INTRODUCTION

Viral infections are often cited as a major environmental factor involved in autoimmune thyroid diseases (ATD). It remains to determine whether they are responsible for ATD or whether they are just innocent bystanders.

Exposure to viral antigens that share amino acid (AA) sequence similar with self-antigens might trigger autoimmune diseases in genetically predisposed individuals and the molecular mimicry theory suggests that epitope mimicry between the virus and human proteins can activate autoimmune disease. Objectives - The purpose of this study is to explore the possible sequence similarity between the amino acid sequences of thyroid self-protein and hepatitis C virus proteins, using databanks of proteins and immunogenic peptides, to explain autoimmune thyroid disease. Methods - Were performed the comparisons between the amino acid sequence of the hepatitis C virus polyprotein and thyroid self-protein human, available in the database of National Center for Biotechnology Information on Basic Local Alignment Search Tool. Results - The sequence similarity was related each hepatitis C virus genotype to each thyroid antigen. The similarities between the thyroid and the viral peptides ranged from 21.0% (31 identical residues out of 147 amino acid in the sequence) to 71.0% (5 identical residues out of 7 amino acid in the sequence). Conclusion - Bioinformatics data, suggest a possible pathogenic link between hepatitis C virus and autoimmune thyroid disease. Through of molecular mimicry is observed that sequences similarities between viral polyproteins and self-proteins thyroid could be a mechanism of induction of crossover immune response to self-antigens, with a breakdown of self-tolerance, resulting in autoimmune thyroid disease.

Bioinformatics program currently available to construct the structure analyses of HCV component provides an essential framework for virus assembly and understanding of your molecular mechanisms. Similarly, molecular modeling has facilitated the understanding of molecular mimicry as a result of cross-immune response to similar bacterial or viral antigen with thyroid self-protein human.

Until recently, little structure information about HCV protein and thyroid self-protein human were available and molecular models has been built on partial structures, with assembly guided by biochemical data. The purpose of this study is to explore the possible sequence similarity between the AA sequences of thyroid self-protein, which are potential B- and T-cell epitopes of these antigens and proteins of HCV, using databanks of proteins and immunogenic peptides to explain ATD.

**METHODS**

Were performed the comparison between the AA sequence of the HCV poly-protein and thyroid self-protein human, available in the database of NCBI (National Center for Biotechnology Information) on Basic Local Alignment Search Tool (BLAST2p)\(^9\). The 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 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.

Was used expect BLAST program for performing the comparisons. The expect value is a parameter that describes the number of hits one can “expect” to see by chance when searching a database of a particular size (expect value considered as statistical significance <0.05). It decreases exponentially as the score of the match increases. The lower expect value, or the closer it is to zero, the more “significant” the match. However, identical short alignments have relatively high E values. This is because the calculation of the E value takes into account the length of the query sequence. These high expect value make sense because shorter sequences have a higher probability of occurring in the database purely by chance. The expect value can also be used as a convenient way to create a significance threshold for reporting results.

**Sequence Similarity searches - Protein database search and analysis**

One protein-protein sequence alignment method, BLAST2p program, was used to search for sequence similarity between HCV and thyroid self-proteins. This method is limited to searching for linear epitope similarities, which will miss three-dimensional conformational similarities and possible cross-reactivity between proteic and non-proteic epitopes, since most of molecular mimicry is likely to involve T-cell mediation, and T cells generally recognize linear peptides 8–20 AA in length.

**HCV genotypes examined**

Were evaluated the following HCV genotypes, with the respective NCBI sequence identification number (GI): hepatitis C virus subtype 1a (HCV1a) (GI:212675089), hepatitis C virus subtype 1b (HCV1b) (GI:190684416), hepatitis C virus subtype 1c (HCV1c) (GI:81992806), hepatitis C virus subtype 2a (HCV2a) (GI:557164188), hepatitis C virus subtype 2b (HCV2b) (GI:312299482), hepatitis C virus subtype 2c (HCV2c) (GI:395147275), hepatitis C virus subtype 2d (HCV2d) (GI:845452), hepatitis C virus subtype 2e (HCV2e) (GI:295684768), hepatitis C virus subtype 2f (HCV2f) (GI:295684656), hepatitis C virus subtype 3a (HCV3a) (GI:75566979), hepatitis C virus subtype 3b (HCV3b) (GI:290352647), hepatitis C virus subtype 3g (HCV3g) (GI:395147253), hepatitis C virus subtype 4a (HCV4a) (GI:110294850), hepatitis C virus subtype 4c (HCV4c) (GI:1369771), hepatitis C virus subtype 4d (HCV4d) (GI:576294948), hepatitis C virus subtype 4f (HCV4f) (GI:156469741), hepatitis C virus subtype 5a (HCV5a) (GI:262036445), hepatitis C virus subtype 6a (HCV6a) (GI:324330489), hepatitis C virus subtype 6b (HCV6b) (GI:295702049), hepatitis C virus subtype 6c (HCV6c) (GI:82617876), and hepatitis C virus subtype 6n (HCV6n) (GI:82617878).

The following proteins of the HCV have been examined: C (capsule, core protein), E1 (envelope glycoprotein 1), E2 (envelope glycoprotein 2, NS1), p7, NS2 (non-structural protein 2), NS3 (non-structural protein 3, protease/helicase), NS4a (non-structural protein 4a), NS4b (non-structural protein 4b), NS5a (non-structural protein 5a), NS5b (non-structural protein 5b, RNA polymerase), RNA_dep_RNAP (RNA-dependent RNA polymerase), HELICc (Helicase superfamily c-terminal domain), DEXDc (DEAD-like helicases superfamily), Peptidase_S29 (hepatitis C virus NS3 protease), and RdRP_3 (Viral RNA dependent RNA polymerase).

**Tissue-specific antigens examined of thyroid gland**

Were evaluated the following tissue-specific thyroid self-protein: thyroglobulin (Tg) (GI:55770862), and thyroid peroxidase (TPO) (GI:332078490), and thyrotropin receptor (TSHr) (GI:5705928).

The following proteins of the Tg polypeptide have been examined: TY superfamily (Thyroglobulin type I repeats), CO esterase (Carboxylesterase family), and GCC2_GCC3 (trinucleotide repeat polymorphism). The following proteins of the TPO polypeptide have been examined: EGF_CA (Calcium-binding EGF-like domain), and CCP (complement control protein). The following proteins of the TSHr polypeptide have been examined: 7tm_2 super family (7 transmembrane receptor - Secretin family), 7tm_1 (7 transmembrane receptor - rhodopsin family), and LRR_5 (Leucine rich repeats).

**Exposed protein databases**

Regarding HCV1a, the ID contained 3023 AA protein sequences\(^19\). In the case of HCV1b, the ID had 3010 AA protein sequences\(^19\). The HCV1c study allowed us to obtain...
an ID that comprised 3011 AA protein sequences (18). Concerning HCV2a, we were able to assemble an ID of 3033 AA protein sequences (21). The HCV2b, the ID contained 191 AA protein sequences (29). In the case of HCV2c, the ID had 3033 AA protein sequences (27). The HCV2d, the ID contained 149 AA protein sequences (34). The HCV2e study allowed us to obtain ID of 119 AA protein sequences (33). In the case of HCV3a, we were able to assemble an ID of 3021 AA protein sequences (19). The HCV3b study allowed us to obtain an ID that comprised 111 AA protein sequences (31). The HCV3g study allowed us to obtain an ID that comprised 3022 AA protein sequences (27). In the case of HCV4a, the ID had 3009 AA protein sequences (13). The HCV4c, the ID contained 319 AA protein sequences (30). In the case of HCV4d, the ID had 3006 AA protein sequences (36). The HCV4f study allowed us to obtain an ID that comprised 3010 AA protein sequences (18). The HCV5a, the ID contained 186 AA protein sequences (22). The HCV6a study allowed us to obtain an ID of 3015 AA protein sequences (36). The HCV6f study allowed us to obtain an ID that comprised 3015 AA protein sequences (22). The HCV6n, the ID contained 3015 AA protein sequences (22).

**Tissue-specific antigens of thyroid gland**

Regarding Tg, the ID contained 2768 AA protein sequences (19). In the case of TPO, the ID had 933 AA protein sequences (11). In the case of TSHr, the ID had 253 AA protein sequences (9). The HCV3b study allowed us to obtain an ID of 3033 AA protein sequences (27). The HCV2b, the ID contained 191 AA protein sequences (29). The HCV2c, the ID had 3033 AA protein sequences (27). The HCV2d, the ID contained 149 AA protein sequences (34). The HCV2e study allowed us to obtain ID of 119 AA protein sequences (33). Concerning HCV2a, we were able to assemble an ID of 3021 AA protein sequences (19). The HCV3b study allowed us to obtain an ID that comprised 111 AA protein sequences (31). The HCV3g study allowed us to obtain an ID that comprised 3022 AA protein sequences (27). In the case of HCV4a, the ID had 3009 AA protein sequences (13). The HCV4c, the ID contained 319 AA protein sequences (30). In the case of HCV4d, the ID had 3006 AA protein sequences (36). The HCV4f study allowed us to obtain an ID that comprised 3010 AA protein sequences (18). The HCV5a, the ID contained 186 AA protein sequences (22). The HCV6a study allowed us to obtain an ID of 3015 AA protein sequences (36). The HCV6f study allowed us to obtain an ID that comprised 3015 AA protein sequences (22). The HCV6n, the ID contained 3015 AA protein sequences (22).

**RESULTS**

The sequence similarity was related each HCV genotype to each thyroid antigen. The similarities between the thyroid and the viral peptides ranged from 21.0% (31 identical residues out of 147AA in the sequence) to 71.0% (5 identical residues out of 7 AA in the sequence).

The AA sequences identities between the Tg, TPO, TSHr and the HCV1a poly-protein showed 60%, 33% and 30% sequences similarities respectively. The sequences similarities are shown in Figure 2. AA appears in standard single letter code. + indicates conserved or semi-conserved substitutions.

The AA sequences identities between the Tg, TPO, TSHr and the HCV1b poly-protein showed 35%, 33% and 58% sequence similarities respectively. The sequences similarities are shown in Figure 3. AA appears in standard single letter code. + indicates conserved or semi-conserved substitutions.

**FIGURE 1.** Sequence analysis and homology modeling og Tg, TPO and TSHr. Structure of the membrane domain homology model. A) Tg; B) TPO; C) TSHr. Source http://www.ncbi.nlm.nih.gov/protein

**FIGURE 2.** The AA sequences homology between HCV1a and Tg; TPO; TSHr.

**FIGURE 3.** The AA sequences homology between HCV1b and Tg; TPO; TSHr.
The AA sequences identities between the Tg, TPO, TSHr and the HCV subtype 2a polyprotein showed 36%, 32% and 55% sequence similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV2b polyprotein showed 41%, 41% and 35% sequence similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV2c polyprotein showed 67%, 32% and 55% sequence similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV2d polyprotein showed 71%, 30% respectively, and no significant sequences similarities with TSHr. The AA sequences similarities between the Tg, TPO, TSHr and the HCV2e polyprotein showed 57%, 56% and 43% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV2f polyprotein showed 67%, 36% and 26% sequences similarities respectively. The percentage AA sequences identities detected between the Tg, TPO, TSHr and the HCV subtype 2 polyprotein are shown in Table 1.

The AA sequences similarities between the Tg, TPO, TSHr and the HCV subtype 3a polyprotein showed 34%, 39% respectively, and no significant sequences similarities with TSHr. The AA sequences similarities between the Tg, TPO, TSHr and the HCV3b polyprotein showed 25%, 47% and 31% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV3g polyprotein showed 40%, 40% respectively, and no significant sequences similarities with TSHr. The percentage AA sequences similarities detected between the Tg, TPO, TSHr and the HCV subtype 3 polyprotein are shown in Table 2.

The AA sequences similarities between the Tg, TPO, TSHr and the HCV subtype 4a polyprotein showed 28%, 26% and 63% sequences similarities respectively. The AA sequences similarities between the Tg and HCV4c polyprotein showed no significant, whereas the AA sequences similarities between the TPO, TSHr and HCV4c polyprotein showed 21% and 50% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV4d polyprotein showed 45%, 27% and 27% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV subtype 4 polyprotein are shown in Table 3.

The AA sequences similarities between the Tg, TPO, TSHr and the HCV subtype 4a polyprotein showed 28%, 26% and 63% sequences similarities respectively. The AA sequences similarities between the Tg and HCV4c polyprotein showed no significant, whereas the AA sequences similarities between the TPO, TSHr and HCV4c polyprotein showed 21% and 50% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV4d polyprotein showed 45%, 27% and 27% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV subtype 4 polyprotein are shown in Table 3.

<table>
<thead>
<tr>
<th>TABLE 1. Percentage AA sequences structural similarity between HCV subtype 2 and thyroid-specific antigens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thyroid-specific antigens</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tg</td>
</tr>
<tr>
<td>TPO</td>
</tr>
<tr>
<td>TSHr</td>
</tr>
</tbody>
</table>

NS: no significant.

<table>
<thead>
<tr>
<th>TABLE 2. Percentage AA sequences structural similarity between HCV subtype 3 and thyroid-specific antigens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thyroid-specific antigens</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tg</td>
</tr>
<tr>
<td>TPO</td>
</tr>
<tr>
<td>TSHr</td>
</tr>
</tbody>
</table>

The AA sequences identities between the Tg, TPO, TSHr and the HCV5a polyprotein showed 44% and 24% respectively, whereas the AA sequences similarities between the TSHr and HCV5a polyprotein showed no significant similarities (Figure 4). AA appears in standard single letter code. + indicates conserved or semi-conserved substitutions.

<table>
<thead>
<tr>
<th>TABLE 3. Percentage AA sequences structural similarity between HCV subtype 4 and thyroid-specific antigens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thyroid-specific antigens</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tg</td>
</tr>
<tr>
<td>TPO</td>
</tr>
<tr>
<td>TSHr</td>
</tr>
</tbody>
</table>

NS: no significant.

| FIGURE 4. The AA sequences identities between HCV5a and Tg; TPO; TSHr. |
The AA sequences identities between the Tg, TPO, TSHr and the HCV subtype 6a polyprotein showed 53%, 50% and 35% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, and the HCV6h polyprotein showed 53%, 28% respectively, whereas the AA sequences similarities between TSHr and HCV subtype 6h polyprotein showed no significant sequences similarities. The AA sequences similarities between the Tg, TPO, TSHr and the HCV6k polyprotein showed 27%, 25% and 47% sequences similarities respectively. The AA sequences similarities between the Tg, TPO, TSHr and the HCV6n polyprotein showed 26%, 31% and 63% sequences similarities respectively (Table 4).

TABLE 4. Percentage AA sequences structural similarity between HCV subtype 6 and thyroid-specific antigens

<table>
<thead>
<tr>
<th>Thyroid-specific antigens</th>
<th>Hepatitis C virus subtype 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HCV6a</td>
</tr>
<tr>
<td>Tg</td>
<td>53%</td>
</tr>
<tr>
<td>TPO</td>
<td>50%</td>
</tr>
<tr>
<td>TSHr</td>
<td>35%</td>
</tr>
</tbody>
</table>

NS: no significant.

DISCUSSION

Several mechanisms have been proposed to explain induction of autoimmunity by infection and three mechanisms appear important in individuals susceptible to developing autoimmune disease: molecular mimicry, polyclonal T cell activation, and MHC class II antigen induction(24).

This study suggests the possible role of molecular mimicry between the AA sequences of thyroid self-protein and proteins of HCV in development of ATD. The hypothesis that foreign antigens of proteins HCV sharing sequence similarity with the thyroid self-protein might drive autoimmunity, via a molecular mimicry mechanism, has been already suggested(23).

Thyroid disorders are common in patients with chronic HCV infection. Anti-thyroid antibodies are present in 5% to 17%, and thyroid disease occurs in 2% to 13% of patients with HCV infection(21). Thyroid dysfunction and autoimmune thyroiditis were observed during antiviral therapy with interferon-alpha (IFN-α) and ribavirin therapy especially in patients with previous positivity for antithyroid antibodies. IFN-α therapy for HCV infection may induce thyroid changes or dysfunction in 2.5% to 20% of treated patients. Up to 40% become thyroid antibody positive and these side-effects can interfere with effective management of HCV(22).

The fact of a microorganisms develop antigen determinants that resemble antigens of the host as a mean of avoiding recognition and elimination by the host was first proposed in 1964 by Damian(23). The molecular mimicry could explain the development of autoimmune diseases, through of cross reactions between epitopes of invading microorganisms and antigens present in the body that promote an adverse autoimmune response(30).

In this work, we analyzed the sequence similarity between the AA sequences of the several subtypes HCV polyprotein and three tissue-specific antigens of human thyroid. We found that HCV and thyroid auto-antigens share AA sequence similarity, wherein some similar regions contain epitopes of both HCV and thyroid auto-antigens very high. A similar study showed that the similarities between the thyroid and the viral peptides were similar to ours, and the frequency of the sequence similarity increased up to 100%, when conservative substitutions were included in the analysis(23).

In our study high similarities were found between Tg and HCV polypeptide subtype 2d, while the lowest percentage of sequence similarity between TPO and HCV polyprotein occurred with subtype 4c.

Given the sequence similarity between HCV sequence and thyroglobulin sequence and the HCV is classified in six major genotypes and over 70 subtypes, has been observed at possible association between genotype viral and susceptibility of patients in developing of ATD(17). The presence of anti-Tg antibodies in the clinical course of the autoimmune thyroiditis in HCV infection treated with IFN-α may be related to the high number of similarities between the Tg antigen and the HCV polyprotein(6). In our study significant similarities were found between Tg and HCV polypeptide subtype 2d, while the lowest percentage of sequence similarity between Tg and HCV polyprotein occurred with subtype 3b.

The TPO, that plays a key role in the biosynthesis of thyroid hormones, contains 933 AA protein sequences, and most of the extracellular region of TPO shows a high degree of sequence similarity with HCV polyprotein(28). AA sequences similarities between HCV polypeptide and TPO have been reported(25), however the similar viral peptides have not been examined for immune-pathogenicity. A study examining 20 amino acid-length peptides showed 41.7 to 58.3% sequence similarities between TPO and HCV5a, increased to 75.0% when including the conservative residues and identical residues(17). Our results showed larger sequence similarity between TPO and HCV polypeptide subtype 2c, while the lowest percentage of sequence similarity between TPO and HCV polypeptide occurred with subtype 4c, and HCV polypeptide subtype 5a.

The TSHr regulates thyroid growth and differentiated functions, is also a member of G protein-coupled receptors and well known to undergo similar desensitization(29). Prior study already demonstrated molecular mimicry between the TSHr and HCV1a(31). Comparing the protein sequences of TSHr and HCV polyprotein, our study found similarities equality between TSHr and HCV polyprotein subtype 4a and 6n, while the lowest percentage of similarities between TSHr and HCV polyprotein occurred with subtype 5a. In our study there were no significant similarities between TSHr and HCV polyprotein subtype 2d, 3a, 3g and 6h.
Despite the limitations posed by the unavailability of complete proteome data for HCV proteins, because the BLAST is a method limited to searching for linear epitope similarities, losing three-dimensional conformational similarities and possible cross-reactivity between proteic and non-proteic epitopes, the similarities between the AA sequences of thyroid self-protein, which are potential B- and T-cell epitopes of these antigens and proteins of HCV were successfully identified. Thus, these observed similarities could be functionally important in molecular mimicry, receptor binding and cell signaling events involved in autoimmunity, and has important implications for the understanding of the occurrence of autoimmune diseases.

In conclusion, it is important to draw attention that bioinformatics data suggest a possible pathogenic link between HCV and ATD. Therefore, through of molecular mimicry is observed that sequences similarities between viral polyproteins and self-proteins thyroid could be a mechanism of induction a cross over immune response to self-antigens resulting in autoimmune disease thyroid. Finally, our result adds not only new data, but also incorporates data to few already published studies.

Authors’ contributions

Conception and design, writing the article, critical revision of the article, final approval of the article: Andrade LJO, Sousa MAC, Paraná R. Analysis and interpretation, data collection, statistical analysis: Andrade LJO, Sousa MAC. Overall responsibility: Andrade LJO.
REFERENCES


