

# Protein molecular modeling of genetic markers for thyroid cancer

## *Modelagem molecular proteica dos marcadores genéticos do câncer de tireoide*

Luis Jesuino O. Andrade<sup>1</sup>; Gabriela C. M. Oliveira<sup>1</sup>; Alcina Maria V. Bittencourt<sup>2</sup>; Paulo Roberto S. Melo<sup>1</sup>

1. Universidade Estadual de Santa Cruz (UESC), Bahia, Brazil. 2. Universidade Federal da Bahia (UFBA), Bahia, Brazil.

### ABSTRACT

**Introduction:** The advances in thyroid molecular biology studies provide not only insight into thyroid diseases but accurate diagnosis of thyroid cancer. **Objective:** Design a tutorial on protein molecular modeling of genetic markers for thyroid cancer. **Methods:** The proteins were selected using the Protein Data Bank sequence and the basic local alignment search tool (BLAST) algorithm. The obtained sequences were aligned with the Clustal W multiple alignment algorithms. For the molecular modeling, three-dimensional structures were generated from this set of constraints with the SWISS-MODEL, which is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server. **Results:** We demonstrated protein analysis, projection of the molecular structure and protein homology of the following molecular markers of thyroid cancer: receptor tyrosine kinase (RET) proto-oncogene; neurotrophic tyrosine kinase receptor 1 (NTRK1) proto-oncogene; phosphatase and tensin homolog (PTEN); tumor protein p53 (TP53) gene; phosphoinositide 3-kinase/threonine protein kinase (PI3K/AKT); catenin beta 1 (CTNNB1); paired box 8-peroxisome proliferator-activated receptor gamma (PAX8-PPARG); rat sarcoma viral oncogene (RAS); B-raf proto-oncogene, serine/threonine kinase (BRAF); and thyroid-stimulating hormone receptor (TSHR). **Conclusion:** This study shows the importance of understanding the molecular structure of the markers for thyroid cancer through bioinformatics, and consequently, the development of more effective new molecules as alternative tools for thyroid cancer treatment.

**Key words:** sequence analysis protein; molecular structure; genetic markers; thyroid neoplasms.

### INTRODUCTION

Thyroid cancer incidence is increasing worldwide, while mortality is stable or decreasing. Advances in thyroid molecular biology studies not only provide insight into development of thyroid diseases but offer an accurate diagnosis of thyroid cancer.

The evolution of knowledge in molecular diagnosis of thyroid cancer has accelerated in the latest years, especially with the advent of high-throughput sequencing technologies, when new molecular abnormalities have been diagnosed. The molecular markers of thyroid cancer are found in more than 70% of differentiated carcinomas, and understanding the molecular mechanisms of thyroid cancer offers new perspectives for its diagnosis and treatment<sup>(1,2)</sup>.

Progress in protein sequencing has led to an increase of protein databases, with a variety of protein sequences never seen before. Many molecular markers of thyroid cancer are proteins

present at very low levels. Similar sequences and protein structures enable their evaluation with the use of computational comparative methods<sup>(3)</sup>.

The bioinformatics program currently available to perform molecular modeling and protein analysis provides a tool for assembly of the molecular markers of thyroid cancer, and understanding of its molecular mechanisms. The purpose of this study is to perform a tutorial about protein molecular modeling of genetic markers for thyroid cancer.

### METHODS

We evaluated a search for protein sequence, projection of the molecular structure and protein homology of the following molecular markers of thyroid cancer: receptor tyrosine kinase (RET) proto-oncogene; neurotrophic tyrosine kinase receptor 1 (NTRK1)

proto-oncogene; phosphatase and tensin homolog (PTEN); tumor protein p53 (TP53) gene; phosphoinositide 3-kinase/threonine protein kinase (PI3K/AKT); catenin beta 1 (CTNNB1); paired box 8-peroxisome proliferator-activated receptor gamma (PAX8-PPARG); rat sarcoma viral oncogene (RAS); B-raf proto-oncogene, serine/threonine kinase (BRAF); and thyroid-stimulating hormone receptor (TSHR).

The proteins were selected using the Protein Data Bank sequence and the basic local alignment search tool (BLAST) algorithm<sup>(4)</sup>. The obtained sequences were aligned with the Clustal W multiple alignment algorithms. For molecular modeling, three-dimensional (3-D) structures were generated from this set of constraints with the SWISS-MODEL (<http://swissmodel.expasy.org>), which is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server.

**Protein database search and sequence analysis – BLAST**

BLAST is a used protein sequence analysis tool available in the public domain (<http://www.ncbi.nlm.nih.gov/Blast/>), and a wide choice of BLAST algorithms is used to search many different sequence databases via BLAST. BLAST was introduced as a sequence alignment heuristic that was an order of magnitude faster than earlier approaches for analyzing biological information. Very quickly, this software became a landmark enabling technique for bioinformatics. Thus, the BLAST 2p refers to a program used to generate alignments between a nucleotide or protein sequence, referred to as a query and nucleotide sequences against other database of nucleotide, referred to as subject sequences.

**Building 3-D molecular models**

The structure and function of proteins are determined by their amino acid sequences, and the 3-D structure prediction still remains a significant challenge, with a great demand for high-resolution structure prediction methods.

Homology modeling is currently the most accurate computational method to generate reliable structural models and is routinely used in many biological applications.

**Modeling with SWISS-MODEL**

SWISS-MODEL is a server for automated comparative modeling of 3-D protein structures. The process begins with the identification of suitable template structures based on their sequence similarity with the target sequence. This is achieved

by searching a database of sequences with known 3-D structures using the sequence alignment tools BLAST and FASTA.

SWISS-MODEL provides several levels of user interaction through its World Wide Web interface: the “first approach mode”, the “alignment mode”, and the “project mode” using DeepView (Swiss-PdbViewer), an integrated sequence-to-structure workbench. The Swiss-PdbViewer not only provides advanced molecular display features and real-time visual feedback during structure modeling, but also features direct submission to the SWISS-MODEL model server.

**RESULTS**

We demonstrated a search for protein sequence, the projection of the molecular structure and protein homology of the following molecular markers of thyroid cancer: RET proto-oncogene, NTRK1 proto-oncogene, PTEN, TP53 gene, PI3K/AKT, CTNNB1, PAX8-PPARG, RAS, BRAF, and TSHR.

**Search for protein sequence of RET proto-oncogene**

The main data source used for reconstructing the RET proto-oncogene was the protein sequence file in FASTA format. The full-length protein of RET proto-oncogene was obtained from the GenBank database under the identifier NCBI: AAH04257.1. The RET proto-oncogene was predicted to encode a 1072-amino acid protein. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. RET proto-oncogene [*Homo sapiens*] protein analysis is shown in **Figure 1**.

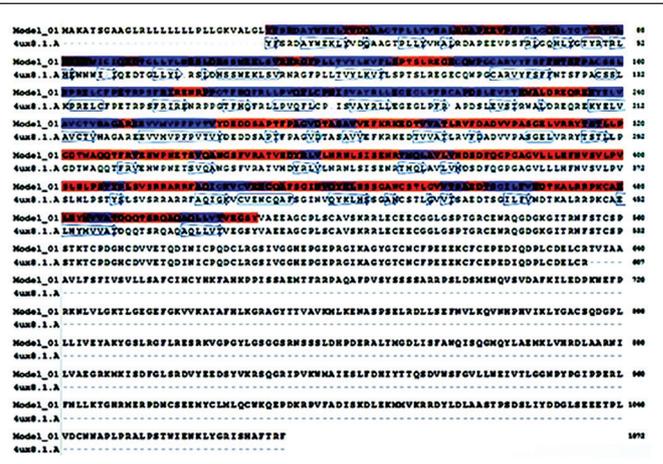


FIGURE 1 – Proto-oncogene tyrosine protein kinase receptor RET – model – template alignment

### Building a 3-D molecular model of RET proto-oncogene

Protein sequences of RET proto-oncogene [*Homo sapiens*] were obtained using BLAST; 3-D modeling was conducted using the Swiss-PdbViewer suite programs, which were adjusted and optimized for alignment between RET proto-oncogene protein and structural templates. On the basis of a sequence alignment between the RET proto-oncogene protein and the template structure, a 3-D model for the target protein was generated. Model quality assessment tools were used to estimate the reliability of the resulting model. Thus, using the SWISS-MODEL automated comparative protein modeling server, we constructed a homology model of the RET proto-oncogene [*Homo sapiens*] (Figure 2).

### Search for protein sequence of NTRK1 proto-oncogene

The NTRK1 proto-oncogene was predicted to encode a 786-amino acid transmembrane tyrosine kinase expressed in neural tissues. The reconstruction of the NTRK1 proto-oncogene was made from a protein sequence file in FASTA format. The full-length protein of NTRK1 proto-oncogene was obtained from the GenBank database under the identifier NCBI: P04629.4. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. NTRK1 proto-oncogene [*Homo sapiens*] protein analysis is shown in Figure 3.

### Building a 3-D molecular model of NTRK1 proto-oncogene

Template search with BLAST was performed against the SWISS-MODEL template library. The NTRK1 sequence was searched with BLAST<sup>(5)</sup>. A total of 500 templates were found. An initial profile was built using the procedure outlined by Remmert *et al.* (2011)<sup>(6)</sup>. The templates with the highest quality were then selected for model building. Ligands present in the template structure were transferred by homology to the model when the proposed criteria were met. The homo-oligomeric structure of the NTRK1 protein is predicted based on the analysis of pairwise interfaces of the identified template structures. According to the above-described criteria, a 3-D model for the NTRK1 proto-oncogene was generated (Figure 4).

### Search for protein sequence of PTEN

We conducted BLAST analysis using the PTEN phosphatase domain as a query, and subsequently identified a predicted

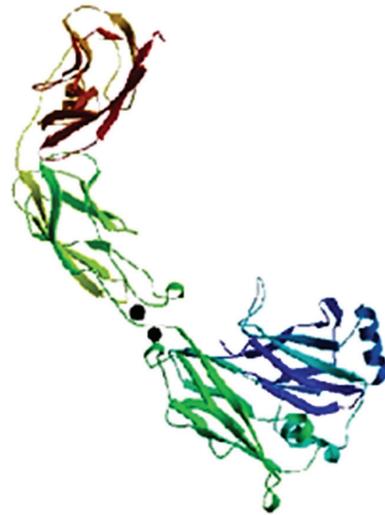


FIGURE 2 – Homology model of the RET proto-oncogene [*Homo sapiens*]

Model_02	MLRGGRRGQLGWNHWAAGPSLLANLILASAGAAPCPDACCPHGSSGLRCTRDCALDLSLHNLPGAENLTK	70
4f01.2.A	-----	
Model_02	LYIENQHQHQLHLELRDLRGLGELRNLTIVKSGLRFPVADAFHPTPLRLSRLNLSFNALSLSMKVTQGLSL	140
4f01.2.A	-----	
Model_02	QELVLSGNPLHCSALRWLQRWEEEGLVGVEPKLQCHGQGPLAHMPNASCQVPTLVKQVFNASVDVGD	210
4f01.2.A	-----	
Model_02	VLLRCQVEGRGLEQAGWILTELEQSATVMKSGGLPSLGLTLANVTSDLNRRKNVTCWAENDVGRAEVSQV	280
4f01.2.A	-----	
Model_02	NVSPFASVQLHTAVEMHHNCIPFSVDGQPAPSLRWLNFNGSVLNETSIFITFELEPAANETVRHGCLRNLQ	350
4f01.2.A	-----	
Model_02	PTHVNGNYTLTLANPFGQASASIMAAFMDNPFEPNEDPDPVPSVSPVDNNTSGDPPVKKEDTFFGVSV	420
4f01.2.A	-----	
Model_02	AVGLVAFACFLSTLLLVLNKGRNRNFGINRPAVLAPEDGLAMSLHFMPTLGGSSLSPTEGKSGQLGHI	490
4f01.2.A	-----	
Model_02	IENPQYFSDA <b>VNIRKRDIVLKWLEGGAPQKVFIAECHMLPEQDQHLVAVKALKEASESARQDFQR</b>	560
4f01.2.A	-----SDACV <b>IKRRIVLKWLEGGAPQKVFIAECHMLPEQDQHLVAVKALKEASESARQDFQR</b>	64
Model_02	<b>ALLTILQHQHVAIFQVCTGSRVLLMVFETMRGDLRFLAENGPDAKLLAGQDVAQVLLGLQLLAV</b>	630
4f01.2.A	<b>AELLTILQHQHVAIFQVCTGSRVLLMVFETMRGDLRFLAENGPDAKLLAGQDVAQVLLGLQLLAV</b>	134
Model_02	<b>AEQVAAGHYVLAGLHFVHRDLATRNCLDQGLIVYKIDGDFGMSRDIVSTIDYRGGRRMIPERMPPESIL</b>	700
4f01.2.A	<b>AEQVAAGHYVLAGLHFVHRDLATRNCLDQGLIVYKIDGDFGMSRDIVSTIDYRGGRRMIPERMPPESIL</b>	204
Model_02	<b>YKFTTESDVMSEFVYVMEIFETAKQPFQSRFEALDCITQGRALATACIFAVYAIMGCGGRRDQ</b>	770
4f01.2.A	<b>YKFTTESDVMSEFVYVMEIFETAKQPFQSRFEALDCITQGRALATACIFAVYAIMGCGGRRDQ</b>	274
Model_02	<b>HSIKDVMHARLQALQAAPPVYLDVLG</b>	796
4f01.2.A	<b>HSIKDVMHARLQALQAAPPVYLDVLG</b>	300

FIGURE 3 – NTRK1 proto-oncogene – model – template alignment



FIGURE 4 – NTRK1 proto-oncogene [*Homo sapiens*]

protein by NCBI accession: AAD13528.1. The PTEN [*Homo sapiens*] from all sources has 403 basic amino acids in the highlighted positions. The reconstruction of the PTEN protein [*Homo sapiens*] was made from a protein sequence file in FASTA format. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. PTEN protein [*Homo sapiens*] analysis is shown in **Figure 5**.

### Building a 3-D molecular model of PTEN

For molecular modeling, a sequence alignment was obtained by using Clustal W and a threading approach. Template search with BLAST was conducted against the SWISS-MODEL template library. On the basis of a sequence alignment between the PTEN protein and the template structure, a 3-D model for the target protein was generated. Model quality assessment tools were used to estimate the reliability of the resulting model. Thus, using the SWISS-MODEL automated comparative protein modeling server, we constructed a homology model of the PTEN protein [*Homo sapiens*] (**Figure 6**).

### Search for protein sequence of TP53 gene

Protein analysis of TP53 gene was performed using BLAST and subsequently identified a predicted protein by NCBI accession: AEX20383.1. The TP53 gene [*Homo sapiens*] from all sources has 354 basic amino acids in the highlighted positions. The reconstruction of TP53 gene [*Homo sapiens*] was made from a protein sequence file in FASTA format. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. TP53 gene [*Homo sapiens*] protein analysis is shown in **Figure 7**.

### Building a 3-D molecular model of TP53 gene

The target sequence was searched with BLAST against the primary amino acid sequence. A total of 97 templates were found. For each identified template, the template quality was predicted from features of the target-template alignment. The template with the highest quality was used for model building.

Structural homology modeling was used to model the TP53 gene [*Homo sapiens*] amino acid sequence, based on the high-resolution crystal structure, and this structural model was calculated using the SWISS-MODEL, producing a high quality 3-D structure with a model-template C- $\alpha$  root mean square deviation of 2.9 angstroms (**Figure 8**).



FIGURE 5 – PTEN protein [*Homo sapiens*] – model – template alignment

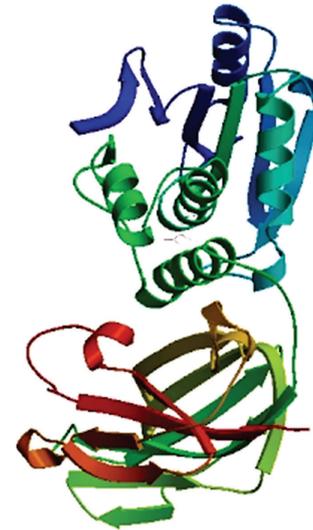


FIGURE 6 – PTEN protein



FIGURE 7 – TP53 gene – model – template alignment

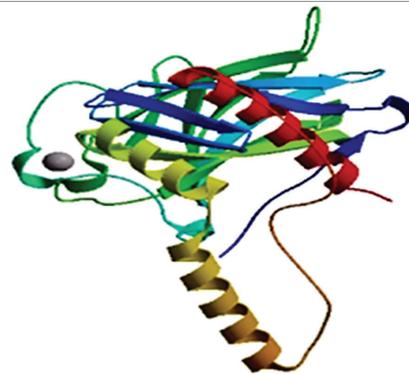


FIGURE 8 – Tumor suppressor TP53 [*Homo sapiens*]

### Search for protein sequence of PI3K/AKT

We conducted BLAST analysis using the PI3K/AKT domain as a query, and subsequently identified a predicted protein by NCBI accession: AGC00787.1. The PI3K/AKT from all sources has 527 basic amino acids in the highlighted positions. The reconstruction of the PI3K/AKT was made from a protein sequence file in FASTA format. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. PI3K/AKT protein analysis is shown in **Figure 9**.

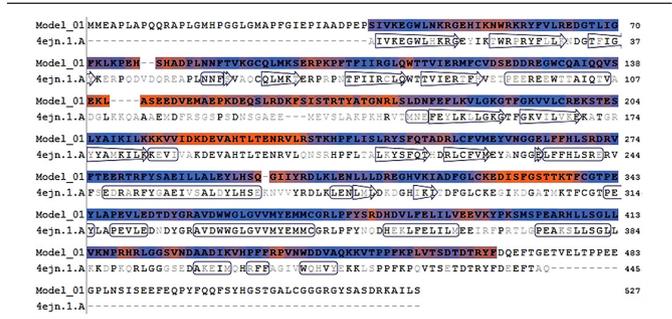


FIGURE 9 – Phosphoinositide 3-kinase (PI3K)/AKT – model – template alignment

### Building a 3-D molecular model of PI3K/AKT

A 3-D model of PI3K/AKT (NCBI GI number: AGC00787.1) comprising 527 amino acids was built using the SWISS-MODEL program based on homology modeling (**Figure 10**). The SWISS-MODEL program automatically provides an all-atom model using alignments between the query sequence and known homologous structures.

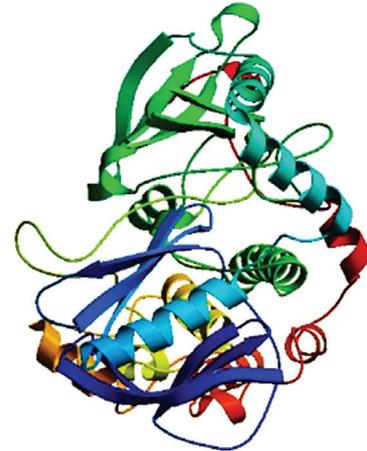


FIGURE 10 – Phosphoinositide 3-kinase (PI3K)/AKT

### Search for protein sequence of CTNNB1

Protein analysis of catenin beta 1 [*Homo sapiens*] was performed using BLAST, and subsequently identified a predicted protein by NCBI Reference Sequence: NP\_001091679.1. The catenin beta 1 [*Homo sapiens*] from all sources has 781 basic amino acids in the highlighted positions. The reconstruction of catenin beta 1 [*Homo sapiens*] was made from a protein sequence file in FASTA format. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. CTNNB1 catenin beta 1 [*Homo sapiens*] protein analysis is shown in **Figure 11**.

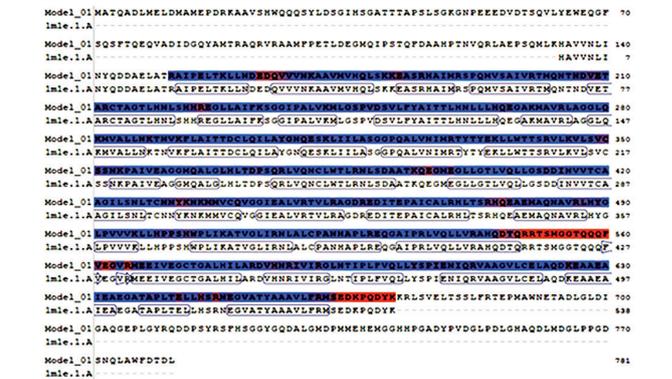


FIGURE 11 – CTNNB1 catenin beta 1 [*Homo sapiens*] – model – template alignment

### Building a 3-D molecular model of CTNNB1

A 3-D model of CTNNB1 by NCBI Reference Sequence: NP\_001091679.1 comprising 781 amino acids was built using the SWISS-MODEL program based on homology modeling. The SWISS-MODEL program automatically provides an all-atom model using alignments between the query sequence and known homologous structures. The homology modeling project submitted to SWISS-MODEL is shown in **Figure 12**.

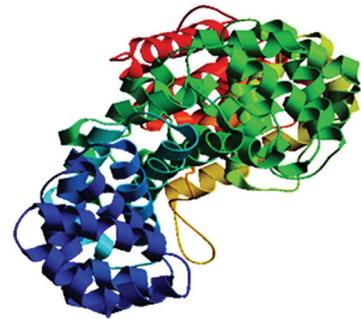


FIGURE 12 – CTNNB1 catenin beta 1 [*Homo sapiens*]

### Search for protein sequence of PAX8/PPARG

Sequences were analyzed with BLAST from the NCBI Reference Sequence: AAU95904.1. The PAX8/PPARG fusion

oncogene was predicted to encode a 505-amino acid protein. The reconstruction of PAX8/PPARG fusion oncogene was made from a protein sequence file in FASTA format. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. The PAX8/PPARG fusion oncogene protein analysis is shown in **Figure 13**.

### Building a 3-D molecular model of PAX8/PPARG

For molecular modeling, a sequence alignment was obtained by using Clustal W and a threading approach. Template search with BLAST was performed against the SWISS-MODEL template library. On the basis of a sequence alignment between the PAX8/PPARG fusion oncogene and the template structure, a 3-D model for the target protein was generated. Model quality assessment tools were used to estimate reliability of the resulting model. Thus, using the SWISS-MODEL automated comparative protein modeling server, we constructed a homology model of the chimeric PAX8/PPARG protein (**Figure 14**).

### Search for protein sequence of RAS

The protein sequence was accessed with the NCBI Reference Sequence: P62834.1 Protein Data Bank using the BLAST tool. Sequence alignments for this protein were generated with the Clustal W program and the alignments scrutinized. The RAS gene was predicted to encode a 184-amino acid protein. The reconstruction of the RAS gene was made from a protein sequence file in FASTA format. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation of the SWISS-MODEL model server. The RAS gene protein analysis is shown in **Figure 15**.

### Building a 3-D molecular model of RAS

For molecular modeling, a sequence alignment was obtained by using Clustal W and a threading approach. Template search with BLAST was performed against the SWISS-MODEL template library. On the basis of a sequence alignment between the RAS gene and the template structure, a 3-D model for the target protein was generated. Model quality assessment tools were used to estimate the reliability of the resulting model. Thus, using the SWISS-MODEL automated comparative protein modeling server, a homology model of the RAS gene (**Figure 16**) was constructed.

```

Model_01 MGETLGDSPIDPESDSTDTLSANISQEMTMVDTMPFFWTFNGISSVDLSVMEHSHSFDIKPFTTVD 70
3e00.1.B -----
Model_01 SSISSTPHYEDIPFTRDTPVADYKYDLKQYQSAIKVPEASPPYYSERTQLYNKPHEEFSNSLMAIECR 140
3e00.1.B -----YQSAIKVPEASPPYYSERTQLYNKPHEEFSNSLMAIECR 54
Model_01 VGDQKASGPHYGVHACGCKOFFRRTIRLKLIDRCGLNCRHKHSKSKGQYGRFQKCLAVGNSHNAIRF 210
3e00.1.B VCGDQKASGPHYGVHACGCKOFFRRTIRLKLIDRCGLNCRHKHSKSKGQYGRFQKCLAVGNSHNAIRF 124
Model_01 RMPGQAEKELKLAISDIDQLNPEADLALAKHLYDSYIKSPFLTKAKARAILDQKTTOKSPFVYVDM 280
3e00.1.B GRMPQAEKELKLAISDIDQLNPEADLALAKHLYDSYIKSPFLTKAKARAILDQKTTOKSPFVYVDM 194
Model_01 NMLMGGQDKIKFKHITPLQEQSKVAIRYFQCGQFRSVEAVQRIEYAKSIQGFVNLDLNDQVTLTKYGV 350
3e00.1.B NSLMGGQDKIKFKHITPLQEQSKVAIRYFQCGQFRSVEAVQRIEYAKSIQGFVNLDLNDQVTLTKYGV 264
Model_01 HEIITYTLASLNKKGVLISGQGFMTREFLKSLRKFQDFPHEKPFPAVKFALELDDSDLAIFIAVI 420
3e00.1.B HEIITYTLASLNKKGVLISGQGFMTREFLKSLRKFQDFPHEKPFPAVKFALELDDSDLAIFIAVI 334
Model_01 LSGDRFGLLVKPIEDIQDNLQALELQLKLNHPSSQLFAKLLQKMTDLRQIVTEHVQLLQVIKKEVTD 490
3e00.1.B LSGDRFGLLVKPIEDIQDNLQALELQLKLNHPSSQLFAKLLQKMTDLRQIVTEHVQLLQVIKKEVTD 404
Model_01 NSLHFLQEIYKDY 505
3e00.1.B NSLHFLQEIYKDY 419
    
```

FIGURE 13 – PAX8-PPARG protein – model – template alignment

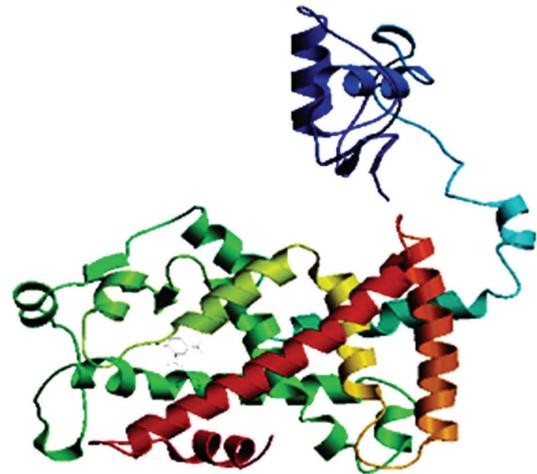


FIGURE 14 – Chimeric PAX8-PPARG protein

```

Model_01 MREYKLVVIGSGGVOKSALTVQVFVQGIYVAKYDPTIILDSYRKQVAVDCQQCMLELDLDTAGTEQFTAMRDL 70
1c1y.1.A MREYKLVVIGSGGVOKSALTVQVFVQGIYVAKYDPTIILDSYRKQVAVDCQQCMLELDLDTAGTEQFTAMRDL 70
Model_01 YHKNQGFALVYSYTAQSTFNDLQDLREQILRVKDYEDVPHILVONKCDLEDRVVOGKQQNLRQWNCN 140
1c1y.1.A YHKNQGFALVYSYTAQSTFNDLQDLREQILRVKDYEDVPHILVONKCDLEDRVVOGKQQNLRQWNCN 140
Model_01 LAFLEPSAKSKINVNATYDYLVRQINRATPVKPKPKKSSCLLL 184
1c1y.1.A LAFLEPSAKSKINVNATYDYLVRQINRATPVKPKPKKSSCLLL 167
    
```

FIGURE 15 – RAS gene – model – template alignment

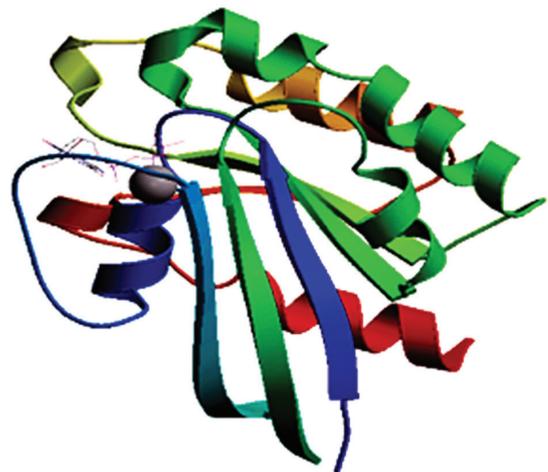


FIGURE 16 – RAS gene

### Search for protein sequence of BRAF

The main data source used for reconstructing BRAF [*Homo sapiens*] was a protein sequence file in FASTA format. The full-length protein of BRAF [*Homo sapiens*] was obtained from the GenBank database under the identifier NCBI: NP\_004324.2. BRAF was predicted to encode a 766-amino acid protein. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation SWISS-MODEL model server. BRAF [*Homo sapiens*] protein analysis is shown in **Figure 17**.

### Building a 3-D molecular model of BRAF

The target sequence was searched with BLAST against the primary amino acid sequence. A total of 500 templates were found. For each identified template, the template quality was predicted from features of the target-template alignment. The template with the highest quality was then selected for model building.

Structural homology modeling was used to model BRAF [*Homo sapiens*] amino acid sequence, based on the high-resolution crystal structure, and this structural model was calculated using the SWISS-MODEL, producing a high-quality 3-D structure.

This document lists the results for the homology modeling project submitted to the SWISS-MODEL workspace. The submitted primary amino acid sequence is given in **Figure 18**.

### Search for protein sequence of TSHR

The main data source used for reconstructing the TSHR protein [*Homo sapiens*] was a protein sequence file in FASTA format. The full-length protein of TSHR [*Homo sapiens*] was obtained from the GenBank database under the identifier NCBI: AAI27629.1. The TSHR [*Homo sapiens*] was predicted to encode a 274-amino acid protein. All coding sequences were selected and exported as amino acids in FASTA format, using the annotation SWISS-MODEL model server. TSHR [*Homo sapiens*] protein analysis is shown in **Figure 19**.

### Building a 3-D molecular model of TSHR

For the molecular modeling, a sequence alignment was obtained by using Clustal W and a threading approach. Template search with BLAST was performed against the SWISS-MODEL template library. On the basis of a sequence alignment between the TSHR protein [*Homo sapiens*] and the template structure, a 3-D model for the target protein was generated. Model quality assessment tools were used to estimate reliability of the resulting

model. Thus, using the SWISS-MODEL automated comparative protein modeling server, we constructed a homology model of the TSHR protein [*Homo sapiens*] (**Figure 20**).

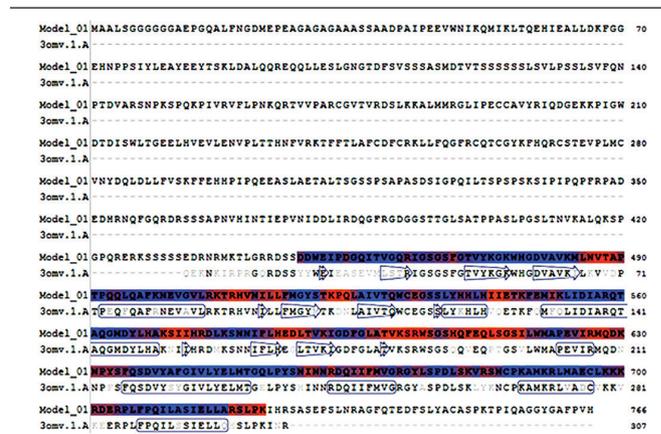


FIGURE 17 – Serine/threonine-protein kinase B-raf [*Homo sapiens*] – model – template alignment

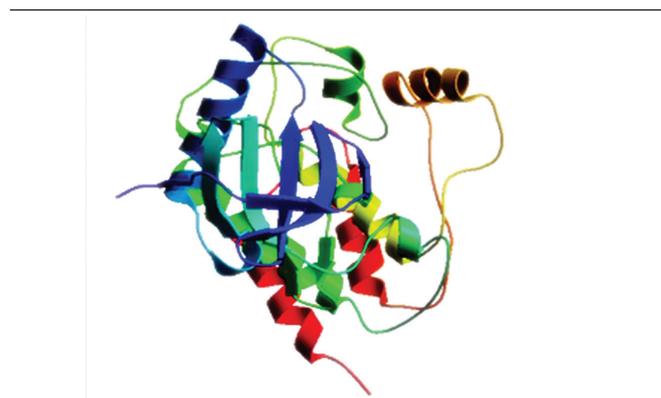


FIGURE 18 – Serine/threonine-protein kinase B-raf [*Homo sapiens*]



FIGURE 19 – TSHR [*Homo sapiens*] – model – template alignment

TSHR: thyroid-stimulating hormone receptor.

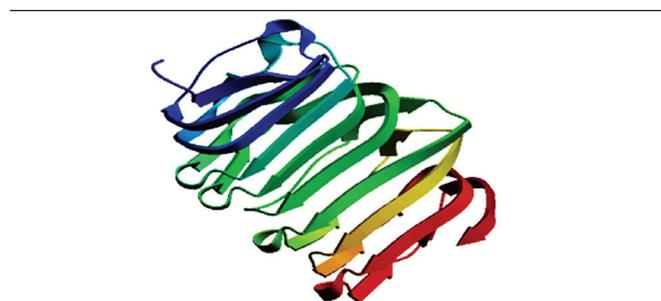


FIGURE 20 – TSHR protein [*Homo sapiens*]

TSHR: thyroid-stimulating hormone receptor.

## DISCUSSION

---

In this study, we developed a tutorial on protein molecular modeling of genetic markers of thyroid cancer. One of the most important objectives of evaluation of the molecular markers in thyroid cancer is accurate diagnosis, besides the possible identification of individuals at greatest risk for thyroid cancer, in order to allow better management, recommendation concerning the most appropriate therapy, and prognosis<sup>(2)</sup>.

The molecular markers for thyroid cancer diagnosis have been an investigative focus that can likely lead to diagnostic improvements, as well as provide more accurate prognostic information before and after surgery.

The RET proto-oncogene, located on chromosome 10q11.2, encodes a membrane receptor of extracellular domain with a tyrosine kinase activity. RET, an abbreviation for “rearranged during transfection”, exhibits an oncogenic potential and plays an important role in thyroid human cancers. In papillary thyroid carcinoma (PTC), genomic rearrangements juxtapose the RET kinase and COOH-terminus encoding domains (exons 11-21) to unrelated genes, thereby creating dominantly transforming oncogenes called RET/PTC<sup>(4,7)</sup>.

A correlation between specific RET mutation type and organ-specific tumor development has been described<sup>(8)</sup>. The RET proto-oncogene is involved in the molecular evolution of sporadic medullary and PTC, but has also been involved in three subtypes of the inherited cancer syndrome multiple endocrine neoplasia type 2 (MEN2), as well as in Hirschsprung’s disease, and each variant of MEN2 results from different RET gene mutations<sup>(9)</sup>.

The post-genomic era is featured by biological information filling the databases, and large-scale biology projects such as the sequencing of the human genome and gene expression surveys using ribonucleic acid (RNA)-seq, microarrays and other technologies have created a wealth of data for biologists. However, the challenge facing scientists is to analyze and even to access these data to extract useful information pertaining to the system being studied. We researched the RET proto-oncogene sequences in the database using BLAST software, identifying all the proteins encoded in RET proto-oncogene genome, and predicted their structures using domain analysis tools.

A study using bioinformatics tools investigated the structural organization of the extracellular region of the RET receptor tyrosine. It used the BLAST tool with multiple sequence alignments of seven vertebrate sequences and one invertebrate RET sequence delineated four distinct N-terminal domains, each of about 110

residues<sup>(10)</sup>. In our analysis, we carried out extensive BLAST searches of the current databases, and a 3-D model of RET proto-oncogene was built the SWISS-MODEL template library.

The use of proteomic technologies in drug discovery and development is called pharmacoproteomics. Proteomic technologies have contributed to molecular diagnosis, which is a basis of personalized medicine, and the individualized therapy may be based on differential protein expression rather than a genetic polymorphism<sup>(11)</sup>. Currently, several kinds of therapeutic approaches have been developed for the treatment of RET-associated cancers, including tyrosine kinase inhibitors, and molecular models have been potential targets for the development of more selective RET inhibitors.

The NTRK1 proto-oncogene is located on the 1q21-q22 chromosome and encodes the high affinity transmembrane receptor for nerve growth factor. Constitutive activation of NTRK1 has been detected in several tumors types. The fusion of the NTRK1 TK domain to 5’ sequences with at least three different genes (TPM3 gene, TPR gene and TFG gene) leads to oncogenic activation of NTRK1. Somatic rearrangements of NTRK1, producing chimeric oncogenes (an activated version of the proto-oncogene was generated by a somatic intrachromosomal rearrangement fusing the tyrosine kinase domain of NTRK1 with 5’ sequences of the non-muscular tropomyosin gene) with constitutive tyrosine kinase activity, have been frequently found in a consistent fraction of PTC<sup>(12)</sup>. However, PTC arising in patients with a history of exposure to elevated levels of ionizing irradiation does not carry these known abnormalities. In the same way NTRK1 rearrangements are rare in cases of sporadic PTC<sup>(13)</sup>. The frequency of NTRK1 rearrangements in post-Chernobyl papillary thyroid tumors was observed to be equivalent to that in sporadic tumors<sup>(14)</sup>.

Among different genetic factors involved in the pathogenesis of the PTC, rearrangements of NTRK1 proto-oncogene are best known. An explanation for the probability of thyrocytes to undergo gene rearrangements of NTRK1 was proposed when the recombination between RET and H4 was shown to be favored by the loci proximity in interphase nuclei<sup>(15)</sup>.

Reviewing the literature, we observed that in only one study a 3-D structural model of the NTRK1<sup>(16)</sup> was built. In the study, the sequencing analysis of the NTRK1 was evaluated, protein modeling was conducted based on the recent data of NTRK1 structure in the Protein Data Bank, and mutation-related residues were positioned in the 3-D structural model using the PyMOL Molecular Graphics System<sup>(17)</sup>. In our study, the protein analysis of NTRK1 was performed in BLAST, and the 3-D modeling used the

Swiss-PdbViewer suite program, which was adjusted and optimized for alignment between NTRK1 and structural templates. Thus, in order to build a 3-D structural model of NTRK1 we used the Protein Data Bank sequence and a structural homology analysis strategy that recognizes structural similarity among proteins with low sequence identity based on 3-D position-specific scoring algorithms. The increased use of 3-D molecular models should promote advances in drug engineering and could also facilitate the development of new therapies and screening tests.

The tumor suppressor phosphatase PTEN, identified in 1997 as a tumor-suppressor gene located on 10q23.3, is a member of the protein tyrosine phosphatase family and, following activating mutations or amplifications of the genes encoding the effector proteins of PI3K/AKT pathway, inhibits PI3K signaling, thereby reducing the level of activated AKT. Structurally, PTEN protein is composed of an N-terminal dual specificity phosphatase-like enzyme domain and a C-terminal regulatory domain, which binds to phospholipid membranes<sup>(18)</sup>. The dysregulation of the PI3K/AKT signaling pathway contributes to many cancers in humans with antiapoptotic action. Recent findings have demonstrated that PTEN also plays a critical role in deoxyribonucleic acid (DNA) damage repair and DNA damage response<sup>(19)</sup>. The importance of PTEN catalytic activity in its tumor suppressor function is underscored by the fact that the majority of PTEN missense mutations detected in tumor specimens target the phosphatase domain and cause a loss in PTEN phosphatase activity. These data suggest that genetic loss of PTEN is sufficient to induce thyroid cancer *in vivo*.

A study showed that the 3-D structure of PTEN provides a number of insights into the potential mechanisms by which phosphatases recognizes and dephosphorylates its 3-phosphate-containing phospholipid substrates<sup>(20)</sup>, and the current model of PTEN, together with its crystal structure, should prove useful for structural predictions that can be tested in enzymatic assays<sup>(21)</sup>. We conducted NCBI-BLAST database searches using the PTEN domain as a query, which was identified by sequencing, and a high-quality structural model of PTEN was built with the SWISS-MODEL program<sup>(22)</sup>. The 3-D models have also been making important contributions to growing appreciation of oncologic therapeutics, including thyroid cancer.

The TP53 gene is located on chromosome 17p13.1 and consists of 11 exons, coding for a nuclear phosphoprotein that can bind to specific DNA sequences, and acts as a transcription factor. The TP53 is recognized as a tumor-suppressor gene because it encodes protein p53 participating in the processes of cell-cycle arrest in the G1 phase of the cell cycle via DNA repair, and also in apoptosis<sup>(23)</sup>. Inactivation of TP53 in immortalized cells results

in an important instability of chromosome structure, including translocations, deletions, telomeric associations, and ring chromosomes<sup>(24)</sup>. The TP53 gene is frequently affected by loss of alleles and by point mutations in almost all cancers. A principle of this function for cancer progression is provided by the fact that mice without functional TP53 gene breed and develop (almost) normally, but die at an early age from multiple cancers<sup>(25)</sup>. Over 20,000 alterations in TP53 have been discovered in human tumors. The role of TP53 in cancer has been studied, and the presence of a TP53 mutation may be predictive of the tumor response to treatment and patient survival<sup>(26)</sup>.

The TP53 mutation is a common event in many cancers, including thyroid carcinoma, and it also has an effect on infiltration, lymphatic metastasis and prognosis of thyroid carcinoma. The mutations of TP53 are detectable in 15% of malignant thyroid tumors and are associated with the evolution from differentiated to anaplastic carcinoma<sup>(27)</sup>. Thus, inactivation of p53 has been considered a hallmark of advanced thyroid tumors.

A study evaluated the TP53 structure through a 3-D computer model constructed with the NOC program<sup>(28)</sup>. In that study, TP53 (residues 219-292) was modeled with the SWISS-MODEL software using the crystal structure of human TP53 (PDB accession code 2qxa, chain B) as a template<sup>(22)</sup>. Similarly, we built a 3-D structural model of TP53 gene [*Homo sapiens*] based on the high-resolution crystal structure, and this structural model was elaborated using SWISS-MODEL.

Computational 3-D screening has been employed for the discovery of novel small-molecule inhibitors of the MDM2-p53 interaction. It includes 3-D pharmacophore searching, structure-based database searching and the combination of both approaches<sup>(29)</sup>.

There are several classes of PI3Ks, among which class I is the best characterized and most important in human tumorigenesis. The PI3K is a lipid kinase that generates a messenger essential for the translocation of AKT to the plasma membrane<sup>(30)</sup>. AKT is a Ser/Thr kinase, and three types of its isoforms are found in human tissues: AKT-1, AKT-2, and AKT-3, with AKT-1, and AKT-2 being the most abundant and important in thyroid cancer. Activation of AKT plays a pivotal role in cellular functions such as cell proliferation and survival by phosphorylating a variety of substrates<sup>(31)</sup>. The PI3K/AKT signaling pathway plays an important role in transmission of cell signals through transduction systems to cell nucleus, where they influence the expression of genes that regulate crucial cellular processes as: cell growth, proliferation and apoptosis. Therefore, a major consequence of activating PI3K/AKT signaling is the inhibition of cell cycle inhibitors. Genetic and

epigenetic alternations, concerning PI3K/AKT signaling pathways, contribute to their activation and interaction in consequence of malignant cell transformation<sup>(32)</sup>. Thus, the alterations to the PI3K/AKT signaling pathway are frequent in human cancer.

Structural templates for the PI3K domain were identified by searching the PDB database with HHpred using default parameters. A study created alignments from the template structure of human phosphatidylinositol 3-kinase catalytic subunit type 3, which resulted in the highest probability score. Next, alignments were manually converted to FASTA alignments, and 3-D models were built with the SWISS-MODEL workspace using the alignment mode. Surface representations of the PI3K domain were created in PyMOL<sup>(17,33)</sup>. In our study, BLAST searches of current databases were followed by the construction of a 3-D model of PI3K/AKT signaling pathway with the SWISS-MODEL template library.

The CTNNB1 is a dominantly acting cancer gene located on chromosome 3p22-p21.3, composed of 16 exons, which has two main functions in cell regulation: as a cadherin-mediated adhesion regulator and as a mediator of WNT/CTNNB1 signaling<sup>(2)</sup>.

Catenin beta 1 is an oncoprotein, encoded by the CTNNB1 gene, and is responsible for anchoring the cadherin with the cytoskeleton. Loss of cadherin-mediated adhesion and activation of the WNT/catenin beta 1 signaling pathway are important steps in the development and progression of many neoplasms<sup>(34)</sup>.

The basic aspects of the role of catenin beta 1 in malignant transformation have been studied in various tumors. Catenin beta 1 is well-characterized as a major player in canonical WNT signaling, and once activated, it goes into the nucleus and stimulates transcription of target genes. The aberrant activation of WNT/catenin beta 1 signaling related to diverse mutations in catenin beta 1 has been demonstrated in thyroid cancer<sup>(35)</sup>.

We built the 3-D molecular model of CTNNB1 using the homology modeling method based on the CTNNB1 sequence and the high-homology structure as the template with SWISS-MODEL web server based on the protein sequence of CTNNB1. The used computational method can be of significant value for knowledge of thyroid cancer pathogenesis and may provide useful structural insights to facilitate rational drug design with personalized therapies. A study using the IntFOLD2, a recognition method<sup>(36)</sup>, built a full atomic 3-D model for the CTNNB1 sequence, and the phosphorylated tyrosine side chain structure was modeled using the Builder tool in PyMOL<sup>(37)</sup>.

The Pax8 belongs to the mammalian Pax protein family, a group of important developmental regulators featured by the presence of a highly conserved DNA-binding motif of 128 amino

acids<sup>(38)</sup>. The peroxisome proliferator activated receptor gamma (PPARG) is one of the nuclear receptors that play an important role in insulin sensitivity and in adipocyte differentiation<sup>(39)</sup>.

The PAX8/PPARG fusion oncogene was created by a balanced translocation between chromosomes 2 and 3, where the 2q13-qter region is translocated to 3p25, resulting in an in-frame fusion between most of the coding sequence of the thyroid-specific paired-box transcription factor PAX8 (2q13) and the entire translated reading-frame of the gene of the liganded nuclear receptor-family member PPARG (3p25). Thus, the fusion of the PAX8 gene with the PPARG gene results in strong overexpression of the chimeric PAX8-PPARG protein<sup>(40)</sup>.

Studies have shown that PAX8-PPARG may be useful in the diagnosis and treatment of thyroid carcinoma. The PAX8/PPARG fusion oncogene is a common genetic alteration in follicular thyroid carcinoma; it is found in 30%-40% of thyroid follicular carcinoma and in 2%-10% of follicular adenomas, and has been reported of low frequency in papillary thyroid carcinoma with < 5% in non-classical thyroid papillary carcinoma<sup>(41)</sup>. However, the PAX8/PPARG rearrangement in itself may not be sufficient for the development of a malignant phenotype, and additional genetic or epigenetic studies may be required to enable the full phenotypic expression of follicular thyroid carcinoma.

In this study, the amino acid sequence of PAX8/PPARG was retrieved from NCBI sequence viewer, and this sequence was converted to FASTA format. The 3-D structure of PAX8/PPARG as built using SWISS-MODEL. We found no studies demonstrating the 3-D structure of PAX8/PPARG fusion oncogene. Thus, in absence of three-dimensional structures for most of the sequenced protein, homology modeling experimentally forms the basis for the resolution of the structure.

The RAS mutations are common in thyroid cancer; however, the incidence of RAS mutations in thyroid tumors and their frequency in specific histologic types vary widely in different series. The RAS mutations appear to be associated with aggressive tumors and point mutations that have been identified and localized at codons 12, 13 or 61 of three RAS genes (Harvey rat sarcoma viral oncogene: H-RAS, neuroblastoma rat sarcoma viral oncogene: N-RAS, and Kristen rat sarcoma viral oncogene: K-RAS), with mutations of N-RAS and H-RAS at codon 61, and of K-RAS at codon 12/13 being the most common<sup>(42)</sup>. The three members of the RAS gene family (H-RAS, N-RAS, AND K-RAS) were identified more than 25 years ago because of their frequent oncogenic activation in human tumors. Constitutive activation of H-RS, N-RAS, and K-RAS mutations has been reported as marker for aggressive thyroid cancer behavior. However, some studies have shown a similar prevalence of RAS mutations in benign and malignant

thyroid neoplasm, suggesting that RAS activation may represent an early event<sup>(43)</sup>. The RAS signaling through downstream ERK and AKT pathways stimulate epithelial to mesenchymal transition in various cells<sup>(44)</sup>. These data suggest that RAS is involved in both thyroid tumor formation and in thyroid cancer progression.

A study of bio-computational analysis was performed using web-based tools and servers. Multiple sequence alignment of selected human RAS subfamily proteins with other homologous sequences revealed highly conserved regions, with the 3-D structure prediction done by using SWISS-MODEL<sup>(45)</sup> online tools. In the present study, RAS protein sequences were taken from NCBI, and we built the 3-D molecular model of RAS using the homology modeling method based on the RAS sequence with SWISS-MODEL based on the protein sequence of RAS.

The BRAF is located on chromosome 7q24 and encodes a serine-threonine kinase, playing a critical role in cell signaling. It is dependent on the mitogen-activated protein kinase (MAPK) signaling pathway, and has been implicated in human cancers, what occurs in approximately 40%-45% of papillary cancers, 10%-15% of poorly differentiated carcinoma, and 20%-30% of anaplastic carcinoma<sup>(46)</sup>. Thus, BRAF has occupied a central role in thyroid cancer pathogenesis since the original identification of BRAF mutations in papillary and anaplastic thyroid cancer. Activated versions of BRAF can also be generated by intrachromosomal inversions that fuse the kinase domain of BRAF to the NH<sub>2</sub>-terminal portion of A-kinase anchoring protein 9 (AKAP9), resulting in BRAF-AKAP fusion proteins that are similar in structure to the RET/PTC fusion proteins, and are found in about 11% of patients whose thyroid cancers are thought to be caused by the Chernobyl nuclear power station disaster in 1986<sup>(47)</sup>.

There are numerous BRAF mutations, but BRAF V600E, resulting from the BRAFT1799A transversion mutation in exon 15, is the more prevalent, where a substitution of a valine by a glutamic acid at position 600 occurs, and the BRAF mutation is almost exclusively detected with RET/PTC rearrangement<sup>(48)</sup>. A large number of studies have tried to evaluate the relevance and the function of the V600E mutation in controlling oncogenesis and progression of thyroid cancer.

The BRAF is composed of three conserved domains characteristic of the Raf kinase family: conserved region 1, a Ras-GTP-binding self-regulatory domain; conserved region 2, a serine-rich hinge region; and conserved region 3, a catalytic protein kinase domain that phosphorylates a consensus sequence on protein substrates. In its active conformation, BRAF forms dimers via hydrogen-bonding and electrostatic interactions of its kinase domains<sup>(49)</sup>. In 2004, the crystal structure of BRAF in complex with Bay 43-9006 was resolved and revealed a distinctive protein–ligand interaction mode<sup>(50)</sup>.

A study of alignment BRAF performed using the automatic mode of the SWISS-MODEL server built a 3-D molecular model of BRAF<sup>(51)</sup>. In our study, we constructed a 3-D model of BRAF with the SWISS-MODEL template library. Thus, the molecular model may predict diverse BRAF inhibitors for clinical use.

The TSHR is a key protein in the control of thyroid function and a major thyroid autoantigen. TSHR regulates thyroid growth and differentiation at late developmental stages. TSHR molecules in the membrane are quite stable, and signaling in the thyrocyte will be controlled mainly through circulating TSH levels. The mature TSHR is encoded by a single gene with 10 exons. The protein contains two subunits: a large ectodomain also called A subunit encoded by exons 1-8 and binds TSH, and an intracellular domain encoded by exons 9-10 called B subunit that will interact with G proteins to initiate signaling<sup>(52)</sup>.

The expression of TSHR gene is not an infrequent feature in thyroid cancers during the process of dedifferentiation that involves perturbation of several nuclear transcription factors<sup>(53)</sup>.

Somatic mutations in the TSHR gene have been identified in benign and malignant thyroid tumors. As a result of these somatic mutations, the TSHR is continuously activated, what could prompt the overgrowth of thyroid cells. Mutations of the TSHR gene are possibly related to thyroid cancer, and the first case of clear cell follicular carcinoma of the thyroid with next generation sequencing analysis reported showed that gain-of-function mutation of TSHR can overstimulate the thyroid follicular cells as the elevated level of TSH does and might have contributed to the development of clear cell morphology<sup>(54)</sup>.

The structural and functional dimensions of the extracellular TSHR region were studied in detail by mutagenesis<sup>(55)</sup>. The use of the crystal structure of the FSHR/FSH complex enabled extended and structurally supported insights into extracellular hormone binding, signal transduction and organization of the TSHR<sup>(56)</sup>. A study combining sequence and structure of molecular modeling of genetic markers for thyroid cancer has been recently published with the objective to predict whether specific missense variants in the protein kinase region have pathological significance<sup>(57)</sup>. Other studies reported the prediction of phenotypic severity of uncertain gene variants in the proto-oncogene using a computational method<sup>(58)</sup>.

The increase insight into the genetics and molecular biology of cancer has resulted in the identification of an increasing number of molecular markers for anticancer drug discovery and development. Currently, a number of recent successful drugs have partially or totally emerged from a structure-based research approach, and several advances, including crystallography and bioinformatics, are behind these successes<sup>(59)</sup>.

Although the methods of molecular modeling of genetic markers for thyroid cancer by bioinformatics have the advantages of manageable dimensionality, ease of testing, and clinical import, they are ultimately hampered by the complexity of their interaction. However, tools of computational models currently developed are of high impact and will be able to function as the bridge between tumor molecular biology and the individualized treatment of cancer patients.

There is growing evidence giving prominence to computational modeling in the field of biomedicine, in special, applied to the *in silico* analysis of cancer dynamics. Bioinformatics and molecular modeling tools have proved powerful towards the prediction of plausible cytotoxic T cell epitopes as well as other epitopes, cutting down time and cost in cancer treatment<sup>(60)</sup>. Thus, in the era of modern medicine, molecular modeling may allow the discovery of new molecular targets useful for the design of novel cancer drugs. That is because the use of advanced computer models allows the simulation of complex biological processes, providing hypotheses and supporting experimental design.

## CONCLUSION

The structure and function of proteins are determined by their amino acid sequences. The 3-D structure prediction still remains a huge challenge, and there is still a great demand for high-resolution structure prediction methods. Thus, with determination of protein structure and construction of a 3-D model, it will be possible to identify the location of binding sites on proteins of fundamental importance for a wide range of applications, including molecular docking, *de novo* drug design, structure identification, and comparison of functional sites.

This study shows the importance of knowledge of molecular structure of the markers for thyroid cancer through bioinformatics, and consequently, the development of more effective new molecules as alternative tools for thyroid cancer treatment.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## RESUMO

**Introdução:** Os avanços nos estudos de biologia molecular da tireoide não fornecem apenas uma visão sobre as doenças tireoidianas, mas um diagnóstico preciso do câncer de tireoide. **Objetivo:** Realizar um tutorial sobre modelagem molecular proteica dos marcadores genéticos do câncer de tireoide. **Métodos:** As proteínas foram selecionadas utilizando sequência do Banco de Dados de Proteínas e algoritmo basic local alignment search tool (BLAST). As sequências obtidas foram alinhadas com os algoritmos de alinhamento múltiplo Clustal W. Para a modelagem molecular, as estruturas tridimensionais foram geradas a partir deste conjunto com o SWISS-MODEL, um servidor de homologia de modelagem de estrutura proteica totalmente automatizado, acessível por meio do servidor web ExPASy. **Resultados:** Demonstramos a análise das proteínas, a projeção da estrutura molecular e a homologia proteica dos seguintes marcadores moleculares de câncer de tireoide: proto-oncogene receptor tyrosine kinase (RET); proto-oncogene neurotrophic tyrosine kinase receptor 1 (NTRK1); phosphatase and tensin homolog (PTEN); gene tumor protein p53 (TP53); phosphoinositide 3-kinase/threonine protein kinase (PI3K/AKT); catenina beta 1 (CTNNB1); paired box 8-peroxisome proliferator-activated receptor gamma (PAX8-PPARG); rat sarcoma viral oncogene (RAS); B-raf proto-oncogene, serine/threonine kinase (BRAF) e thyroid-stimulating hormone receptor (TSHR). **Conclusão:** Este estudo mostra a importância do conhecimento da estrutura molecular dos marcadores de câncer da tireoide por meio da bioinformática e, conseqüentemente, o desenvolvimento de novas moléculas mais eficazes como ferramentas alternativas para o seu tratamento.

**Unitermos:** análise de sequência de proteína; modelos moleculares; marcadores genéticos; neoplasias da glândula tireoide.

## REFERENCES

1. Buffet C, Groussin L. Molecular perspectives in differentiated thyroid cancer. *Ann Endocrinol (Paris)*. 2015; 76(1 Suppl 1): 1S8-15. PubMed PMID: 26826484.
2. Andrade LJO, Melo PRS, Neto WA, et al. Molecular markers in thyroid cancer. *Rev Argent Endocrinol Metab*. 2013; 50(2): 84-98.
3. Eswar N, John B, Mirkovic N, et al. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res*. 2003; 31(13): 3375-80. PubMed PMID: 12824331.

4. Strausberg RL, Feingold EA, Grouse LH, et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci U S A*. 2002; 99(26): 16899-903. PubMed PMID: 12477932.
5. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25(17): 3389-402. PubMed PMID: 9254694.
6. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 2011; 9(2): 173-5. PubMed PMID: 22198341.
7. Prescott JD, Zeiger MA. The RET oncogene in papillary thyroid carcinoma. *Cancer*. 2015; 121(13): 2137-46. PubMed PMID: 25731779.
8. Eng C. RET proto-oncogene in the development of human cancer. *J Clin Oncol*. 1999; 17(1): 380-93. PubMed PMID: 10458257.
9. Komminoth P. RET proto-oncogene and thyroid cancer. *Endocr Pathol*. 1997 Autumn; 8(3): 235-9. PubMed PMID: 12114728.
10. Lesueur F, Cebrian A, Robledo M, et al. Polymorphisms in RET and its coreceptors and ligands as genetic modifiers of multiple endocrine neoplasia type 2A. *Cancer Res*. 2006; 66(2): 1177-80. PubMed PMID: 16424056.
11. Jain KK. Role of pharmacoproteomics in the development of personalized medicine. *Pharmacogenomics*. 2004; 5(3): 331-6. PubMed PMID: 15102547.
12. Pierotti MA, Greco A. Oncogenic rearrangements of the NTRK1/NGF receptor. *Cancer Lett*. 2006; 232(1): 90-8. PubMed PMID: 16242838.
13. Greco A, Miranda C, Pierotti MA. Rearrangements of NTRK1 gene in papillary thyroid carcinoma. *Mol Cell Endocrinol*. 2010; 321(1): 44-9. PubMed PMID: 19883730.
14. Shaw AT, Hsu PP, Awad MM, Engelman JA. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer*. 2013; 13(11): 772-87. PubMed PMID: 24132104.
15. Roccatto E, Bressan P, Sabatella G, et al. Proximity of TPR and NTRK1 rearranging loci in human thyrocytes. *Cancer Res*. 2005; 65(7): 2572-6. PubMed PMID: 15805251.
16. Gao L, Guo H, Ye N, et al. Oral and craniofacial manifestations and two novel missense mutations of the NTRK1 gene identified in the patient with congenital insensitivity to pain with anhidrosis. *PLoS One*. 2013; 8(6): e66863. PubMed PMID: 23799134.
17. DeLano WL. 2008. The PyMOL molecular graphics system. DeLano Scientific LLC, Palo Alto, CA.
18. Torres J, Pulido R. The tumor suppressor PTEN is phosphorylated by the protein kinase CK2 at its C terminus. Implications for PTEN stability to proteasome-mediated degradation. *J Biol Chem*. 2001; 276(2): 993-8. PubMed PMID: 11035045.
19. Ming M, He YY. PTEN in DNA damage repair. *Cancer Lett*. 2012; 319(2): 125-9. PubMed PMID: 22266095.
20. Lee JO, Yang H, Georgescu MM, et al. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell*. 1999; 99(3): 323-34. PubMed PMID: 10555148.
21. Wu Y, Dowbenko D, Pisabarro MT, Dillard-Telm L, Koeppen H, Lasky LA. PTEN 2, a Golgi-associated testis-specific homologue of the PTEN tumor suppressor lipid phosphatase. *J Biol Chem*. 2001 Jun 15; 276(24): 21745-53. PubMed PMID: 11279206.
22. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006; 22(2): 195-201. PubMed PMID: 16301204.
23. Soussi T, Wiman KG. TP53: an oncogene in disguise. *Cell Death Differ*. 2015; 22(8): 1239-49. PubMed PMID: 26024390.
24. Schwartz JL, Jordan R, Evans HH, Lenarczyk M, Liber H. The TP53 dependence of radiation-induced chromosome instability in human lymphoblastoid cells. *Radiat Res*. 2003; 159(6): 730-6. PubMed PMID: 12751955.
25. Kenzelmann Broz D, Attardi LD. In vivo analysis of p53 tumor suppressor function using genetically engineered mouse models. *Carcinogenesis*. 2010; 31(8): 1311-8. PubMed PMID: 20097732.
26. Odell AF, Odell LR, Askham JM, Alogheli H, Ponnambalam S, Hollstein M. A novel p53 mutant found in iatrogenic urothelial cancers is dysfunctional and can be rescued by a second-site global suppressor mutation. *J Biol Chem*. 2013; 288(23): 16704-14. PubMed PMID: 23612969.
27. Garufi A, D'Orazi V, Crispini A, D'Orazi G. Zn(II)-curc targets p53 in thyroid cancer cells. *Int J Oncol*. 2015; 47(4): 1241-8. PubMed PMID: 26314369.
28. Zheng CX, Gu ZH, Han B, et al. Whole-exome sequencing to identify novel somatic mutations in squamous cell lung cancers. *Int J Oncol*. 2013; 43(3): 755-64. PubMed PMID: 23799614.
29. Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. *Annu Rev Pharmacol Toxicol*. 2009; 49: 223-41. PubMed PMID: 18834305.
30. Franke TF. PI3K/Akt: getting it right matters. *Oncogene*. 2008; 27(50): 6473-88. PubMed PMID: 18955974.
31. Ghosh MK, Sharma P, Harbor PC, Rahaman SO, Haque SJ. PI3K-AKT pathway negatively controls EGFR-dependent DNA-binding activity of Stat3 in glioblastoma multiforme cells. *Oncogene*. 2005; 24(49): 7290-300. PubMed PMID: 16007122.
32. Chang F, Lee JT, Navolanic PM, et al. Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. *Leukemia*. 2003; 17(3): 590-603. PubMed PMID: 12646949.
33. Knutson BA, Hahn S. Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Mol Cell Biol*. 2011; 31(4): 818-31. PubMed PMID: 21149579.
34. Osterheld MC, Bian YS, Bosman FT, Benhattar J, Fontollet C. Beta-catenin expression and its association with prognostic factors in adenocarcinoma developed in Barrett esophagus. *Am J Clin Pathol*. 2002; 117(3): 451-6. PubMed PMID: 11888085.
35. Garcia-Rostan G, Tallini G, Herrero A, D'Aquila TG, Carcangiu ML, Rimm DL. Frequent mutation and nuclear localization of beta-catenin in anaplastic thyroid carcinoma. *Cancer Res*. 1999; 59(8): 1811-5. PubMed PMID: 10213482.
36. Buenavista MT, Roche DB, McGuffin IJ. Improvement of 3D protein models using multiple templates guided by single-template model

- quality assessment. *Bioinformatics*. 2012; 28(14): 1851-7. PubMed PMID: 22592378.
37. Tucci V, Kleefstra T, Hardy A, et al. Dominant  $\beta$ -catenin mutations cause intellectual disability with recognizable syndromic features. *J Clin Invest*. 2014; 124(4): 1468-82. PubMed PMID: 24614104.
38. Vincenzi M, Camilot M, Ferrarini E, et al. Identification of a novel pax8 gene sequence variant in four members of the same family: from congenital hypothyroidism with thyroid hypoplasia to mild subclinical hypothyroidism. *BMC Endocr Disord*. 2014; 14: 69. PubMed PMID: 25146893.
39. Katayama K, Wada K, Nakajima A, et al. A novel PPAR gamma gene therapy to control inflammation associated with inflammatory bowel disease in a murine model. *Gastroenterology*. 2003; 124(5): 1315-24. PubMed PMID: 12730872.
40. Eberhardt NL, Grebe SK, McIver B, Reddi HV. The role of the PAX8/PPAR gamma fusion oncogene in the pathogenesis of follicular thyroid cancer. *Mol Cell Endocrinol*. 2010; 321(1): 50-6. PubMed PMID: 19883731.
41. Fallahi P, Giannini R, Miccoli P, Antonelli A, Basolo F. Molecular diagnostics of fine needle aspiration for the presurgical screening of thyroid nodules. *Curr Genomics*. 2014; 15(3): 171-7. PubMed PMID: 24955024.
42. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res*. 2012; 72(10): 2457-67. PubMed PMID: 22589270.
43. Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer*. 2013; 13(3): 184-99. PubMed PMID: 22589270.
44. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci Signal*. 2014; 7(344): re8. PubMed PMID: 25249658.
45. Jain D, Rawat R, Jatav VK. Phylogenetic analysis of RAS subfamily proteins. *Int J Pharm Sci Res*. 2016; 7(3): 1070-80.
46. Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol*. 2011; 7(10): 569-80. PubMed PMID: 21878896.
47. Williams D. Cancer after nuclear fallout: lessons from the Chernobyl accident. *Nat Rev Cancer*. 2002; 2(7): 543-9. PubMed PMID: 12094241.
48. Tavares C, Melo M, Cameselle-Teijeiro JM, Soares P, Sobrinho-Simões M. Endocrine tumours: genetic predictors of thyroid cancer outcome. *Eur J Endocrinol*. 2016; 174(4): R117-26. PubMed PMID: 26510840.
49. Bollag G, Tsai J, Zhang J, et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov*. 2012; 11(11): 873-86. PubMed PMID: 23060265.
50. Wan PT, Garnett MJ, Roe SM, et al. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell*. 2004; 116(6): 855-67. PubMed PMID: 15035987.
51. Göring S, Taymans JM, Baekelandt V, Schmidt B. Indolinone based LRRK2 kinase inhibitors with a key hydrogen bond. *Bioorg Med Chem Lett*. 2014; 24(19): 4630-7. PubMed PMID: 25219901.
52. García-Jiménez C, Santisteban P. TSH signalling and cancer. *Arq Bras Endocrinol Metabol*. 2007; 51(5): 654-71. PubMed PMID: 17891229.
53. Brabant G, Maenhaut C, Köhrle J, et al. Human thyrotropin receptor gene: expression in thyroid tumors and correlation to markers of thyroid differentiation and dedifferentiation. *Mol Cell Endocrinol*. 1991; 82(1): R7-12. PubMed PMID: 1761161.
54. Tong GX, Mody K, Wang Z, Hamele-Bena D, Nikiforova MN, Nikiforov YE. Mutations of TSHR and TP53 genes in an aggressive clear cell follicular carcinoma of the thyroid. *Endocr Pathol*. 2015; 26(4): 315-9. PubMed PMID: 26260781.
55. Kleinau G, Mueller S, Jaeschke H, et al. Defining structural and functional dimensions of the extracellular thyrotropin receptor region. *J Biol Chem*. 2011; 286(25): 22622-31. PubMed PMID: 21525003.
56. Krause G, Kreuchwig A, Kleinau G. Extended and structurally supported insights into extracellular hormone binding, signal transduction and organization of the thyrotropin receptor. *PLoS One*. 2012; 7(12): e52920. PubMed PMID: 23300822.
57. George Priya Doss C, Rajith B, Chakraborty C, et al. In silico profiling and structural insights of missense mutations in RET protein kinase domain by molecular dynamics and docking approach. *Mol Biosyst*. 2014; 10(3): 421-36. PubMed PMID: 24336963.
58. Crockett DK, Piccolo SR, Ridge PG, et al. Predicting phenotypic severity of uncertain gene variants in the RET proto-oncogene. *PLoS One*. 2011; 6(3): e18380. PubMed PMID: 21479187.
59. Geromichalos GD. Importance of molecular computer modeling in anticancer drug development. *J BUON*. 2007; 12 Suppl 1: S101-18. PubMed PMID: 17935268.
60. Mishra S, Sinha S. Immunoinformatics, molecular modeling, and cancer vaccines. *Methods Mol Biol*. 2014; 1184: 513-21. PubMed PMID: 25048143.

**CORRESPONDING AUTHOR****Luis Jesuino de Oliveira Andrade**

Rua Nações Unidas, 511; Centro; CEP: 45600-673; Itabuna-BA, Brasil; e-mail: luis\_jesuino@yahoo.com.br