

79603\_Auto\_Edited.docx

***Basic Study*****Three-dimensional model of antigens with serodiagnosis potential for leprosy: an in silico study**

Bianca Luiza Melo de Assis, Rafaela Viana Vieira, Ian Theodoro Rudenco Gomes Palma, Matheus Bertolini Coutinho, Juliana de Moura, Gabrielle Caroline Peiter, Kádima Nayara Teixeira

**Abstract****BACKGROUND**

Leprosy is a disease caused by *Mycobacterium leprae*, an intracellular pathogen that has tropism and affects skin and nervous system cells. The disease, simplified, has two forms of presentation: paucibacillary and multibacillary, with different clinical and immunological manifestations. Unlike what occurs in the multibacillary form of the disease, the diagnostic tests for the paucibacillary form are nonspecific and not very sensitive, allowing the existence of infected individuals without treatment, which contributes to the spread of the pathogen in the population. To mitigate this contamination, more sensitive diagnostic tests capable of detecting paucibacillary patients are needed.

**AIM**

Predict the three-dimensional structure models of *M. leprae* antigens with serodiagnostic potential for leprosy.

**METHODS**

In this *in silico* study, a selection of satisfactory templates was made in the Protein Data Bank (PDB) using BLAST (Basic Local Alignment Search Tool) to predict the structural templates of ML2038, ML0286, ML0050 and 85B antigen by comparative modeling. The templates were selected according to general criteria such as sequence identity, coverage, X-ray resolution, GMQE value and phylogenetic relationship; Clustal X 2.1 software was used in this analysis. Molecular modeling was done using the software Modeller 9v13. Visualization of the models was possible using ViewerLite 4.2 and PyMol software, and analysis of the quality of the predicted models was performed using the QMEAN and Z-score. Finally, the three-dimensional models were validated using the MolProbity and Verify 3D platform.

## RESULTS

The three-dimensional structure models of ML2038, ML0286, ML0050 and 85B antigen of *M. leprae* were predicted using the templates PDB: 3UOI (90.51% identity), PDB: 3EKL (87.46% identity), PDB: 3FAV (40.00% identity) and PDB: 1F0N (85.21% identity), respectively. The QMEAN and Z-score values indicated the good quality of the structure models. These data refer to the monomeric units of antigens, since some of these antigens have quaternary structure. The validation of the models was performed with the final three-dimensional structure - monomer (ML0050 and 85B antigen) and quaternary structures (ML2038 and ML0286). A majority of amino acid residues were observed in favorable and allowed regions in the Ramachandran plot, indicating correct positioning of the side chain and absence of steric impediment. The MolProbity score value and Verify 3D results of all models indicated a satisfactory prediction.

## CONCLUSION

The polarized immune response against *M. leprae* creates a problem in leprosy detection. The selection of immunodominant epitopes is essential for the development of more sensitive serodiagnostic tests, for this it is important to know the three-dimensional structure of the antigens, which can be predicted by bioinformatics tools.

## **INTRODUCTION**

Leprosy is a chronic transmissible and infectious disease caused by the bacillus *Mycobacterium leprae*, which persists as a major public health problem in Brazil. The etiological agent of this disease mainly affects the skin, peripheral nerves and eyes, presents slow evolution and causes deformity and physical disability, when not correctly diagnosed and treated<sup>[1]</sup>.

According to the World Health Organization (WHO)<sup>[1]</sup>, in 2019, there were 202,185 cases of *M. leprae* infection worldwide, 93% of which were reported in the Americas. In addition, according to the WHO, Brazil contributed 27,864 new notifications of the disease in 2019, placing it in second place among the countries with the highest number of leprosy cases, behind only India.

Clinically, leprosy may manifest in two different forms, which differ mainly by the immune response developed by the host against the pathogen. The current classification of leprosy comes from 1982, from the WHO Committee, which proposed a simplified and operational classification of paucibacillary and multibacillary individuals, based on the likely relationship between clinical form and smear<sup>[2]</sup>.

The manifestations of leprosy are distinguished both by clinical presentations and characterization of the immune response. Such presentations are considered antagonistic, which guides the understanding of the pattern of dual response observed between Th1 and Th2 Lymphocytes. In leprosy, the paucibacillary form is characterized by the Th1 response, which confers some resistance of the organism to the pathogen, thereby this is a milder form of the infection; the multibacillary form, a more severe infection, is characterized by the development of the Th2 pattern, which does not have an effective mechanism against the pathogen<sup>[3]</sup>.

Due to the variety of clinical manifestations of leprosy, WHO<sup>[4]</sup> recommends that the diagnosis be based on loss of sensitivity of a hypopigmented or reddish skin area, visualization of thickening or augmentation of a peripheral nerve, accompanied by weakness of muscle tissue supplied by the nerve - which can only be observed - or

visualization of acid-alcohol fast bacilli (AAFB) by means of the intradermal sample smear technique. In addition, ELISA and PCR tests can be used to complement the diagnosis, especially in suspected multibacillary form. Thus, it is noted that the recognition of the disease can be based both on factors arising from the clinical examination (anamnesis and physical examination) as well as from complementary tests, requiring the expertise of the professional who will perform and interpret the procedure, especially with regard to histological techniques, and reliability of the tests available.

However, considering the paucibacillary form, the tests used to diagnose the multibacillary form, such as serological tests by ELISA, become ineffective, since the paucibacillary patient presents low titers of antibodies, which are not detected. Due to the high production of pro-inflammatory cytokines, there is little bacterial proliferation, which makes it difficult to find bacilli at the site of the lesion in the intradermal smear technique. Consequently, when applied to patients with paucibacillary leprosy, the recommended tests have low sensitivity and specificity, which may lead to false-negative results, contributing to the spread of the disease<sup>[5]</sup>. Thus, tests aimed at improving these parameters and diagnosing both severe and mild forms of leprosy are necessary.

To support the development of diagnostic tests with greater sensitivity for any type of clinical manifestation of leprosy, Santana<sup>[6]</sup> mapped epitopes of *M. leprae*, constructed by the spot synthesis technique, using serum from leprosy patients, based on a previous study<sup>[7]</sup>. Of 12 selected proteins, seven were promising candidates for immunodominant antigens. Bioinformatics tools were used to verify molecular mass, isoelectric point, hydrophobicity and acid-basic characteristics<sup>[5]</sup>, however, none of the proteins has experimentally resolved three-dimensional structure, which hinders a complete biochemical characterization.

The three-dimensional structure of a protein is fundamental to determine the selection of the best candidates in order to obtain a functional epitope, so that the peptides hidden in the protein core are not of interest, since the antibodies do not have access to

this region. In turn, antigenic determinants actively participate in antigen-antibody binding, which, when established, generate conformational and structural changes, being such important structures in the process of antigenic neutralization and processing by the immune system<sup>[6]</sup>.

In this study, *M. leprae* antigens selected by Santana<sup>[6]</sup> had their three-dimensional structures predicted by bioinformatics tools in order to provide support for the elaboration of possible serological diagnostic tests for both multibacillary and paucibacillary clinical forms of leprosy.

## **MATERIALS AND METHODS**

### Template selection

The FASTA sequences of antigens ML2038, ML0286, ML0050 and 85B antigen - GenBank WP\_010908683.1, WP\_010907650, WP\_010907488.1 and CAA43269.1, respectively, available at The National Center for Biotechnology Information (ncbi.nlm.nih.gov) were used to select templates for molecular modeling. The selection was made at Protein Data Bank (PDB) (rcsb.org), through local alignment using the BLAST (Basic Local Alignment Search Tool) <sup>[8]</sup>, aiming to find templates with (1) three-dimensional structures resolved, (2) high sequence identity and coverage values as much as possible, (3) high X-ray resolution, (4) phylogenetic proximity to the target proteins. The choice of templates also took into account the GMQE (Global Model Quality Estimate) value, which predicts the overall quality of the model and ranges from 0 to 1, with higher numbers indicating higher quality. The identity between primary sequence and template, as well as sequence coverage, were verified through global alignment using ClustalX 2.1 software <sup>[9]</sup>.

### Molecular modeling and validation

The antigens whose template was satisfactory were modeled by the comparative modeling methodology using the Modeller 9v13 <sup>[10]</sup> software. The selection of the models released by Modeller was performed using the DOPE (Discrete Optimized

Protein Energy) method, which predicts the lowest energy models, therefore the most stable ones. The selected three-dimensional models were visualized and analyzed by the ViewerLite 4.2 (Accelrys Inc.) and PyMol (Schrödinger Inc.) softwares and the quality was verified by the QMEAN <sup>[11]</sup> and Z-score. The validation of the models was performed using the MolProbity <sup>[12]</sup> and Verify 3D platforms <sup>[13]</sup>.

## **RESULTS**

Regarding the effector mechanisms of the immune system for mycobacteria and other pathogens, it would be feasible that surface antigens and those secreted would be better candidates for serodiagnostic tests, since these would be more exposed to the components of the immune system. However, this placement needs to be better evaluated, at least in relation to *M. leprae*, in which a polarity of response is observed. Therefore, it is important to know the biological functionality of the antigen and also its three-dimensional structure in order to be able to evaluate the localization of epitopes within the structure, and thus try to understand the behavior of the antibody response against *M. leprae*, which is exploited by serological tests. In this context, the prediction of satisfactory structural models is able to collaborate with the development of these tests.

## **DISCUSSION**

The infection caused by the *M. leprae* - the leprosy - causes marked lesions on the skin, peripheral nerves and when not properly treated leads to irreversible and disfiguring sequelae in the patient, such as destruction of cartilage and leonine facies <sup>[15]</sup>. In addition, the great social stigma related to the disease is added, which contributes even more to the worsening of the psychological issue of these patients <sup>[16]</sup>. It is also worth noting the high prevalence of this pathology in the Americas, especially in Brazil, which in 2017 was responsible for 92.3% of the cases reported on the continent, evidencing the clinical and epidemiological importance of the pathology for the country <sup>[16]</sup>. Thus, there is a need for tests that identify the disease early, to prevent its spread and enable the

early institution of the recommended therapy, avoiding or reducing the associated sequelae.

However, variations in the clinical manifestation of leprosy - paucibacillary/tuberculoid form with predominance of Th1 response, better combating of pathogen and lower presence of lesions, and multibacillary/Lepromatous with Th2 response, greater dissemination of pathogen in the body and greater number of lesions - hinders the development of specific and sensitive diagnostic tests for its identification. Moreover, as there is little presence of initial signs and symptoms, and there may be asymptomatic presentation and similarity with other diseases, the purely clinical diagnosis of leprosy is impaired<sup>[5, 16]</sup>.

In addition, there is also the difficulty of growing *M. leprae* in artificial or cellular media, constituting one of the major obstacles to leprosy research. The available forms of Hansen's bacillus cultivation are inoculation in captive armadillo and inoculation in the paw pads of immunocompetent mice (Shepard method) and immunodeficient mice (Prabhakaran method). The use of armadillos for leprosy research is limited by the difficult management of the animal, while research with different types of mice is extensive, taking up to eight months for adequate immunological manifestation<sup>[17-19]</sup>.

Therefore, in silico studies aimed at the identification and mapping of B-cell epitopes <sup>[20]</sup> which predict the type of interaction and the affinity energy between the epitope and the Fab region of the antibodies, are valid for the development of serological tests for the diagnosis of leprosy. This kind of study allows the recognition of immunodominant protein domains and their selection for the preparation of immunohistochemical tests, as well as for the manufacture of synthetic peptides, which are low cost and easy to manipulate<sup>[16,21]</sup>. Thus, there is a decreased possibility of cross-reactivity with antigens from other pathogens, increasing the specificity and sensitivity of the test.

This study aimed to predict the three-dimensional structure models of *M. leprae* antigens that, according to previous studies <sup>[6,7]</sup>, have the potential to be used in serodiagnostic tests for both paucibacillary and multibacillary leprosy. One of these antigens, ML2038, is the Bacterioferritin protein or Major Membrane Protein II (MMP-



II), which is transcribed by the BFRA gene. This protein is present in the cell membrane of *M. leprae* and is responsible for iron storage in restricted situations. In addition, MMP-II has a high identity with its homologue found in *M. tuberculosis* and has the ability to activate TCD4+ and TCD8+ cells [6,16,22].

Due to the high primary sequence identity of the template (91.14%), it was possible to predict the three-dimensional structure of the ML2038 antigen by comparative modeling. The quality of the model was evaluated by the QMEAN score - a parameter that aims to analyze the structure obtained and compare it with others already known, in relation to physicochemical and evolutionary characteristics, in order to identify problematic regions for subsequent correction [23, 24].

The QMEAN score of the three-dimensional structure of ML2038 was 0.56. This value was applied to the Z-score plot, a tool that combines the QMEAN score with those of proteins with structure deposited in the PDB, demonstrating that the model obtained by the study is similar to what is expected for native proteins with molecular mass similarity [24, 25].

From these data, it was possible to correct the errors found in the configuration of amino acid residues using tools of the MolProbity platform, for subsequent validation of the three-dimensional model. Thus, correcting the outliers, the MolProbity score equal to 1.31 was obtained. This score evaluates the log of the clashscore, percentage of amino acids in unfavorable regions in the Ramachandran chart and bad rotations of the lateral chains; MolProbity score values lower than the crystallographic resolution of the template (1.90 Å) indicate three-dimensional model quality[26].

The results of the Ramachandran plot also contribute to the reliability of the three-dimensional organization of the ML2038 model, which predicts that 98.48% of residues are in permitted regions, indicating the stability of the predicted structure.

In a study conducted by Maeda *et al* [25], in which sera from paucibacillary and multibacillary patients, sera from patients with tuberculosis and sera from healthy individuals vaccinated with BCG (*Bacillus of Calmette-Guérin*) were tested, a better sensitivity of the serological test was observed for individuals with the multibacillary

form (82.4%, 95.0% CI; 71.8 - 90.3) than those with the paucibacillary form of leprosy (39.0%, 95.0% CI 28.8 - 50.1), when screened for antibodies against MMP-II. Furthermore, healthy individuals or individuals with tuberculosis showed high specificity, indicating the low influence of homologous *M. bovis* antigens, and ancestry between *M. leprae* and *M. tuberculosis* on the diagnostic test result.

Similar results were found by Kau *et al*<sup>[27]</sup>, in which paucibacillary and multibacillary patients tested positive - 47.6% and 85.1%, respectively, when investigated for antibodies against ML2038, and 20.2% and 57.0% when investigated for antibodies against PGL-I (Phenolic Glycolipid-I), one of the first isolated *M. leprae*-specific antigens currently instituted in serological tests for leprosy<sup>[28]</sup>.

In this way Tsukamoto *et al*<sup>[29]</sup> conducted clinical research that associated the search for anti-MMP-II and anti-MMP-I antibodies. This association allowed the increase in the sensitivity of the test and the rescue of patients with false-negative results obtained when researched for antigens purely against ML2038. Thus, it is perceived that the literature converges in relation to the possibility and advantage of the use of Bacterioferritin in the diagnostic test of leprosy patients.

In the face of these studies, Santana<sup>[6]</sup> points out, through a study that aimed to identify antigen-antibody recognition for some antigens of the *M. leprae*, that the ML2038, as well as ML0286 and ML0050, has no reactivity to the serum from leprosy patients. The author points out that the possible reasons for this non-recognition would be the presence of conformational immunodominant protein domains, which depend on the novel structure of the protein to act as epitopes, and the variability of sera from patients of different ethnicities and regions.

The ML0286 antigen (Putative Fructose Bisphosphate Aldolase), encoded by the FBA gene, is a cytosolic enzyme, soluble and integral to the energy metabolism of *M. leprae*. This enzyme<sup>1</sup> catalyzes the conversion of fructose 1-6-diphosphate into glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate by metabolizing this sugar. Puckett *et al*<sup>[30]</sup> suggested that this enzyme regulates the glycolytic and gluconeogenic metabolism of carbohydrates in mycobacteria, with indications of participation in aminoglycoside

resistance in strains resistant to these antibiotics. The overexpression of Rv0363c, another classification for this aldolase, can maintain the energy supply for resistant *M. leprae* strains, providing ATP during the state of antimicrobial resistance. Homology modeling revealed that ML0286 is a homothetramer, with 96.47% of residues in favorable regions, and is therefore feasible for the elaboration of serological tests for recognition of the protein epitope [7,32,33].

The ML0050 antigen, also called ESAT-6-like protein *esxB* or Culture filtrate protein (CFP-10), acts on the virulence and pathogenicity of *M. leprae* together with ESAT-6 protein, which is secreted by the ESX-1 secretion system and improves the stability of other externalized antigens [16,19,34]. In addition, ML0050 is encoded by the *esxB* gene, having recognized its expression only in the *Mycobacterium* genus [19]. For such evolutionary conservation it is understandable the existence of homology among such proteins in distinct species of mycobacteria, such as between *M. leprae* and *M. tuberculosis*, increasing the chances of cross-reactivity in serological tests.

According to Geluk *et al* [35] although the specificity of ML0050 is high for the *Mycobacterium* genus, the absence of differentiation between both paucibacillary and multibacillary leprosy is a problem to the use of serological tests based on this antigen. Moreover, even if the author points out the properties of this antigen as a measure of disease progression and effectiveness of leprosy treatment, it is noted that there are obstacles to its use.

The 85B antigen (MPT59), also called ML2028 or Diacylglycerol acetyltransferase is encoded by the FBPB gene, having function in cell wall synthesis through its Mycolyltransferase activity [16,36]. Studies have shown that the 85B antigen induces the proliferation and release of high levels of IFN-Gamma in T lymphocyte cultures of immunized mice. According to Spencer *et al* [7] ML2028 may represent a biomarker of disease progression because a patient who developed leprosy had the strongest response to the antigen about 19 mo before clinical diagnosis, due to this incubation time and delay in epitope reactivity, the author reports that this characteristic is dependent on bacterial burden.

The molecular modeling of the 85B antigen was possible due to the similarity between both homologous proteins of *M. leprae* and *M. tuberculosis*. Thus, the identity rate between the two reaches 85.21%, with the antigen being 11 beta-sheet and 6 alpha-helix regions. According to Santanal<sup>[6]</sup>, the external location of reactive peptides facilitates the action of antibodies in the recognition of the antigenic target.

According to Serafin-Lopez<sup>[37]</sup> the specificity of 85B antigen reaches 100%, so that control patients or tuberculosis patients showed no reactivity against the antigen. The results also showed a high degree of antigenicity in leprosy patients, regardless of the clinical classification, and is therefore an important candidate for serological markers.

## **CONCLUSION**

### Selected templates

Templates have been selected for target proteins ML2038, ML0286, ML0050 and 85B antigen - 3UOI (90.51% identity and 1.90 Å resolution), 3EKL (87.46% identity and resolution of 1.51 Å), 3FAV (40.00% identity and 2.15 Å resolution) and 1F0N (85.21% identity and 1.80 Å resolution), respectively. In addition, 100.00% coverage was obtained for ML2038 and ML0050 proteins, 99.00% for ML0286 and 87.00% for 85B antigen (Table 1).

### Comparative molecular modeling

The primary sequences of the target proteins were submitted to comparative molecular modeling using the Modeller 9v13 software which generated the three-dimensional structures shown in Figure 1 and Figure 2. The comparative modeling of ML0050 was performed with a pipeline designed for low identity templates <sup>[14]</sup>. QMEAN scores for the structural models were 0.13 (ML2038), 0.56 (ML0286), -0.62 (ML0050), -0.39 (85B antigen); this score typically ranges from 0 to 4.0, and values closer to 0 indicate a good model. QMEAN score is referring to the monomeric subunits that were modeled, however, the ML2038 and ML0286 antigens are quaternary proteins. ML0050 can dimerize, but functionally, the protein is a monomer. Validation of the predicted

structural models was performed with the functional structure of each antigen, i.e., ML0050 and 85B antigen in monomeric form and ML2038 and ML0286 in quaternary form. The Z-score of all structure models were  $0.5 < \text{Z-score} < 1.0$ . Figure 3 shows the Z-score plot which indicates the quality of the three-dimensional models obtained.

#### Validation of three-dimensional models

Validation was performed using MolProbity software; parameters related to the geometry of the angles of the chemical groups of amino acid residues were evaluated in the models obtained by molecular modeling. The MolProbity score, outlier residues and the percentage of residues in favorable regions in the Ramachandran plot were considered (Table 2). On Verify3D software, 3D/1D score for all models was more than 80%; it means that the models have good quality.

### **ARTICLE HIGHLIGHTS**

#### ***Research background***

The research perspective is to continue the study - map the epitopes, evaluate them through experimental studies.

#### ***Research motivation***

The structural models of Mycobacterium leprae antigens are considered high-quality models by validation parameters and can be used for the mapping of epitopes candidates for serodiagnostic tests.

#### ***Research objectives***

The results obtained in this study were considered of good quality; no important parameters, such as steric impediment and lack of stability were observed. Therefore, for the purpose for which they are intended - research of immunodominant epitopes, the models were considered satisfactory.

### ***Research methods***

The methods used were classic methods of bioinformatics, but very well founded and with proven reliability. Comparative modeling is the simplest methodology of molecular modeling, which can be used in this study due to antigen conditions. Once the input data is well filtered, the results are very satisfactory - this can be proven by the similarity of the structures with the homologous ones.

### ***Research results***

The aim of this study was to provide reliable three-dimensional structure models for the analysis of immunodominant epitopes that can be tested later, in the form of synthetic peptides, as possible candidates for the development of diagnostic tests that can detect patients with paucibacillary leprosy. The structure and location of the epitope within the antigen structure is important to understand the behavior of the humoral response of patients.

### ***Research conclusions***

This study is motivated by our group's previous studies on diagnostic tests for leprosy. Promising and relevant data have been achieved in experimental research with patient serum and have been confirmed by bioinformatics analyses. This study predicting *Mycobacterium leprae* antigen models is only one step towards future research on the development of more sensitive diagnostic tests for leprosy.

### ***Research perspectives***

Our research group always tries to use bioinformatics tools as allies of experimental research in order to corroborate and confirm data. So far the results have been very satisfactory and have saved research time and financial costs, many times.

ORIGINALITY REPORT

1 %

SIMILARITY INDEX

PRIMARY SOURCES

1	<a href="http://www.researchsquare.com">www.researchsquare.com</a> Internet	15 words — < 1 %
2	<a href="http://f6publishing.blob.core.windows.net">f6publishing.blob.core.windows.net</a> Internet	13 words — < 1 %

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS