

## Research Article

# A Comprehensive *In Silico* Analysis of the Functional and Structural Impact of SNPs in the *IGF1R* Gene

S. A. de Alencar<sup>1,2</sup> and Julio C. D. Lopes<sup>2</sup>

<sup>1</sup> Departamento de Bioquímica e Imunologia, Bioinformática, Universidade Federal de Minas Gerais, Av. Antonio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil

<sup>2</sup> Chemoinformatics Group, NEQUIM, Departamento de Química, Universidade Federal de Minas Gerais, Av. Antonio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil

Correspondence should be addressed to S. A. de Alencar, [sergiodealencar@gmail.com](mailto:sergiodealencar@gmail.com) and Julio C. D. Lopes, [jlopes.ufmg@gmail.com](mailto:jlopes.ufmg@gmail.com)

Received 1 February 2010; Accepted 28 April 2010

Academic Editor: Ravindra N. Chibbar

Copyright © 2010 S. A. de Alencar and J. C. D. Lopes. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Insulin-like growth factor 1 receptor (IGF1R) acts as a critical mediator of cell proliferation and survival. Many single nucleotide polymorphisms (SNPs) found in the *IGF1R* gene have been associated with various diseases, including both breast and prostate cancer. The genetics of these diseases could be better understood by knowing the functions of these SNPs. In this study, we performed a comprehensive analysis of the functional and structural impact of all known SNPs in this gene using publicly available computational prediction tools. Out of a total of 2412 SNPs in *IGF1R* retrieved from dbSNP, we found 32 nsSNPs, 58 sSNPs, 83 mRNA 3' UTR SNPs, and 2225 intronic SNPs. Among the nsSNPs, a total of six missense nsSNPs were found to be damaging by both a sequence homology-based tool (SIFT) and a structural homology-based method (PolyPhen), and one nonsense nsSNP was found. Further, we modeled mutant proteins and compared the total energy values with the native IGF1R protein, and showed that a mutation from arginine to cysteine at position 1216 (rs61740868) on the surface of the protein caused the greatest impact on stability. Also, the FASTSNP tool suggested that 31 sSNPs and 3 intronic SNPs might affect splicing regulation. Based on our investigation, we report potential candidate SNPs for future studies on *IGF1R* mutations.

## 1. Introduction

Single nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome is altered. SNPs make up about 90% of all human genetic variation, occurring every 100–300 bases along the 3-billion-base human genome, although their density vary between regions [1]. SNPs are found in both coding (gene) and noncoding regions of the genome. Many SNPs have no effect on cell function; however, others could predispose people to disease or influence their response to a drug. Nonsynonymous SNPs (nsSNPs) that lead to an amino acid residue substitution in the protein product are of particular interest because they are responsible for nearly half of the known genetic variations related to human inherited disease [2]. Coding synonymous SNPs (sSNPs) and SNPs occurring outside gene promoter or coding regions

may nevertheless still have consequences for gene expression, splicing, or transcription-factor binding [3, 4].

The identification of SNPs responsible for specific phenotypes appears to be a problem that is very difficult to solve, requiring multiple testing of hundreds or thousands of SNPs in candidate genes [5]. However, the question of how to choose the set of SNPs to be screened is critical to the success of association studies. A possible way to overcome this problem would be to prioritize SNPs according to their functional significance [6, 7] by using Bioinformatics prediction tools, which may help discriminate neutral SNPs from SNPs of likely functional importance and could also be useful to reveal the structural basis of disease mutations. Without any careful preselection of SNPs to be screened, a huge number of individuals might be required to detect association at a reasonable level of statistical significance [5].

Although wetlab-based approaches used to identify disease-associated SNPs from a large number of neutral SNPs remain crucial evidence for the functional role of SNPs [8], numerous disease associations published could not be confirmed by subsequent independent studies [6, 9]. Hence, independent evidence of functionality of SNPs obtained by using prediction tools could also serve as additional argument to discriminate true associations from false positives [5], as shown recently by the functional SNP analysis of the *BRCA1*, *ABL1*, *ERBB2*, *CFTR*, and *EGFR* genes [10–14].

Insulin-like growth factor 1 receptor (IGF1R) is a growth factor receptor tyrosine kinase that acts as a critical mediator of cell proliferation and survival. This receptor is implicated in several cancers, including both breast and prostate cancer [15, 16]. Evidence suggests that IGF1R signaling is required for survival and growth when prostate cancer cells progress to androgen independence [17], as increased levels of the receptor are expressed in the majority of primary and metastatic prostate cancer patient tumors [18]. There have also been studies showing associations of *IGF1R* polymorphisms in dementia and ischemic stroke [19, 20].

Although there are presently several articles describing the association of SNPs in the *IGF1R* gene with different types of diseases, computational analysis has not yet been undertaken on the functional consequences of SNPs in this gene. We applied different publicly available computational algorithms, namely, Sorting Intolerant From Tolerant (SIFT) [21], Polymorphism Phenotyping (PolyPhen) [22], and Function Analysis and selection tool for single nucleotide polymorphisms (FASTSNP) to identify likely deleterious SNPs which could affect protein function [23].

The SIFT algorithm predicts whether an amino acid substitution affects protein function based on sequence homology among related genes and domains over evolutionary time, and the physical-chemical properties of the amino acid residues [24–26]. Sequence conservation and the nature of the amino acid residues involved are also incorporated by PolyPhen, but it also values the location of the substitution within known structures and structural features of the protein available in the annotated database SwissProt [5, 27]. By accessing a variety of heterogeneous biological databases and analytical tools, FASTSNP is able to identify SNPs most likely to have functional effects, such as changes to the transcriptional level and pre-mRNA splicing [23].

SIFT and PolyPhen were approximately 80% successful in benchmarking studies employing amino acid substitutions assumed to have a major negative impact on the residual activity of the variant protein as the test set [22, 25, 27–29] and it has been estimated that the “false negative” and “false positive” error rates of SIFT is 31% and 20%, and 31% and 9% for PolyPhen [26]. FASTSNP was used to analyze 1569 SNPs from the SNP500 cancer database, and results showed that SNPs with a high predicted risk exhibited low allele frequencies for the minor alleles, which is consistent with the finding that a strong selective pressure exist for functional polymorphisms [23, 30].

As the majority of disease mutations affect protein stability [31, 32], we also proposed modeled protein structures for the mutant proteins and compared them with the native protein in order to evaluate stability changes.

## 2. Materials and Methods

**2.1. Evaluation of the Functional Impact of Coding nsSNPs Using a Sequence Homology Tool (SIFT).** SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence (<http://sift.jcvi.org>) [21]. It is a multistep procedure that, given a protein sequence, (1) searches for similar sequences, (2) chooses closely related sequences that may share similar function, (3) obtains the multiple alignment of these chosen sequences, and (4) calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a tolerance index of 0.05 are predicted to be intolerant or deleterious; those greater than or equal to 0.05 are predicted to be tolerated [24, 26].

The analysis was performed by allowing the algorithm to search for homologous sequences using the default settings (UniProt-TrEMBL 39.6 database, median conservation of sequences of 3.00, and allowance to remove sequences more than 90% identical to query sequence). The IGF1R FASTA amino acid sequence of the NCBI Protein accession id NP\_000866.1 was used as the query sequence, and a total of 24 *IGF1R* nsSNPs filtered from the dbSNP database were analyzed.

**2.2. Evaluation of the Functional Impact of Coding nsSNPs Using a Structural Homology-Based Method (PolyPhen).** PolyPhen prediction is based on straightforward empirical rules which are applied to the sequence, phylogenetic and structural information characterizing the substitution [5]. The online input form available at <http://coot.embl.de/PolyPhen> was filled with the IGF1R amino acid sequence in FASTA format (NCBI Protein accession id NP\_000866.1), and the position and substitution of each of the 24 nsSNPs analyzed by SIFT were also submitted for PolyPhen analysis. PolyPhen then searched for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases, calculated position-specific independent counts (PSIC) scores for each of the two amino acid residues entered (the original residue and the nsSNP), and then computed the PSIC scores difference of the two residues. The higher a PSIC score difference, the higher functional impact a particular amino acid substitution is likely to have. A PSIC score difference of 1.5 and above is considered to be damaging. The query options were left with default values.

**2.3. Functional Significance of SNPs in Regulatory Regions.** The online tool FASTSNP [23] was used to determine the impact of the sSNPs, 3' UTR regions SNPs and intronic

TABLE 1: List of nsSNPs that were analysed by SIFT and PolyPhen.

dbSNP ID	Alleles	AA change	Tolerance index	PSIC	Heterozygosity	Validation
rs70958401	C/T	Arg/Trp	0.18	<b>1.892</b>	0.039	
rs70958396	G/A	Ala/Thr	0.41	0.011	0.039	
rs61740877	G/A	Val/Ile	0.77	0.019	n/a	
rs61740868	C/T	Arg/Cys	<b>0.00</b>	<b>2.609</b>	n/a	1
rs61731172	G/A	Arg/Gln	0.74	0.137	n/a	
rs56248469	G/A	Arg/His	0.57	0.613	n/a	
rs45611935	A/G	Asn/Ser	0.77	0.387	n/a	
rs45597432	T/C	Ile/Thr	0.96	0.079	n/a	
rs45578132	T/C	Val/Ala	<b>0.00</b>	<b>2.027</b>	n/a	
rs45553041	G/A	Arg/His	<b>0.00</b>	<b>2.196</b>	0.012	
rs45526336	G/A	Glu/Lys	<b>0.00</b>	1.470	n/a	
rs45524940	A/G	Thr/Ala	<b>0.01</b>	<b>2.296</b>	n/a	
rs45522834	C/T	Thr/Ile	0.29	1.220	n/a	
rs45512296	G/A	Arg/His	<b>0.01</b>	<b>2.128</b>	n/a	
rs45504297	T/C	Leu/Pro	<b>0.00</b>	<b>2.372</b>	n/a	
rs45493995	G/T	Ser/Ile	0.30	0.400	n/a	
rs45475702	G/A	Val/Ile	0.55	0.296	n/a	
rs45451896	G/T	Arg/Leu	0.25	0.305	n/a	
rs45445894	G/A	Val/Met	<b>0.03</b>	0.947	0.011	
rs35224135	G/A	Ala/Thr	0.31	1.026	0.005	2
rs34516635	G/A	Arg/His	1.00	1.339	0.005	1; 2
rs34102392	G/A	Ala/Thr	0.17	1.097	n/a	
rs34061581	A/G	His/Arg	0.25	1.346	0.005	2
rs33958176	G/A	Arg/Gln	0.59	<b>1.503</b>	n/a	1; 2

Prediction scores found to be functionally significant by SIFT and PolyPhen are shown in bold. Validation Status Description: (1) validated by multiple, independent submissions to the refSNP cluster; (2) validated by frequency or genotype data: minor alleles observed in at least two chromosomes.

SNPs on the regulation of the *IGF1R* gene. The FAST-SNP server (<http://FASTSNP.ibms.sinica.edu.tw>) follows the decision tree principle with external Web service access to TFSearch, which predicts whether a non-coding SNP alters the transcription factor binding site of a gene. The score is given on the basis of levels of risk with a ranking of 0, 1, 2, 3, 4, or 5. This signifies the levels of no, very low, low, medium, high, and very high effect, respectively.

**2.4. Modeling of nsSNPs on Protein Structures and Calculation of their RMSD Difference.** Structural analysis was performed in order to evaluate and compare the stability of native and mutant structures. Information about mapping the nsSNPs in the protein structure was obtained from dbSNP [33]. The highest resolution (2.00 Å) native structure of the IGF1R protein available in the Protein Data Bank (PDB) [34] has an id of 2oj9 [35]. The positions of the studied nsSNPs mutations on PDBid 2oj9 were confirmed by pairwise alignment between the FASTA amino acid sequence of the IGF1R protein obtained from the NCBI (NP\_000866.1) and the 2oj9 FASTA amino acid sequence, using the Sequence Manipulation Suite [36]. The amino acid residue substitutions were performed using the Swiss-Pdb Viewer [37], followed by energy minimization of the modeled 3D structures using the GROMACS software version 4.0 [38]. The algorithms used for energy minimization

were the steepest descent (1000 steps), followed by conjugate gradient (1500 steps) alternating with the steepest descent every 100 steps. The comparison between the resulting native and modeled structures was made by the calculation of the potential energy and RMSD values.

### 3. Results and Discussion

**3.1. SNP Dataset.** Polymorphism data of the *IGF1R* gene investigated in this paper was retrieved from the dbSNP database [33]. It contained a total of 2412 SNPs, out of which 32 (1.3%) were nsSNPs, 58 (2.4%) were sSNPs, 83 (3.4%) occurred in the mRNA 3' UTR, and 2225 (92.2%) occurred in intronic regions. SNPs in the 5' UTR region were not found. It can be seen from the distribution in Figure 1 that the vast majority of SNPs occur in the intronic region, and that there are more 3' UTR region SNPs than nsSNPs or sSNPs. We selected missense nsSNPs, sSNPs, 3' UTR SNPs, and intronic SNPs for our investigation.

**3.2. Deleterious nsSNPs by SIFT Program.** Protein sequence with mutational position and amino acid residue variants associated to 24 missense nsSNPs were submitted as input to the SIFT server, and the results are shown in Table 1, along with the corresponding heterozygosity and validation status description for each SNP, when available from

TABLE 2: List of SNPs predicted to be functionally significant by FASTSNP.

dbSNP ID	Nucleotide change	Region	Level of risk	Possible functional effect	Heterozygosity	Validation
rs45437300	A/T	coding	Very High-Very High (5-5)	Nonsense	n/a	
rs55895813	A/G	intronic	Medium-High (3-4)	Splicing site	n/a	
rs36108138	A/C	intronic	Medium-High (3-4)	Splicing site	n/a	
rs45495500	C/T	intronic	Medium-High (3-4)	Splicing site	n/a	
rs34226328	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.006	2
rs35041862	C/G	coding	Low-Medium (2-3)	Splicing regulation	0.017	2
rs55770488	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs35385418	A/G	coding	Low-Medium (2-3)	Splicing regulation	0.022	1; 2
rs45504194	A/G	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs45582234	G/T	coding	Low-Medium (2-3)	Splicing regulation	0.012	
rs17847210	G/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	1
rs56013396	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs35171849	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.011	1; 2
rs35812156	A/C	coding	Low-Medium (2-3)	Splicing regulation	n/a	1
rs55954954	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs45506098	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.013	1; 4; 5
rs45598332	G/T	coding	Low-Medium (2-3)	Splicing regulation	0.013	
rs45615734	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs45486504	C/G	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs3743262	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.255	1; 4; 5
rs45627636	A/G	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs45443393	A/G	coding	Low-Medium (2-3)	Splicing regulation	0.011	
rs45459793	A/G	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs56400113	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs35449468	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.006	1
rs17847208	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.005	1; 2
rs2229765	A/G	coding	Low-Medium (2-3)	Splicing regulation	0.458	1; 2; 3; 4
rs28664854	A/G	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs35362396	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.005	

TABLE 2: Continued.

dbSNP ID	Nucleotide change	Region	Level of risk	Possible functional effect	Heterozygosity	Validation
rs45598038	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	1
rs34364279	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.006	
rs45468291	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs56020698	C/T	coding	Low-Medium (2-3)	Splicing regulation	n/a	
rs17847203	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.170	1; 2
rs45453791	C/T	coding	Low-Medium (2-3)	Splicing regulation	0.039	1

Validation Status Description: (1) Validated by multiple, independent submissions to the refSNP cluster; (2) Validated by frequency or genotype data: minor alleles observed in at least two chromosomes; (3) All alleles have been observed in at least two chromosomes apiece; (4) Genotyped by HapMap project; (5) SNP has been sequenced in 1000 Genome project.

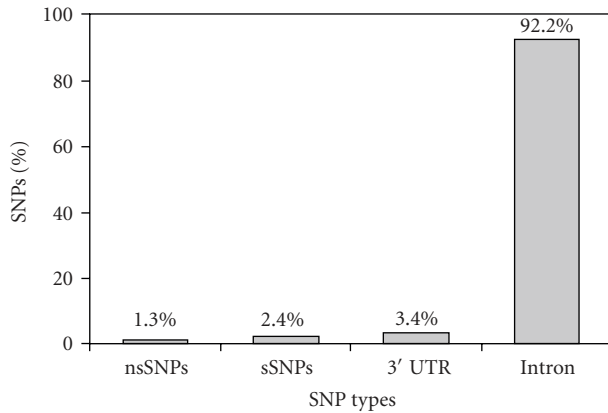


FIGURE 1: Distribution of *IGF1R* non-synonymous SNPs (nsSNPs), synonymous SNPs (sSNPs), 3' UTR SNPs, and intronic SNPs.

dbSNP. According to the classification proposed by Ng and Henikoff [24] and Xi et al. [28], the lower the tolerance index, the higher the functional impact a particular amino acid residue substitution is likely to have and vice versa. Among the 24 nsSNPs analyzed, 8 nsSNPs were identified to be deleterious with a tolerance index score  $\leq 0.05$ . Five nsSNPs (rs61740868, rs45578132, rs45553041, rs45526336, and rs45504297) showed a highly deleterious tolerance index score of 0.00. The remaining deleterious nsSNPs showed tolerance index scores of 0.01 (rs45524940 and rs45512296) and 0.03 (rs45445894). Four deleterious nsSNPs showed a nucleotide change from G/A, four a change from C/T, two a change from T/C, and one a change from A/G.

**3.3. Damaged nsSNPs by PolyPhen Server.** All the 24 protein sequences of missense nsSNPs submitted to SIFT were also submitted to the PolyPhen server. A PSIC score difference of 1.5 and above is considered to be damaging. Eight nsSNPs (rs70958401, rs61740868, rs45578132, rs45504297,

rs45553041, rs45512296, rs45524940, and rs33958176) were considered to be damaging and exhibited a range of PSIC score difference between 1.503 and 2.609 (Table 1). Out of these damaging nsSNPs, two changed from positively charged amino acid in the native protein to hydrophobic amino acid in the mutant type, two from aliphatic nonpolar amino acid to non-polar amino acid, two from positively charged amino acid to aromatic positively charged amino acid, one from polar amino acid to non-polar amino acid, and one from positively charged to polar amino acid, respectively. It can be seen from Table 1 that there was significant correlation between the results obtained from the evolutionary-based approach SIFT and the structural-based approach PolyPhen for six nsSNPs predicted to be damaging by PolyPhen, suggesting that these nsSNPs may disrupt both the protein function and structure. The most damaging nsSNP (rs61740868) showed a PSIC score of 2.609, due to a mutation from arginine to cysteine.

**3.4. SNPs in Regulatory Regions.** According to FASTSNP, out of 58 sSNPs in the *IGF1R* gene, 31 sSNPs were predicted to be damaging with a risk ranking of 2-3, and a possible functional effect on splicing regulation (Table 2). Among these, the A/G polymorphism (rs2229765) has been shown experimentally to affect the susceptibility to ischemic stroke in Chinese population [19] to be associated with higher plasma concentrations of circulating IGF1R and premature pubarche [39, 40] and adult height variation in the human population [41]. Out of 2225 SNPs which occur in the intronic region of the *IGF1R* gene, 3 SNPs (rs55895813, rs36108138 and rs45495500) were predicted to affect the splicing site (3-4 risk) (Table 2).

It can be seen from Table 2 that a coding nonsense SNP (rs45437300) due to a nucleotide change from A to T was detected and showed a very high (5-5) level of risk, as it can truncate and even inactivate the IGF1R protein, causing disease as a result.



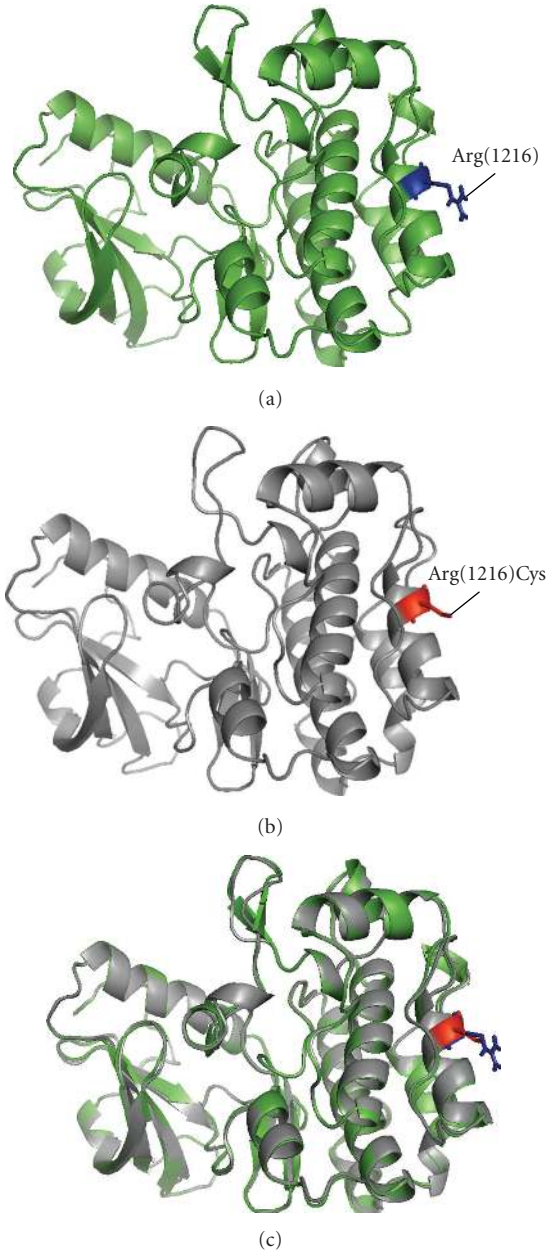


FIGURE 2: (a) Native structure (2jo9) showing arginine at position 1216. (b) Mutant modeled structure (2jo9 R1216C) showing cysteine residue at position 1216. (c) Superimposed structure of native structure (2jo9) (green) with mutant modeled structure (2jo9 R1216C) (gray).

**3.5. Structural Analysis of Mutant Structures.** Out of eight nsSNPs predicted to be deleterious by SIFT or PolyPhen, four (rs61740868, rs45526336, rs45512296, and rs45504297) were mapped to the PDB ID 2jo9 native structure. The amino acid residue substitutions were performed by Swiss-Pdb Viewer independently to get four mutant modeled structures (2jo9 R1216C, 2jo9 E1253K, 2jo9 R1216H, and 2jo9 L1211P, respectively). Then, energy minimizations were performed by GROMACS for the native structure (2jo9) and the mutant modeled structures.

TABLE 3: RMSD and total energy of native structure (2jo9) and mutant modeled structures.

dbSNP ID	Amino Acid change	RMSD between native and mutant structures	Total energy after minimization (KJ/mol)
rs61740868	Arg1216Cys	0,48	-13343.28
rs45526336	Glu1253Lys	0,38	-13887.05
rs45512296	Arg1216His	0,46	-13483.34
rs45504297	Leu1211Pro	0,22	-13782.33

Total energy of native structure (2jo9) after energy minimization: -13841.67.

The total energy for the native structure (2jo9) and the four mutant modeled structures 2jo9 R1216C, 2jo9 E1253K, 2jo9 R1216H, and 2jo9 L1211P was -13841.67, -13343.28, -13887.05, -13483.34, and -13782.33 KJ/mol, respectively (Table 3). Three out of four mutant modeled structures (2jo9 R1216C, 2jo9 R1216H, and 2jo9 L1211P) showed an increase in energy (less favorable change) in comparison with the native structure. This result correlates with the structural homology method (PolyPhen) results, which predicted all these three mutants to be deleterious (PSIC scores 2.609, 2.128, and 2.372, resp.) (Table 1). The mutant model 2jo9 R1216C showed the greatest increase in energy, which may be explained by the energetically unfavorable substitution of a positively charged arginine amino acid residue to a non-polar cysteine amino acid residue at the surface of the protein structure (Figure 2).

It can be seen from Table 3 that the RMSD values between the native structure (2jo9) and the mutant modeled structures are all similar, ranging from 0.22 Å to 0.48 Å. Because these values are low, we can suggest that these mutations do not cause a significant change in the mutant structures with respect to the native protein structure.

#### 4. Conclusion

In this paper, we investigated the functional and structural impact of SNPs in the *IGF1R* gene using computational prediction tools. Out of a total of 2412 SNPs in the *IGF1R* gene, 32 SNPs were found to be non-synonymous, 58 were synonymous, 83 occurred in the mRNA 3' UTR, and 2225 were found in intronic regions. Out of 24 missense nsSNPs, eight were found to be deleterious by SIFT, and eight were found to be damaging by the PolyPhen tool. A total of six nsSNPs were found to be damaging by both SIFT and PolyPhen tools. The structural analysis results showed that the amino acid residue substitutions which had the greatest impact on the stability of the IGF1R protein were mutations 2jo9 R1216C (rs61740868) and R1216H (rs45512296). Among the nsSNPs studied, a nonsense SNP (rs45437300) was found. Out of 58 sSNPs, 31 were predicted to affect splicing regulation by FASTSNP, including an sSNP (rs2229765) associated with several diseases. In the intronic region, 3 SNPs (rs55895813, rs36108138, and rs45495500) were predicted to affect splicing regulation. Based on our

results, we conclude that these SNPs should be considered important candidates in causing diseases related to *IGF1R* malfunction.

## Acknowledgments

This work was supported by an FAPEMIG fellowship (S.A. de Alencar) and a CNPQ grant (J.C.D. Lopes).

## References

- [1] J.-E. Lee, J. H. Choi, J. H. Lee, and M. G. Lee, "Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis," *Mutation Research*, vol. 573, no. 1-2, pp. 195–204, 2005.
- [2] M. Krawczak, E. V. Ball, I. Fenton et al., "Human gene mutation database—a biomedical information and research resource," *Human Mutation*, vol. 15, no. 1, pp. 45–51, 2000.
- [3] L. Prokunina and M. E. Alarcón-Riquelme, "Regulatory SNPs in complex diseases: their identification and functional validation," *Expert Reviews in Molecular Medicine*, vol. 6, no. 10, 2004.
- [4] P. D. Stenson, M. Mort, E. V. Ball, et al., "The human gene mutation database: 2008 update," *Genome Medicine*, vol. 1, no. 1, p. 13, 2009.
- [5] V. Ramensky, P. Bork, and S. Sunyaev, "Human non-synonymous SNPs: server and survey," *Nucleic Acids Research*, vol. 30, no. 17, pp. 3894–3900, 2002.
- [6] T. Emahazion, L. Feuk, M. Jobs et al., "SNP association studies in Alzheimer's disease highlight problems for complex disease analysis," *Trends in Genetics*, vol. 17, no. 7, pp. 407–413, 2001.
- [7] N. J. Schork, D. Fallin, and J. S. Lanchbury, "Single nucleotide polymorphisms and the future of genetic epidemiology," *Clinical Genetics*, vol. 58, no. 4, pp. 250–264, 2000.
- [8] C. G. P. Doss, C. Sudandiradoss, R. Rajasekaran et al., "Applications of computational algorithm tools to identify functional SNPs," *Functional and Integrative Genomics*, vol. 8, no. 4, pp. 309–316, 2008.
- [9] N. J. Risch, "Searching for genetic determinants in the new millennium," *Nature*, vol. 405, no. 6788, pp. 847–856, 2000.
- [10] R. Rajasekaran, C. Sudandiradoss, C. G. P. Doss, and R. Sethumadhavan, "Identification and in silico analysis of functional SNPs of the BRCA1 gene," *Genomics*, vol. 90, no. 4, pp. 447–452, 2007.
- [11] C. G. P. Doss, C. Sudandiradoss, R. Rajasekaran, R. Purohit, K. Ramanathan, and R. Sethumadhavan, "Identification and structural comparison of deleterious mutations in nsSNPs of ABL1 gene in chronic myeloid leukemia: a bio-informatics study," *Journal of Biomedical Informatics*, vol. 41, no. 4, pp. 607–612, 2008.
- [12] R. Rajasekaran, C. G. P. Doss, C. Sudandiradoss, K. Ramanathan, R. Purohit, and R. Sethumadhavan, "Effect of deleterious nsSNP on the HER2 receptor based on stability and binding affinity with herceptin: a computational approach," *Comptes Rendus Biologies*, vol. 331, no. 6, pp. 409–417, 2008.
- [13] C. G. P. Doss, R. Rajasekaran, C. Sudandiradoss, K. Ramanathan, R. Purohit, and R. Sethumadhavan, "A novel computational and structural analysis of nsSNPs in CFTR gene," *Genomic Medicine*, vol. 2, no. 1-2, pp. 23–32, 2008.
- [14] R. Rajasekaran and R. Sethumadhavan, "In Silico identification of significant detrimental missense mutations of EGFR and their effect with 4-anilinoquinazoline-based drugs," *Applied Biochemistry and Biotechnology*, vol. 160, no. 6, pp. 1723–1733, 2010.
- [15] G. S. Warshamana-Greene, J. Litz, E. Buchdunger, C. García-Echeverría, F. Hofmann, and G. W. Krystal, "The insulin-like growth factor-I receptor kinase inhibitor, NVP-ADW742, sensitizes small cell lung cancer cell lines to the effects of chemotherapy," *Clinical Cancer Research*, vol. 11, no. 4, pp. 1563–1571, 2005.
- [16] H. E. Jones, L. Goddard, J. M. W. Gee et al., "Insulin-like growth factor-I receptor signalling and acquired resistance to gefitinib (ZD1839; Iressa) in human breast and prostate cancer cells," *Endocrine-Related Cancer*, vol. 11, no. 4, pp. 793–814, 2004.
- [17] S. L. Krueckl, R. A. Sikes, N. M. Edlund et al., "Increased insulin-like growth factor I receptor expression and signaling are components of androgen-independent progression in a lineage-derived prostate cancer progression model," *Cancer Research*, vol. 64, no. 23, pp. 8620–8629, 2004.
- [18] G. O. Hellawell, G. D. H. Turner, D. R. Davies, R. Poulsom, S. F. Brewster, and V. M. Macaulay, "Expression of the type 1 insulin-like growth factor receptor is up-regulated in primary prostate cancer and commonly persists in metastatic disease," *Cancer Research*, vol. 62, no. 10, pp. 2942–2950, 2002.
- [19] J. Cheng, J. Liu, X. Li, et al., "Insulin-like growth factor-1 receptor polymorphism and ischemic stroke: a case-control study in Chinese population," *Acta Neurologica Scandinavica*, vol. 118, no. 5, pp. 333–338, 2008.
- [20] J. Garcia, A. Ahmadi, A. Wonnacott, et al., "Association of insulin-like growth factor-1 receptor polymorphism in dementia," *Dementia and Geriatric Cognitive Disorders*, vol. 22, no. 5-6, pp. 439–444, 2006.
- [21] P. C. Ng and S. Henikoff, "SIFT: predicting amino acid changes that affect protein function," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3812–3814, 2003.
- [22] S. Sunyaev, V. Ramensky, and P. Bork, "Towards a structural basis of human non-synonymous single nucleotide polymorphisms," *Trends in Genetics*, vol. 16, no. 5, pp. 198–200, 2000.
- [23] H.-Y. Yuan, J.-J. Chiou, W.-H. Tseng et al., "FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization," *Nucleic Acids Research*, vol. 34, pp. W635–W641, 2006.
- [24] P. C. Ng and S. Henikoff, "Predicting deleterious amino acid substitutions," *Genome Research*, vol. 11, no. 5, pp. 863–874, 2001.
- [25] P. C. Ng and S. Henikoff, "Accounting for human polymorphisms predicted to affect protein function," *Genome Research*, vol. 12, no. 3, pp. 436–446, 2002.
- [26] P. C. Ng and S. Henikoff, "Predicting the effects of amino acid substitutions on protein function," *Annual Review of Genomics and Human Genetics*, vol. 7, pp. 61–80, 2006.
- [27] S. Sunyaev, W. Lathe III, and P. Bork, "Integration of genome data and protein structures: prediction of protein folds, protein interactions and "molecular phenotypes" of single nucleotide polymorphisms," *Current Opinion in Structural Biology*, vol. 11, no. 1, pp. 125–130, 2001.
- [28] T. Xi, I. M. Jones, and H. W. Mohrenweiser, "Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function," *Genomics*, vol. 83, no. 6, pp. 970–979, 2004.
- [29] D. Chasman and R. M. Adams, "Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation," *Journal of Molecular Biology*, vol. 307, no. 2, pp. 683–706, 2001.

- [30] M. Cargill, D. Altshuler, J. Ireland et al., "Characterization of single-nucleotide polymorphisms in coding regions of human genes," *Nature Genetics*, vol. 22, no. 3, pp. 231–238, 1999.
- [31] Z. Wang and J. Moulton, "SNPs, protein structure, and disease," *Human Mutation*, vol. 17, no. 4, pp. 263–270, 2001.
- [32] N. Tokuriki, F. Stricher, L. Serrano, and D. S. Tawfik, "How protein stability and new functions trade off," *PLoS Computational Biology*, vol. 4, no. 2, Article ID e1000002, 2008.
- [33] S. T. Sherry, M.-H. Ward, M. Kholodov et al., "dbSNP: the NCBI database of genetic variation," *Nucleic Acids Research*, vol. 29, no. 1, pp. 308–311, 2001.
- [34] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.
- [35] U. Veluparthi, M. Wittman, P. Liu et al., "Discovery and initial SAR of 3-(1H-benzo[d]imidazol-2-yl)pyridin-2(1H)-ones as inhibitors of insulin-like growth factor 1-receptor (IGF-1R)," *Bioorganic and Medicinal Chemistry Letters*, vol. 17, no. 8, pp. 2317–2321, 2007.
- [36] P. Stothard, "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences," *BioTechniques*, vol. 28, no. 6, pp. 1102–1104, 2000.
- [37] N. Gueux, A. Diemand, and M. C. Peitsch, "Protein modelling for all," *Trends in Biochemical Sciences*, vol. 24, no. 9, pp. 364–366, 1999.
- [38] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.
- [39] M. Bonafè, M. Barbieri, F. Marchegiani et al., "Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: cues for an evolutionarily conserved mechanism of life span control," *Journal of Clinical Endocrinology and Metabolism*, vol. 88, no. 7, pp. 3299–3304, 2003.
- [40] M. B. Roldan, C. White, and S. F. Witchel, "Association of the GAA1013 → GAG polymorphism of the insulin-like growth factor-1 receptor (IGF1R) gene with premature pubarche," *Fertility and Sterility*, vol. 88, no. 2, pp. 410–417, 2007.
- [41] V. M. Chia, L. C. Sakoda, B. I. Graubard et al., "Risk of testicular germ cell tumors and polymorphisms in the insulin-like growth factor genes," *Cancer Epidemiology Biomarkers and Prevention*, vol. 17, no. 3, pp. 721–726, 2008.



