[44] respectively.

In Silico Functional Analyses of *DEFA5* and *DEFB1* Missense SNPs

In order to evaluate the potential functional impact of the obtained *DEFA5* and *DEFB1* missense SNPs, a total of 16 computational tools were used, divided into four different groups, according to their methodological approaches. We filtered all SNPs that were classified as deleterious by at least three tools in each of the four groups, and denominated these as convergent deleterious predicted SNPs, as described by Porto et al.[33]

Sequence homology-Based Methods. The following methods based on sequence homology principles were used to produce missense SNP functional predictions: Sorting Intolerant From Tolerant (SIFT),[45] Provean,[46] Mutation Assessor,[47] and Panther.[48]

Supervised Learning Methods. Supervised learning algorithms used for missense SNP impact prediction included neural networks (SNAP)[49] and support vector machines (PhD-SNP,[50] SuSPect,[51] and SNPs&GO)[52].

Protein Sequence and Structure-Based Methods. The following methods either combine information from protein sequence and structure or use protein structural information alone to analyze missense variants: PolyPhen,[53] Site Directed Mutator (SDM),[54] HOTMuSiC,[55] and PoPMuSiC.[56]

Consensus-Based Methods. In order to obtain a consensus score based on many different SNP impact prediction strategies, the following types of consensus software were used: Condel,[57] Meta-SNP,[58] PON-P2,[59] and PredictSNP.[60]

Molecular Modeling

One hundred molecular models for each variant were constructed by comparative molecular modeling through MODELLER 9.14,[61] using the structures of HD5 and HBD1 (PDB IDs: 1ZMP[43] and 1KJ5,[44] respectively). The models were constructed using the default methods of automodel and environ classes from MODELLER. The final models were selected according to the discrete optimized protein energy score (DOPE score). This score assesses the energy of the model and indicates the best probable structures. The best models were evaluated through PROSA II[62] and PROCHECK.[63] PROCHECK checks the stereochemical quality of a protein structure through the Ramachandran plot, where good quality models are expected to have more than 90% of amino acid residues in most favored and additional allowed regions, while PROSA II indicates the fold quality. Structure visualization was done in PyMOL (http://www.pymol.org).

Molecular Dynamics Simulations

The molecular dynamics simulations were carried out according to Porto et al.[39] Native wild-type and variant structures were simulated in water environment, using the Single Point Charge water model.[64] The analyses were performed by using the GROMOS96 43A1 force field and computational package GROMACS 4.[65] The dynamics used the HD5 and HBD1 tridimensional structures or the three-dimensional models from their respective variants as initial structures, immersed in water, in cubic boxes with a minimum distance of 0.7 nm between the proteins and the edges of the boxes. This distance provides the smallest possible box without generating artifacts in the simulation; and requires lower computational time, since less bulk solvent molecules are simulated. In addition, Souza and Ornstein[66] demonstrated that a larger box size (1.0 or 1.5 nm) does not afford any extra benefit over a smaller box. Chlorine ions were also inserted at the complexes with positive charges in order to neutralize the system charge. Geometry of water molecules was constrained by using the SETTLE algorithm.[67] All atom bond lengths were linked by using the LINCS algorithm.[68] Electrostatic corrections were made by Particle Mesh Ewald algorithm,[69] with a cut-off radius of 1.4 nm in order to minimize the computational time. The same cut-off radius was also used for van der Waals interactions. The list of neighbors of each atom was updated every 20 simulation steps of 2 fs. The system underwent an energy minimization using 50,000 steps of the steepest descent algorithm. After that, the system temperature was normalized to 310 K for 100 ps, using the velocity rescaling thermostat (NVT ensemble). Then the system pressure was normalized to 1 bar for 100 ps, using the Parrinello-Rahman barostat (NPT ensemble). The systems with minimized energy, balanced temperature and pressure were simulated for 100 ns by using the leap-frog algorithm. The initial and the final structures were compared through RMSD and TM-Score,[70] where TM-Scores above 0.5 indicate that the two structures share the same fold.[71] Each simulation was repeated three times.

Prediction of Median Lethal Dose against *Escherichia coli*

The predictions of LD_{50} were performed according to Porto et al.[39] One snapshot from each trajectory replicate was taken at 100 ns for solvation potential energy calculation. The snapshots were taken using the utility trjconv from GROMACS. The conversion of pdb files into pqr files was perfomed by the utility PDB2PQR using the AMBER force field.[72] The grid dimensions for APBS calculation were also determined by PDB2PQR. Solvation potential energy was calculated by APBS.[73] Then, the prediction of LD_{50} of variants was performed using the linear regression models described by Porto et al.,[39] where the predicted LD_{50} is described as a function of solvation potential energy by Eqs. (1) and (2) for HD5 and HBD1, respectively:

$$pLD50 = -4.402 esolv \times 10^{-3} + 7.626315 \quad (1)$$

$$pLD50 = -6.634 esolv \times 10^{-3} + 13.998061 \quad (2)$$

Such linear regression models have a good assessment according to the R^2 values (0.7353 and 0.8236 for HD5 and HBD1, respectively).[39] A one-sided Student's *t*-test was applied for verifying the differences between the predicted LD_{50} of wild-type defensin structures and their SNP variants with a critical value of 0.05, with the null hypothesis that the LD_{50}.of the variants are not greater (less potent/less active) than the wild-type one. The variants with significant statistical differences were considered as less active. Statistics were calculated using the *R* package for statistical computing (http://www.r-project.org).

RESULTS

DEFA5 and *DEFB1* Have a Similar Number of SNPs in Exons, but *DEFB1* Has More Missense SNPs in the Mature Defensin Product

Only *DEFA5* (which encodes HD5) and *DEFB1* (which encodes HBD1) genes fitted our selection criteria (Figure 1). Therefore, to obtain a complete picture of SNP variation within the *DEFA5* and *DEFB1* genes, the distribution of all SNPs identified in the coding and non-coding regions was analyzed. A total of 1,431 SNPs were identified in *DEFA5* and 7,433 in *DEFB1*, accounting for a SNP frequency of 1/11bp and 1/15bp, respectively. The SNPs were divided into different classes, based on their location (up or downstream, exonic, intronic, 5′ UTR and 3′ UTR). It can be seen from Figure 2 that the distribution of SNPs in these two genes differs considerably. The *DEFA5* gene has a considerably higher percentage of SNPs

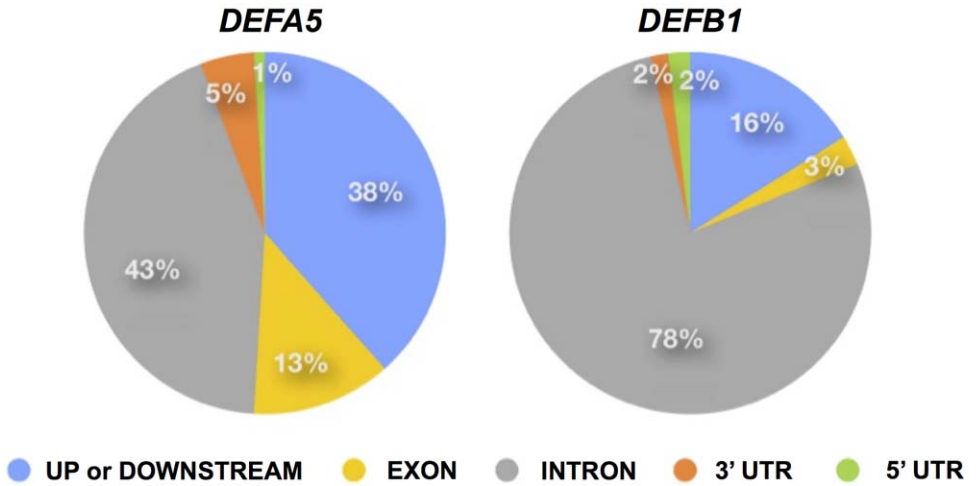

FIGURE 2 Distribution of SNPs within *DEFA5* and *DEFB1* genes. The distribution was based on the SNP locations (Up or Downstream, exonic, intronic, 5′ UTR and 3′ UTR).

located in up/downstream regions and in exons, compared to the *DEFB1* gene, which in turn has a much higher percentage of SNPs located in introns. However, in exon regions, they showed similar amounts, in absolute values being 186 SNPs for *DEFA5* and 223 for *DEFB1*.

The Majority of SNPs in Mature Defensin Products Is Population Restricted

Considering only SNPs in the mature defensin products, two SNPs in the *DEFA5* gene (both conservative) and eight SNPs in the *DEFB1* gene (two conservative, four non-conservative and two nonsense) were retrieved from the initial screening (Table I). From these, six rare SNPs identified within *DEFA5* and *DEFB1* occurred in only one of the four populations analyzed (Table I). As showed in Table II, according to the Weir and Cockeram's fixation index (Fst), all of the SNPs showed little population differentiation, with the exception of one variant allele (rs2738047:C > T), which showed moderate differentiation in the African/Ad Mixed American, African/East Asian + South Asian and African/European population

Table I SNPs Retrieved From Data Mining Process and Their Allele Frequencies

Gene	SNP rs#	Amino Acid Change[a]	Change in Mature Product[b]	SNP Type	Global Minor Allele Frequency (%)[c]	Global Allele Count	Relative Allele Frequency (%)[d]			
							AFR	AMR	EAS + SAS	EUR
DEFA5	rs142540429:G>T	p.Leu88Ile	L26I	Conservative	0.12	6	0.38	0.14	–	–
	rs148244786:C>T	p.Arg75His	R13H	Conservative	0.02	1	–	–	0.05	–
DEFB1	rs201508504:T>C	p.Lys65Arg	K33R	Conservative	0.02	1	–	0.14	–	–
	rs2738047:C>T	p.Val38Ile	V06I	Conservative	3.97	199	13.84	0.43	0.55	0.20
	rs201260899:C>T	p.Cys44Tyr	C12Y	Non-Conservative	0.02	1	–	–	–	0.10
	rs1800968:A>T	p.Cys67Ser	C35S	Non-Conservative	0.08	4	–	0.29	–	0.20
	rs144111819:T>G	p.Lys63Thr	K31T	Non-Conservative	0.08	4	0.30	–	–	–
	rs147178531:T>C	p.Arg61Gly	R29G	Non-Conservative	0.02	1	0.08	–	–	–
	rs5743490:G>T	p.Cys37Ter[e]	C05[e]	Nonsense	0.20	10	0.68	–	–	0.10
	rs140403947:G>T	p.Tyr60Ter[e]	Y28[e]	Nonsense	0.08	4	–	–	0.20	–

[a] *DEFA5* and *DEFB1* nucleotide sequences are both relative to the open reading frame from GenBank Accession numbers NT_023736.17, and amino acid positions are relative to GenBank Accession numbers NP_066290.1 and NP_005209.1, respectively.

[b] The amino acid positions of *DEFA5* and *DEFB1* are relative to PDB IDs 1ZMP and 1KJ5, respectively.

[c] The global minor allele frequency was obtained by analyzing genotype data from 2,504 individuals. All frequency data were obtained from the 1000 Genomes Project (phase 3) browser (http://browser.1000genomes.org).

[d] The populations analyzed represent five super populations: African (AFR), Ad Mixed American (AMR), East Asian (EAS), South Asian (SAS) and European (EUR). The relative allele frequencies were obtained from the allele frequencies of 661, 347, 993 and 503 individuals sequenced in the AFR, AMR, EAS + SAS and EUR super populations, respectively.

[e] This symbol means that the variation changed the codon to a termination codon, which causes premature termination of the protein).

Table II Weir and Cockeram's Fixation indices (Fst) obtained from six pairwise population comparisons for the selected SNPs

Gene	SNP rs#[b]	Amino Acid Change in Mature Product[c]	Pairwise population Fst[d]					
			AFR/ AMR	AFR/ EAS+SAS	AFR/ EUR	AMR/ EAS+SAS	AMR/ EUR	EAS+SAS/ EUR
DEFA5	rs142540429:G>T	L26I	−0.0006	0.0003	0.0009	–	–	–
	rs148244786:C>T	R13H	0.0009	–	–	0.0013	0.0022	–
DEFB1	rs1800968:A>T	C35S	0.0041	–	0.0005	0.0049	−0.0009	0.0008
	rs144111819:T>G	K31T	0.0047	0.0069	0.0087	–	–	–
	rs201508504:T>C	K33R	0.0009	–	–	0.0013	0.0022	–
	rs147178531:T>C	R29G	0.0011	0.0025	0.0035	–	–	–
	rs2738047:C>T	V06I	0.0981	0.0910	0.1388	0.0064	–	0.0107
	rs201260899:C>T	C12Y	–	–	−0.0006	–	−0.0011	−0.0004
	rs140403947:G>T	Y28[a]	–	0.0030	–	0.0020	–	0.0046
	rs5743490:G>T	C05[a]	0.0074	–	0.0005	0.0084	0.0016	0.0008

[a] The "–" symbol indicates that the SNP was not observed in either of the two populations in the pairwise comparison.
[b] *DEFA5* and *DEFB1* nucleotide sequences are both relative to the open reading frame from GenBank Accession numbers NT_023736.17.
[c] The amino acid positions of *DEFA5* and *DEFB1* are relative to PDB IDs 1ZMP and 1KJ5, respectively.
[d] The populations analyzed represent four super populations: African (AFR), Ad Mixed American (AMR), East Asian (EAS) plus South Asian (SAS) and European (EUR)

pairwise comparisons (Fst values of 0.0981, 0.0910 and 0.1388, respectively).

Prediction of Convergent Deleterious SNPs

The *in silico* missense SNP impact analysis using 16 different softwares did not result in any convergent deleterious predicted SNPs for the two DEFA5 SNPs analyzed, while for DEFB1 two SNPs were predicted as convergent deleterious (C12Y and C35S) (Table III).

HD5 Variants Seem to Maintain the Antibacterial Activity

Two conservative variations were found in *DEFA5*, and these generate the variants L26I and R13H (Figure 3). The two amino acid variations did not provide clear changes in the physicochemical properties of HD5, since the change from Leu to Ile and from Arg to His conserves the amino acid characteristic, keeping a hydrophobic and a positively charged residue, respectively. In addition, HD5 forms a homodimer and the variations are located away from the dimeric interface (data not shown). The conservative character from both variations is reflected in the protein structure, where no significant modifications were observed.

The HD5's three-dimensional structure (PDB ID: 1ZMP) was determined by high-resolution X-ray diffraction[74] and is composed of a homodimer containing two antiparallel β-sheets with three β-strands (Figure 3). This native wild-type structure was used as a template for constructing the structure of L26I, since the R13H variant has a three-dimensional structure solved by X-ray diffraction (PDB ID: 3I5W).[20] The validation parameters for the L26I modeling are summarized in Supporting Information Table S1. Structural changes were then evaluated by molecular dynamics simulations. Using the native wild-type structure as a reference, it was observed that the variations did not exert great impact on structure, since the simulation behaviors were similar, since the RMSD values for HD5 and variants were below 3Å (Table IV). In addition, the TM-Scores for each structure were above 0.5, which indicates that the wild-type and the variant structures are in the same fold (Table IV). This data is in agreement with previous reports.[30,39]

In this context, in the absence of structural modification and conservation of physicochemical properties, both variants could have similar activities compared to the native wild-type structure. Using the linear regression model for HD5 variants,[39] the predicted LD_{50} value against *E. coli* for R13H presented a statistical difference in relation to wild-type, however, the difference is biologically irrelevant (Figure 3). Similarly, no significant statistical difference was observed between wild-type and L26I.

HBD1 Variants Could Show Different Antibacterial Activities

Eight variations were found for HBD1, being two conservative (K33R and V06I), four non-conservative (K31T, R29G, C12Y and C35S) and two nonsense (Y28* and C05*) (Figure 4). Initially, K33R and V06I are in a similar context to HD5 variants, not causing great changes in the physicochemical properties of HBD1, taking into account charge and hydrophobicity,

Table III Prediction Results of *DEFA5* and *DEFB1* Missense SNPs Analyzed by 16 Bioinformatics Tools Classified in Four Different Groups

Gene	SNP rs#	Amino Acid Change in Mature Product[a]	Sequence-based[b]					SLM-based[b]				Consensus-based[b]				Structure-based[b]		
			SIFT	Provean	Mutation Assessor	Panther	PhD-SNP	SNP&GO	SNAP	SuSPect	Condel	MetaSNP	PON-P2	Predict SNP	PolyPhen	SDM	HOT MuSiC	PoP MuSiC
DEFA5	rs142540429:G>T	p.Leu88Ile	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	rs148244786:C>T	p.Arg75His	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
DEFB1	rs201508504:T>C	p.Lys65Arg	N	N	N	N	N	N	N	N	N	N	-	N	N	N	D	D
	rs2738047:C>T	p.Val38Ile	N	N	N	N	N	N	D	N	N	N	N	N	D	N	D	D
	rs201260899:C>T	**p.Cys44Tyr**	D	D	D	D	D	D	D	D	D	D	D	D	D	N	D	D
	rs1800968:A>T	**p.Cys67Ser**	D	D	D	D	N	D	D	D	D	N	D	D	D	D	D	D
	rs144111819:T>G	p.Lys63Thr	D	D	N	D	N	D	D	N	N	N	N	D	D	N	D	N
	rs147178531:T>C	p.Arg61Gly	N	N	N	N	N	N	D	N	N	N	N	N	N	D	N	D

Convergent deleterious SNPs are in bold face.

[a] *DEFA5* and *DEFB1* nucleotide sequences are both relative to the open reading frame from GenBank Accession numbers NT_023736.17, and amino acid positions are relative to GenBank Accession numbers NP_066290.1 and NP_005209.1, respectively.

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

respectively. On the other hand, the four non-conservative variants could cause great modifications in HBD1 properties: two of them changing physicochemical properties (K31T and R29G) and the other two, the structure itself (C12Y and C35S). In the first case, the two variations cause a reduction in the net charge, from +5 to +4; and in the second case, despite the conservation of net charge, one disulfide bridge is broken, which could make the structure more flexible. Finally, the nonsense variations, Y28* and C05*, undoubtedly cause the most drastic changes, since they generate truncated proteins (Figure 4).

The structure of HBD1 (PDB ID: 1KJ5) was determined by nuclear magnetic resonance spectroscopy,[44] and it is composed of a monomeric β-sheet with three β-strands (Figure 4). Its structure was used as a template to construct the variant structures (except for C05*, which was not constructed). The validation parameters for molecular modeling are summarized in Supporting Information Table S1. Initially, the wild-type simulations showed a very dynamic behavior, with RMSD values varying from 2.28 to 3.18 Å and, according to TM-Scores, the fold was maintained only in one simulation (Table IV). This characteristic was maintained in the conservative and non-conservative variants simulations, in which at least in one simulation they were not in the same fold as the initial structure, according to the TM-Score (Table IV). However, the two nonsense SNPs have the greatest impact on the HBD1 structure. The C05* variant is the more drastic of the two, since it deletes most of the mature sequence, resulting in only four amino acids residues. The other nonsense SNP, Y28*, generates a variant with 23 amino acid residues. This variation has a great impact on the structure, since it lacks two disulfide bonds. RMSD values above 6Å and a TM-Score below 0.5 were observed for Y28* in the three simulations (Table IV), indicating that this structure does not share the same folding as that seen at the beginning of the simulation.

Using the linear regression model for HBD1 variants,[39] the predicted LD_{50} values against *E. coli* were very similar among wild-type and the variants with net charge of +5: V06I, K33R, C35S, and C12Y (Figure 4). For the variants with charge reduction, R29G and K31T, the predicted LD_{50} values were statistically higher than wild-type predicted LD_{50} (Figure 4), indicating that antibacterial activity could be affected by the amino acid variations. In the case of Y28*, it showed the highest predicted LD_{50} among the variants, however, its structure is completely different from the wild-type HBD1, and our prediction model should not work properly for this truncated HBD1 variant.

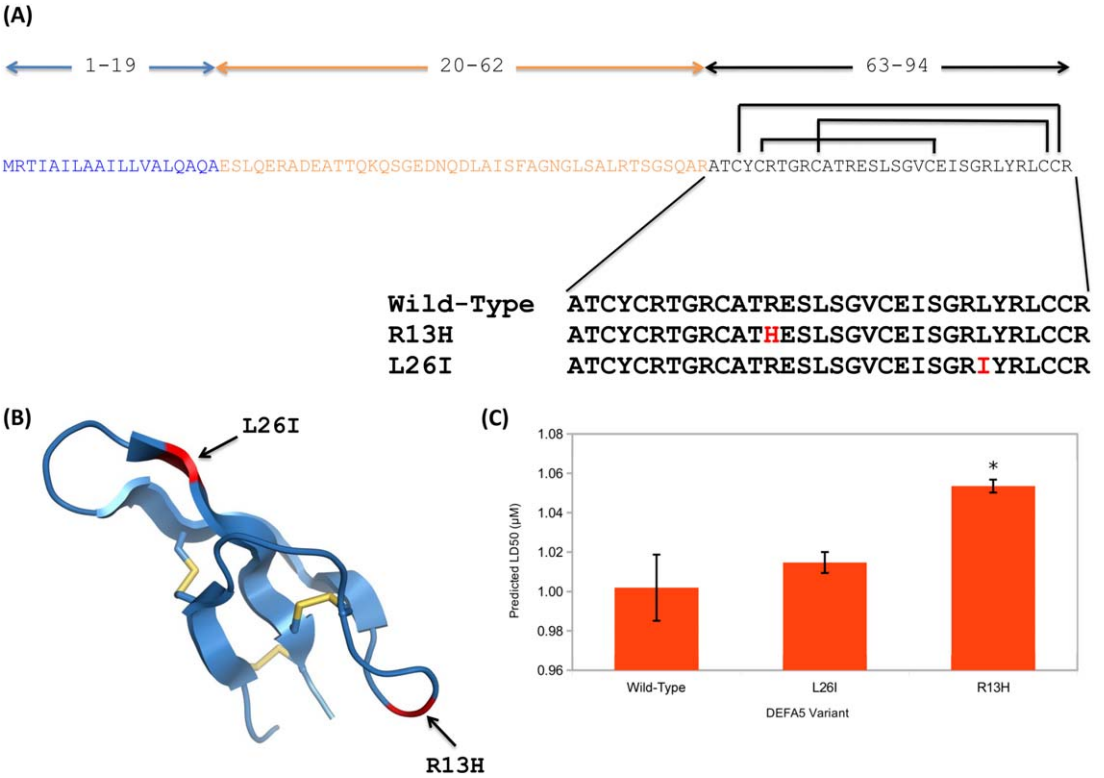

FIGURE 3 HD5 precursor and its variants. A: The positions of the signal peptide sequence (blue), pro-peptide (brown + black), mature peptide (black) and cysteine residues (bold) are shown above the primary sequence. The three disulfide bridges (shown as brackets) are formed at cysteines 65-92, 67-78 and 72-91. Each of the amino acid residue changes caused by missense SNPs mapped to these regions is highlighted in red in its respective position. The amino acid positions are relative to GenBank Accession number NP_066290.1. B: The three-dimensional structure of HD5 (PDB ID: 1ZMP) and the amino acid residue changes caused by missense SNPs are mapped into the structure. C: Predicted LD_{50} values for HD5 variants, the star indicates a statistical difference between the predicted LD_{50} from wild-type and the variant ($P < 0.05$). The amino acid positions of precursor HD5 are relative to GenBank Accession number NP_066290.1.

Table IV Summary of Charge and the Molecular Dynamics Results of HBD1 and Its Variants

Defensin	Variant[b]	Net Charge	RMSD (Å)[a]			TM-Score[a]		
			Simulation 1	Simulation 2	Simulation 3	Simulation 1	Simulation 2	Simulation 3
HD5	Wild-Type	4	2.17	1.37	2.37	0.590	0.593	0.531
	L26I	4	2.32	2.41	2.16	0.509	0.526	0.593
	R13H	4	2.58	2.64	1.50	0.514	0.529	0.615
HBD1	Wild-Type	5	2.49	2.28	3.18	0.549	0.422	0.508
	C12Y	5	3.66	3.17	3.13	0.483	0.419	0.493
	C35S	5	3.63	3.67	2.02	0.449	0.414	0.528
	K31T	4	2.74	3.23	4.25	0.566	0.452	0.485
	K33R	5	3.95	4.69	3.43	0.408	0.504	0.386
	R29G	4	2.21	2.26	3.13	0.548	0.546	0.406
	V06I	5	3.68	2.83	2.78	0.392	0.482	0.501
	Y28[a]	1	5.90	6.60	5.20	0.183	0.160	0.173
	C05[a]	0	n/a	n/a	n/a	n/a	n/a	n/a

[a] Data generated by comparing the structure at 0ns and 100ns.
[b] The amino acid positions of HD5 and HBD1are relative to PDB IDs 1ZMP and 1KJ5.

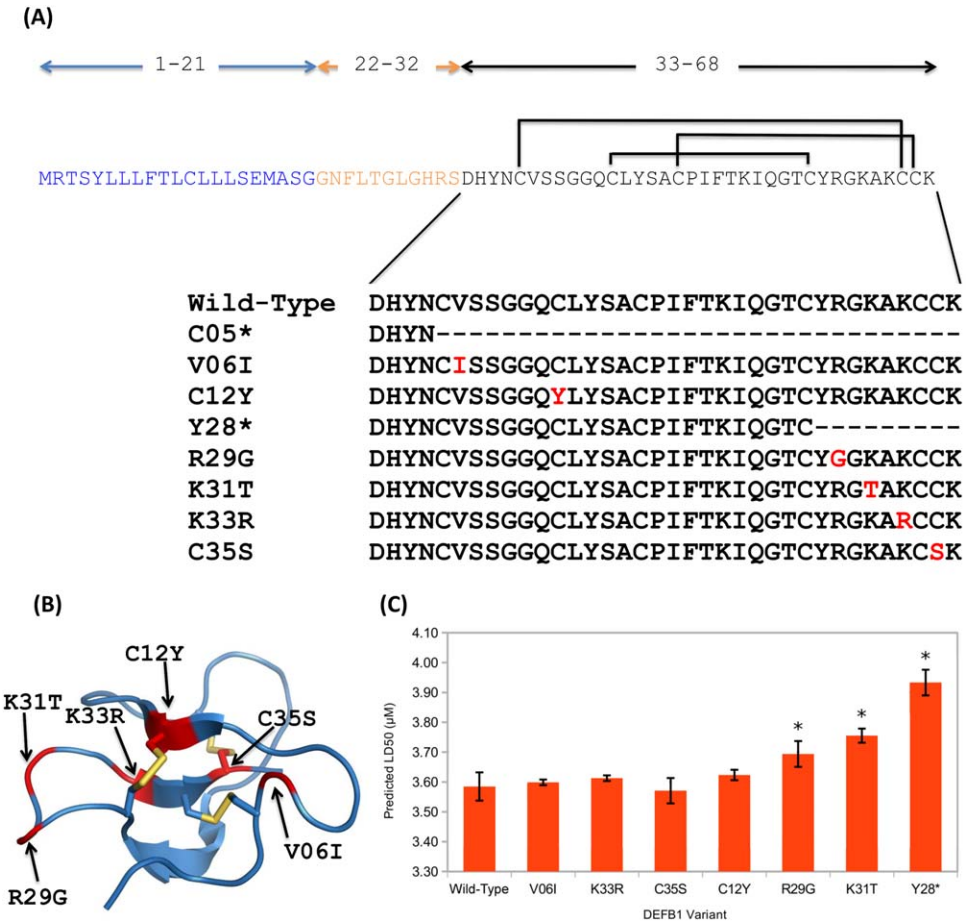

FIGURE 4 HBD1 precursor and its variants. A: The positions of the signal peptide sequence (blue), pro-peptide (brown + black), mature peptide (black) and cysteine residues (bold) are shown above the primary sequence. The three disulfide bridges (shown as brackets) are formed at cysteines 37–66, 44–59, and 49–67. Each of the amino acid residue changes caused by missense SNPs mapped to these regions is highlighted in red in its respective position. The amino acid positions are relative to GenBank Accession number NP_005209.1. B: The three-dimensional structure of HBD1 (PDB ID: 1KJ5) with the amino acid residue changes caused by missense SNPs are mapped into the structure. C: Predicted LD_{50} values for HBD1 variants, the star indicates a statistical difference between the predicted LD_{50} from wild-type and the variant ($P < 0.05$). The amino acid positions of precursor HBD1 are relative to GenBank Accession number NP_005209.1.

DISCUSSION

The explosive growth in database data requires techniques and tools to transform this huge amount of data into useful information and knowledge. Thus, knowledge discovery in databases has been recognized as a hot spot in different areas.[75,76] Data mining processes have been carried out to discover novel biological data such as candidates for functional prediction,[76] novel translatable sequences[77] and also antimicrobial peptides.[2,78–81] In this work, information on four distinct databases (NCBI Gene, PDB, dbSNP and 1000 Genomes Project) was crossed for selecting missense SNPs with population frequencies which lead to alterations into the mature product of human defensin genes, and also two linear regression models were used for evaluation of the effect of such modifications on antibacterial activity.

Only two defensin genes meet the selection criteria: *DEFA5* and *DEFB1* (Figure 1). The distribution of all SNPs shows that *DEFB1* has a lower percentage of SNPs in regulatory and coding regions than *DEFA5* (Figure 2). These data could be an indication that *DEFB1* is under higher selective pressure than *DEFA5*, as SNPs occurring in those regions are more likely to have an impact on protein function. However, considering only SNPs that alter the mature defensin product, the context is reversed, since the majority of missense SNPs were found in

DEFB1 (Table I). Seven out of nine rare SNPs identified in *DEFA5* and *DEFB1* genes occurred in only one of the four populations analyzed (Table I). Considering SNPs with global minor allele frequency $> 1\%$ as common and $< 1\%$ as rare, it has been shown recently that the majority of human genomic variable sites are rare and exhibit little sharing among diverged populations.[82]

Interestingly, it was observed that, in the mature defensin product, the majority of SNPs is restricted to single populations (Table I). It has been previously shown that five beta-defensin genes, including *DEFB1*, have undergone positive selection in primates.[83] Also, the positions in the gene where variations may occur have been shown to be directly involved in antimicrobial activity and signaling, and charge and polarity have been modulated by natural selection in response to the microbial environment.[11] However, the nature of the selective pressure remains unclear and is likely to differ between populations.[83]

Rare SNPs with large functional effects could contribute significantly to the genetic architecture of diseases.[84] Up to date, there are several methods for predicting the effect of one amino acid substitution in the protein function. In a previous study, the concept of convergent deleterious prediction was drawn, where the variation is considered deleterious if it is predicted in three different programs in four classes of predictors.[33] Here, applying such approach, it was demonstrated that only two variants from HDB1 (C12Y and C35S) could be considered convergent deleterious (Table III). However, such prediction methods could not be the most adequate way to evaluate a defensin variant, since previous reports have indicated that variations of one amino acid do not cause structural changes in overall defensin fold.[11,30]

On the other hand, amino acid variations in defensins could change some structural properties such as radius of gyration, solvent accessible surface area and flexibility, but such changes do not correlate with antibacterial activity.[39] These amino acid variations change the solvation potential energy, which, on the other hand, is correlated with antibacterial activity against *E. coli*.[39] Therefore, the solvation potential energy could be used for predicting such activity of variants driven by SNPs.

For HD5, the R13H variant showed a slightly higher predicted LD_{50} in comparison to wild-type defensin, however without biological relevance. In a previous study,[20] the R13H variant was observed in a Crohn's disease patient and its activity against *E. coli* was very similar to wild-type HD5, but its activity against *Staphylococcus aureus*, *Enterobacter aerogenes*, and *Bacilus cereus* was reduced.[20]

For HBD1, the R29G and K31T variants also showed higher predicted LD_{50} in comparison to HBD1 wild-type. Despite the absence of literature data that associate these SNPs with human diseases, their antibacterial activity could really differ from the wild-type, since these variations have a net charge of $+4$, while the native wild-type HBD1 has a net charge of $+5$ (Table IV). The charge reduction in β-defensins could lead to antibacterial activity depletion. Pazgier *et al.*[11] tested the antibacterial activity of several HBD1 variants and found that some that have reduced charge are essentially inactive against *E. coli*.

Interestingly, the HBD1 C12Y and C35S variants induce a disulfide bridge break, but their predicted LD_{50} is similar to wild-type HBD1. In a previous report, antibacterial activity of the C35S variant was characterized against *Pseudomonas aeruginosa* and clinical isolates of a methicillin-resistant *Staphylococcus aureus*, *Cryptococcus neoformans*, and *Candida albicans*. This variant was more active against these microorganisms than the native wild-type HBD1.[85] In addition, its structure was evaluated by circular dichroism and this indicated that the structure is very similar to wild-type. This corroborates with the molecular dynamics simulation data and also the prediction model, since the C35S variant showed a lower value of predicted LD_{50}, although it is not statistically significant (Figure 4). In addition, it has been shown that the reduction of all disulfide bridges from HBD1 does not alter its antibacterial activity.[86] On the other hand, the Y28* variant has a statistically different predicted LD_{50}, despite this variation having two disulfide bridges broken; however, its charge is reduced to $+1$, since it lacks nine C-Terminal residues. In fact, in a previous study,[86] the depletion of seven C-Terminal residues is sufficient to abolish the HBD1 antibacterial activity, corroborating with the data from the Y28* variant.

Although the prediction models have successfully shown the differences between the HBD1 Y28* variant to its wild-type structure, and also the similarity between HD5 R13H and HBD1 C35S variants to their wild-type structures, their associations to human diseases and antibacterial activity are unknown. Therefore, the predictions must be carefully analyzed when associating them to human diseases, mainly because there are other defensin genes that can cover the role of a less active defensin variant. Interestingly, even if the activity of a defensin variant has no statistical difference to the wild-type one, such variant could be associated to human diseases, as observed for HDB1 V06I (rs2738047:C > T). Recently, it was also shown that the frequency of the HBD1 V06I variant was noticeably higher in African sub-Saharan and African-American HIV patients, suggesting that this SNP is distributed differently among different population groups, although further statistical significance tests are required.[22] In addition, this same variant was significantly associated with chronic obstructive pulmonary disease.[87] Together with LD_{50} prediction, these

data indicates that V06I differs from the wild-type in some of the other multiple activities.

In conclusion, the data here presented indicate that at least one and four SNPs for *DEFA5* and *DEFB1*, respectively, could lead to less active defensins, and the data generated here for HD5 and HBD1 could be helpful for novel research. An in depth knowledge of the functional and structural impact of these SNPs may be essential to reduce the number of variations to be screened in genetic association studies, and data here reported could lead to a better understanding of functional aspects of human defensins.

The authors are grateful to the Center for Scientific Computing (NCC/GridUNESP) of São Paulo State University (UNESP) and Dr. T. Joachims, from the Department of Computer Science of Cornell University, USA.

REFERENCES

1. Ganz, T. C R Biol 2004, 327, 539–549.
2. Porto, W. F.; Fensterseifer, G. M.; Franco, O. L. J Mol Model 2014, 20, 2339.
3. Mygind, P. H.; Fischer, R. L.; Schnorr, K. M.; Hansen, M. T.; Sönksen, C. P.; Ludvigsen, S.; Raventós, D.; Buskov, S.; Christensen, B.; De Maria, L.; Taboureau, O.; Yaver, D.; Elvig-Jørgensen, S. G.; Sørensen, M. V.; Christensen, B. E.; Kjaerulff, S.; Frimodt-Moller, N.; Lehrer, R. I.; Zasloff, M.; Kristensen, H. H. Nature 2005, 437, 975–980.
4. Cândido, E. S.; Porto, W. F.; Amaro, D. S.; Viana, J. C.; Dias, S. C.; Franco, O. L. In Science Against Microbial Pathogens: Communicating Current Research and Technological Advances; Méndez-Vilas, A., Ed.; Formatex, Extremadura, Spain, 2011; pp 951–960.
5. Klotman, M. E.; Chang, T. L. Nat Rev Immunol 2006, 6, 447–456.
6. Bensch, K. W.; Raida, M.; Mägert, H. J.; Schulz-Knappe, P.; Forssmann, W. G. FEBS Lett 1995, 368, 331–335.
7. Sass, V.; Schneider, T.; Wilmes, M.; Körner, C.; Tossi, A.; Novikova, N.; Shamova, O.; Sahl, H. G. Infect Immun 2010, 78, 2793–2800.
8. Yang, D.; Chertov, O.; Bykovskaia, S. N.; Chen, Q.; Buffo, M. J.; Shogan, J.; Anderson, M.; Schröder, J. M.; Wang, J. M.; Howard, O. M.; Oppenheim, J. J. Science 1999, 286, 525–528.
9. Chileveru, H. R.; Lim, S. A.; Chairatana, P.; Wommack, A. J.; Chiang, I. L.; Nolan, E. M. Biochemistry 2015, 54, 1767–1777.
10. Mathew, B.; Nagaraj, R. Peptides 2015, 71, 128–140.
11. Pazgier, M.; Prahl, A.; Hoover, D. M.; Lubkowski, J. J Biol Chem 2007, 282, 1819–1829.
12. Chen, C.; Yadav, P. K.; Wang, X.; Liu, Z. Open J Immunol 2012, 02, 78–84.
13. Levy, H.; Raby, B. A.; Lake, S.; Tantisira, K. G.; Kwiatkowski, D.; Lazarus, R.; Silverman, E. K.; Richter, B.; Klimecki, W. T.; Vercelli, D.; Martinez, F. D.; Weiss, S. T. J Allergy Clin Immunol 2005, 115, 252–258.
14. Leung, T. F.; Li, C. Y.; Liu, E. K. H.; Tang, N. L. S.; Chan, I. H. S.; Yung, E.; Wong, G. W. K.; Lam, C. W. K. Genes Immun 2006, 7, 59–64.
15. Sandrin-Garcia, P.; Brandão, L. A. C.; Guimarães, R. L.; Pancoto, J. A. T.; Donadi, E. A.; Lima-Filho, J. L.; de.,; Segat, L Crovella, S. Lupus 2012, 21, 625–631.
16. Kim, E.; Lee, J. E.; Namkung, J. H.; Kim, P. S.; Kim, S.; Shin, E. S.; Cho, E. Y.; Yang, J. M. J Dermatol Sci 2009, 54, 25–30.
17. Prado-Montes de Oca, E.; García-Vargas, A.; Lozano-Inocencio, R.; Gallegos-Arreola, M. P.; Sandoval-Ramírez, L.; Dávalos-Rodríguez, N. O.; Figuera, L. E. Int Arch Allergy Immunol 2007, 142, 211–218.
18. Hollox, E. J.; Huffmeier, U.; Zeeuwen, P. L. J. M.; Palla, R.; Lascorz, J.; Rodijk-Olthuis, D.; van de Kerkhof, P. C. M.; Traupe, H.; de Jongh, G.; den Heijer, M.; Reis, A.; Armour, J. A. L.; Schalkwijk, J. Nat Genet 2008, 40, 23–25.
19. Underwood, M. A.; Bevins, C. L. Pediatrics 2010, 125, 1237–1247.
20. de Leeuw, E.; Rajabi, M.; Zou, G.; Pazgier, M.; Lu, W. FEBS Lett 2009, 583, 2507–2512.
21. Braida, L.; Boniotto, M.; Pontillo, A.; Tovo, P. A.; Amoroso, A.; Crovella, S. Aids 2004, 18, 1598–1600.
22. Mehlotra, R. K.; Zimmerman, P. A.; Weinberg, A.; Jurevic, R. J. Int J Immunogenet 2013, 40, 261–269.
23. Diao, R.; Fok, K. L.; Chen, H.; Yu, M. K.; Duan, Y.; Chung, C. M.; Li, Z.; Wu, H.; Li, Z.; Zhang, H.; Ji, Z.; Zhen, W.; Ng, C. F.; Gui, Y.; Cai, Z.; Chan, H. C. Sci Transl Med 2014, 6, 249ra108.
24. Jones, D. E.; Bevins, C. L. J Biol Chem 1992, 267, 23216–23225.
25. Quayle, A. J.; Porter, E. M.; Nussbaum, A. A.; Wang, Y. M.; Brabec, C.; Yip, K. P.; Mok, S. C. Am J Pathol 1998, 152, 1247–1258.
26. Frye, M.; Bargon, J.; Dauletbaev, N.; Weber, A.; Wagner, T. O.; Gropp, R. J Clin Pathol 2000, 53, 770–773.
27. Hiratsuka, T.; Nakazato, M.; Ihi, T.; Minematsu, T.; Chino, N.; Nakanishi, T.; Shimizu, A.; Kangawa, K.; Matsukura, S. Nephron 2000, 85, 34–40.
28. Ericksen, B.; Wu, Z.; Lu, W.; Lehrer, R. I. Antimicrob Agents Chemother 2005, 49, 269–275.
29. Goldman, M. J.; Anderson, G. M.; Stolzenberg, E. D.; Kari, U. P.; Zasloff, M.; Wilson, J. M. Cell 1997, 88, 553–560.
30. Rajabi, M.; Ericksen, B.; Wu, X.; de Leeuw, E.; Zhao, L.; Pazgier, M.; Lu, W. J Biol Chem 2012, 287, 21615–21627.
31. Abecasis, G. R.; Auton, A.; Brooks, L. D.; DePristo, M. A.; Durbin, R. M.; Handsaker, R. E.; Kang, H. M.; Marth, G. T.; McVean, G. A. Nature 2012, 491, 56–65.
32. Kryukov, G. V.; Pennacchio, L. A.; Sunyaev, S. R. Am J Hum Genet 2007, 80, 727–739.
33. Porto, W. F.; Franco, O. L.; Alencar, S. A. Peptides 2015, 69, 92–102.
34. Kamaraj, B.; Purohit, R. Cell Biochem Biophys 2013, 68, 97–109.
35. Kumar, A.; Rajendran, V.; Sethumadhavan, R.; Purohit, R. PLoS One 2013, 8, e77453.
36. Liu, M.; Wang, L.; Sun, X.; Zhao, X. Sci Rep 2014, 4, 5095.
37. Jia, M.; Yang, B.; Li, Z.; Shen, H.; Song, X.; Gu, W. PLoS One 2014, 9, e104311.
38. Padhi, A. K.; Jayaram, B.; Gomes, J. Sci Rep 2013, 3, 1225.
39. Porto, W. F.; Nolasco, D. O.; Pires, ÁS.; Fernandes, G. R.; Franco, O. L.; Alencar, S. A. Biopolymers 2016, 106, 43–50.

40. Do, R.; Kathiresan, S.; Abecasis, G. R. Hum Mol Genet 2012, 21, R1–R9.
41. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C. A.; Banks, E.; DePristo, M. A.; Handsaker, R. E.; Lunter, G.; Marth, G. T.; Sherry, S. T.; McVean, G.; Durbin, R. Bioinformatics 2011, 27, 2156–2158.
42. Weir, B. S.; Cockerham, C. C. Evolution (N. Y) 1984, 38, 1358–1370.
43. Szyk, A.; Wu, Z.; Tucker, K.; Yang, D.; Lu, W.; Lubkowski, J. Prot Sci 2006, 15, 2749–2760.
44. Schibli, D. J.; Hunter, H. N.; Aseyev, V.; Starner, T. D.; Wiencek, J. M.; McCray, P. B.; Tack, B. F.; Vogel, H. J. J Biol Chem 2002, 277, 8279–8289.
45. Kumar, P.; Henikoff, S.; Ng, P. C. Nat Protoc 2009, 4, 1073–1081.
46. Choi, Y.; Sims, G. E.; Murphy, S.; Miller, J. R.; Chan, A. P. PLoS One 2012, 7, e46688.
47. Reva, B.; Antipin, Y.; Sander, C. Nucleic Acids Res 2011, 39, e118.
48. 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. Nucleic Acids Res 2005, 33, D284–D288. (Database issue)
49. Bromberg, Y.; Yachdav, G.; Rost, B. Bioinformatics 2008, 24, 2397–2398.
50. Capriotti, E.; Calabrese, R.; Casadio, R. Bioinformatics 2006, 22, 2729–2734.
51. Yates, C. M.; Filippis, I.; Kelley, L. A.; Sternberg, M. J. E. J Mol Biol 2014, 426, 2692–2701.
52. Capriotti, E.; Calabrese, R.; Fariselli, P.; Martelli, P. L.; Altman, R. B.; Casadio, R. BMC Genomics 2013, 14 (Suppl 3), S6.
53. Adzhubei, I. A.; Schmidt, S.; Peshkin, L.; Ramensky, V. E.; Gerasimova, A.; Bork, P.; Kondrashov, A. S.; Sunyaev, S. R. Nat Methods 2010, 7, 248–249.
54. Worth, C. L.; Preissner, R.; Blundell, T. L. Nucleic Acids Res 2011, 39, W215–W222. (Web Server issue)
55. Gonnelli, G.; Rooman, M.; Dehouck, Y. J Biotechnol 2012, 161, 287–293.
56. Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. BMC Bioinformatics 2011, 12, 151.
57. González-Pérez, A.; López-Bigas, N. Am J Hum Genet 2011, 88, 440–449.
58. Capriotti, E.; Altman, R. B.; Bromberg, Y. BMC Genomics 2013, 14, S2.
59. Niroula, A.; Urolagin, S.; Vihinen, M. PLoS One 2015, 10, e0117380.
60. Bendl, J.; Stourac, J.; Salanda, O.; Pavelka, A.; Wieben, E. D.; Zendulka, J.; Brezovsky, J.; Damborsky, J. PLoS Comput Biol 2014, 10, e1003440.
61. Webb, B.; Sali, A. Curr Protoc Bioinformatics 2014, 47, 561–565.6.32.
62. Wiederstein, M.; Sippl, M. J. Nucleic Acids Res 2007, 35, W407–W410. (Web Server issue)
63. Laskowski, R.; Macarthur, M.; Moss, D.; Thornton, J. J Appl Cryst 1993, 26, 283–291.
64. Berendsen, H.; Postma, J.; van Gunsteren, W.; Hermans, J. In Intermolecular Forces; Pullman, B., Ed.; Springer, 1981; pp 331–342.
65. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. J Chem Theory Comput 2008, 4, 435–447.
66. de Souza, O. N.; Ornstein, R. L. Biophys J 1997, 72, 2395–2397.
67. Miyamoto, S.; Kollman, P. A. J Comput Chem 1992, 13, 952–962.
68. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J Comput Chem 1997, 18, 1463–1472.
69. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089.
70. Zhang, Y.; Skolnick, J. Proteins 2004, 57, 702–710.
71. Xu, J.; Zhang, Y. Bioinformatics 2010, 26, 889–895.
72. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. Nucleic Acids Res 2004, 32, W665–W667. (Web Server issue)
73. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Proc Natl Acad Sci USA 2001, 98, 10037–10041.
74. Zhang, Y.; Doherty, T.; Li, J.; Lu, W.; Barinka, C.; Lubkowski, J.; Hong, M. J Mol Biol 2010, 397, 408–422.
75. Chen, M. S.; Han, J.; Yu, P. S. IEEE Trans Knowl Data Eng 1996, 8, 866–883.
76. Porto, W. F.; Maria-Neto, S.; Nolasco, D. O.; Franco, O. L. J. Proteom Bioinform 2014, 07, 203–213.
77. Desler, C.; Suravajhala, P.; Sanderhoff, M.; Rasmussen, M.; Rasmussen, L. J. BMC Bioinformatics 2009, 10, 289.
78. Zhu, S. Mol Immunol 2008, 45, 828–838.
79. Fernandes, F. C.; Porto, W. F.; Franco, O. L. In Lecture Notes in Computer Science; Guimarães, K. S., Panchenko, A., Przytycka, T. M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 5676 .
80. Porto, W. F.; Souza, V. A.; Nolasco, D. O.; Franco, O. L. Peptides 2012, 38, 127–136.
81. Cândido, E.; de, S.; Fernandes, G.; da, R.; de Alencar, S. A.; Cardoso, M. H e S.; Lima, S. M.; de, F.; Miranda, V.; de, J.; Porto, W. F.; Nolasco, D. O.; de Oliveira-Júnior, N. G.; Barbosa, A. E. A.; de, D.; Pogue, R. E.; Rezende, T. M. B.; Dias, S. C.; Franco, O. L. PLoS One 2014, 9, e90487.
82. Gravel, S.; Henn, B. M.; Gutenkunst, R. N.; Indap, A. R.; Marth, G. T.; Clark, A. G.; Yu, F.; Gibbs, R. A.; Bustamante, C. D. Proc Natl Acad Sci U S A 2011, 108, 11983–11988.
83. Hollox, E. J.; Armour, J. A. L. BMC Evol Biol 2008, 8, 113.
84. Manolio, T. A.; Collins, F. S.; Cox, N. J.; Goldstein, D. B.; Hindorff, L. A.; Hunter, D. J.; McCarthy, M. I.; Ramos, E. M.; Cardon, L. R.; Chakravarti, A.; Cho, J. H.; Guttmacher, A. E.; Kong, A.; Kruglyak, L.; Mardis, E.; Rotimi, C. N.; Slatkin, M.; Valle, D.; Whittemore, A. S.; Boehnke, M.; Clark, A. G.; Eichler, E. E.; Gibson, G.; Haines, J. L.; Mackay, T. F. C.; McCarroll, S. A.; Visscher, P. M. Nature 2009, 461, 747–753.
85. Circo, R.; Skerlavaj, B.; Gennaro, R.; Amoroso, A.; Zanetti, M. Biochem Biophys Res Commun 2002, 293, 586–592.
86. Schroeder, B. O.; Wu, Z.; Nuding, S.; Groscurth, S.; Marcinowski, M.; Beisner, J.; Buchner, J.; Schaller, M.; Stange, E. F.; Wehkamp, J. Nature 2011, 469, 419–423.
87. Matsushita, I.; Hasegawa, K.; Nakata, K.; Yasuda, K.; Tokunaga, K.; Keicho, N. Biochem Biophys Res Commun 2002, 291, 17–22.