BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **20**, No 6
Special Issue on New Developments in Scalable Computing
Sofia • 2020 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2020-0073

Immunoinformatic Analysis of Human Thyroglobulin

Irini Doytchinova[1], *Andrey Tchorbanov*[2], *Stefan Ivanov*[3]

[1]*Faculty of Pharmacy, Medical University of Sofia, 1000 Sofia, Bulgaria*
[2]*Institute of Microbiology, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria*
[3]*Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD 20850, USA*
E-mails: idoytchinova@pharmfac.mu-sofia.bg tchorban@microbio.bas.bg sivanov@ddg-pharmfac.net

Abstract: *The AutoImmune ThyroiDitis* (*AITD*)*, known as Hashimoto's disease, is a chronic autoimmune thyroid disease progressively developed to hypothyroidism. The AITD is characterized by the formation of autoantibodies targeting two specific thyroid antigens, Thyroglobulin* (*Tg*) *and Thyroid PerOxidase* (*TPO*)*. Tg is a precursor of the thyroid hormones while TPO catalyses their synthesis. The AITD has a strong genetic predisposition. During the last years, it was found that the susceptibility to AITD is associated with certain Human Leukocyte Antigens* (*HLA*) *class II genes of loci DR and DQ. In the present study, we applied in-house immunoinformatic tools to identify peptides originating from Tg and binding to AITD susceptible alleles: HLA-DR3, HLA-DR4, HLA-DR5, HLA-DQ2 and HLA-DQ8. Five peptide fragments containing promiscuous overlapping binders were selected. These were p470, p949, p1948, p2348 and p2583. Only one of them contains a known epitope* (*p1948*)*. The rest have not been reported yet. The selected peptide fragments will be coupled to monoclonal antibodies specific to inhibitory B cell receptors designed to suppress the production of Tg autoantibodies.*

1. Introduction

The protein Thyroglobulin (Tg) is a precursor of the thyroid hormones Triiodothyronine (T3) and Tetra-iodothyronine (T4). The synthesis of T3 and T4 from Tg proceeds in the thyroid gland by iodination and coupling of pairs of Tyr residues positioned in close proximity on Tg. The process of iodination is catalysed by the enzyme Thyroid PerOxidase (TPO).

Tg is a homodimer of glycoproteins (Fig. 1). Its structure was recently resolved by cryo-electron microscopy (pdb code: 6SCJ) [1]. Each subunit has a size of 330 kDa and contains 2768 amino acids. Among them, 66 are Tyr and half of them are iodinated but only small number are hormonogenic, i.e., are able to convert in thyroid hormones [1].

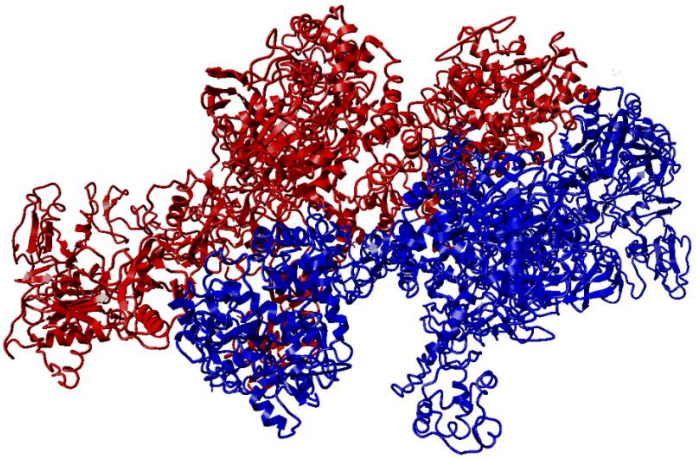

Fig. 1. Homodimer of human Thyroglobulin (Tg) (pdb code: 6SCJ). The monomers are given in red and blue

In human, Tg and TPO are potential autoantigens. In some patients, they are recognized by the T lymphocytes, followed by the production of CD4+ T cells, CD8+ T cells and Immunoglobulin G (IgG) autoantibodies which attack and destroy progressively the thyrocytes and lead to hypothyroidism. These processes underlie the AutoImmune ThyroiDitis (AITD) known as well as Hashimoto's disease [2]. AITD is one of the most common human autoimmune disease and affects around 2% of the general population [3]. The onset of AITD depends on a combination of genetic and environmental factors like iodine intake [4], bacterial and viral infections [5], and pregnancy [6]. Among the genetic factors, the role of Human Leukocyte Antigens (HLA) is the most prominent [2]. The HLA genes encode intracellular proteins, which are involved in the antigen processing [7]. The role of HLA proteins in the cell is to bind peptides originating from foreign or self-proteins and to present them to T cells. The peptides recognized by the T cells are assigned as epitopes. The activation of T cells initiates an immune response leading to production of specific antibodies targeting and destroying the Antigen-Presenting Cells (APC). If the peptides fail to bind to HLA, no immune reaction is initiated. If peptides originating from self-proteins bind to HLA, the activation of T cells triggers an autoimmune disease like AITD.

HLA are extremely polymorphic. By September 2020, in the IMGT/HLA database are described 20192 HLA class I and 7407 HLA class II genes. The role of selected HLA class II genes susceptible to AITD has been clarified during the last decades. HLA-DR3, -DR4 and -DR5 have been reported to be associated to AITD in Caucasians [2, 8]. Additionally, HLA-DQ2 and -DQ8 are present in half of the patients with AITD [9].

The asymptomatic and subclinical AITD has no specific therapy. When the disease progresses to hypothyroidism, a thyroid hormone replacement with synthetic T4 is administered. During the last years, many attempts have been made to control and modulate the autoimmune diseases by supressing and elimination of

autoantibodies targeting autoantigens [10]. Recently, we constructed a chimeric protein molecule containing a specific monoclonal antibody (mAb) coupled with peptide epitopes originating from the type 1 diabetes mellitus autoantigen GAD65 [11, 12]. The chimeric molecule was able to modulate the activity of GAD65-specific B-lymphocytes and to supress the production of anti-GAD65 antibodies.

In the present study, we analyse the protein sequence of human Tg by in-house immunoinformatic tools and identify peptides binding selectively to HLA proteins susceptible to AITD. Next, these peptides will be loaded to a chimeric molecule designed to suppress the production of Tg autoantibodies.

2. Protein sequences and servers for HLA class II binding prediction

2.1. Tg proteins

The protein sequence of the human Tg (UniProtKB: P01266) was retrieved from UniProt knowledgebase [13]. Tg consists of 2768 residues. It starts with a signal peptide of 19 amino acids, which is removed in the mature protein. There are two isoforms of Tg – isoform 1 (major) and isoform 2 (minor). The isoform 1 is defined as a canonical sequence because is the most prevalent and the most similar to orthologous sequences found in other species [13]. In the present study, we used the isoform 1 with identifier P01266-1. As the preclinical studies of the constructed chimeric antibodies will be tested on humanized NOD-SCID mice, similarity search was performed with the mouse Tg (UniProtKB: O08710).

2.2. Peptide binding prediction

The peptides binding to HLA-DR3 (DRB1*03:01), HLA-DR4 (DRB1*04:01, DRB1*04:04 and DRB1*04:05), HLA-DR5 (DRB1*11:01 and DRB1*12:01), HLA-DQ2 (DQA1*05:01/ DQB1*02:01) and HLA-DQ8 (DQA1*03:01/DQB1*03:02) were predicted by the in-house servers EpiTOP and EpiDOCK.

EpiTOP [14] is a server for peptide binding prediction based on proteochemometrics [15]. It predicts binding to 24 most frequent HLA-DR, -DQ and -DP alleles in human population [16]. The predicted affinity is given as pIC_{50} ($-\log IC_{50}$), where IC_{50} is the concentration inhibiting 50% of the binding of a radiolabeled standard peptide to detergent-solubilized MHC molecules. EpiTOP is freely accessible at **http://www.ddg-pharmfac.net/** EpiTOP3.

EpiDOCK [17] is a server for peptide binding prediction based on quantitative matrices derived by molecular docking of peptide combinatorial libraries on HLA class II proteins and predicts binding to 23 most frequent HLA-DR, -DQ and -DP alleles (same alleles as EpiTOP with DQA1*05:01/DQB1*02:01 but without DPA1*01:03/DPB1*03:01 and DPA1*03:01/DPB1*04:02). The predicted affinity is given as a score assessing the free energy of binding. EpiDOCK is accessible at **http://www.ddg-pharmfac.net/epidock**.

Both servers predict the peptide binding core consisted of nine successive amino acid residues.

3. Results and discussion

The signal peptide (residues 1-19) was removed and the mature protein sequence was presented as a set of overlapping nonamers (peptides consisting of 9 amino acids). The nonamers correspond to the binding cores of peptides binding to HLA class II proteins (Fig. 2). Thus, 2741 nonamer peptides were generated.

EpiTOP and EpiDOCK were applied to predict the binding affinities of peptides to AITD-susceptible HLA alleles: HLA-DR3 (DRB1*03:01), HLA-DR4 (DRB1*04:01, DRB1*04:04 and DRB1*04:05), HLA-DR5 (DRB1*11:01 and DRB1*12:01), HLA-DQ2 (DQA1*05:01/DQB1*02:01) and HLA-DQ8 (DQA1*03:01/DQB1*03:02). The top 10 best predicted binders for each allele were mapped on the Tg sequence. Initially, 54 peptide fragments containing between 1 and 15 binders were identified. The fragments include overlapping or non-overlapping binding cores and 4 flanking residues – two of each end (Fig. 2).

Fig. 2. Peptide bound into the binding site of HLA-DQ8 (pdb code: 1JK8). The HLA protein is given in green. The binding core consists of 9 amino acids (given in blue). The flanking residues are given in cyan

Next, only fragments containing peptides binding to at least 5 of 8 alleles were considered. Thus, 7 promiscuous fragments were selected. The binder identities in human and mouse Tgs were compared and only 5 peptide fragments showed 75% or higher identity. These were p470, p949, p1948, p2348, and p2583. They were selected for analysis in the present study (Table 1). Only one of them contains a known epitope (p1948) [18]. The rest have not been reported yet.

The peptide p470 consists of 41 residues and contain 15 predicted binders to 5 DR and 1 DQ alleles with 78% identity between human and mouse Tgs. The peptide p949 has 20 residues and 8 predicted binders to 5 DR and 1 DQ alleles with 88% identity. The peptide p1948 covers 6 alleles with 6 predicted binders and 75% identity. It contains the known T-cell epitope ^{1956}ILEDKVKNF^{1964} associated with

AITD [18]. The peptide p2348 covers 5 DR alleles with 82% binder identity with mouse Tg. The p2583 contains 6 binders to 6 alleles with 75% identity.

Table 1. Human Tg peptide fragments (epitopes) selected in the present study. The predicted binding cores are given in bold. Peptides are encoded by single amino acid letters

Position	Peptide	Binders	Covered HLA	Identity to mouse Tg
470	NL**FGGKFLVNVGQFNLSGAL GTRGTFNFSQFFQQLGLA**SFL	15	DR3, DR4, DR5, DQ8	78%
949	SR**FPLGESFLVAKGIRLR**NE	8	DR3, DR4, DR5, DQ8	88%
1948	KA**LFRKKVILEDKVKNFY**TR	6	DR4, DR5, DQ2	75%
2348	GE**VSGNWGLLDQVAA**LT**WV QTHIRGF**GG	5	DR4, DR5	82%
2583	AT**RDYFIICPIIDMASAWAKR ARGNV**FM	6	DR4, DR5, DQ8	75%

The selected Tg epitopes will be coupled to a specific mAb (CD35) suppressing the production of anti-Tg antibodies by B cells (Fig. 3). By the Tg epitopes, the constructed chimera cross-links the B-Cell Receptor (BCR) on the B cell secreting anti-Tg antibodies and the suppressive CD35 molecule. This chimeric molecule is expected to suppress the B cell differentiation to anti-Tg antibody-secreting plasma cells as well as to increase the apoptosis of already differentiated B cells.

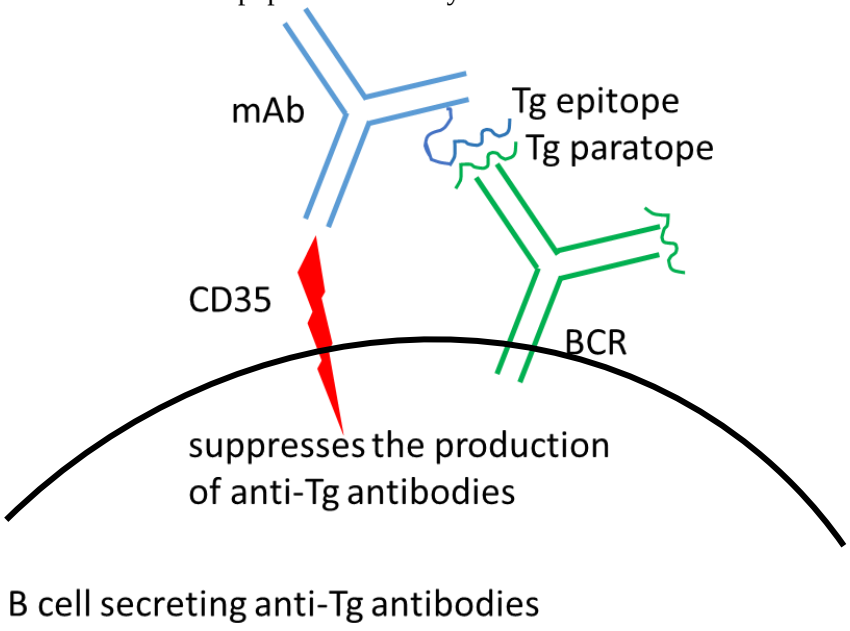

Fig. 3. A putative mechanism of suppressing the production of anti-Tg antibodies by a chimeric protein molecule (mAb) consisting of CD35 mAb and Tg epitope (in blue). The mAb cross-links the B-Cell Receptor (BCR) with Tg paratope (in green) and the suppressive CD35 molecule (in red) on the surface of a specific B cell

In conclusion, in the present study we applied immunoinformatic tools to identify T-cell epitopes originating from the human Tg and binding to HLA class II

proteins associated with susceptibility to AITD. The epitopes will be utilized in the development of a specific immunotherapeutic drug against AITD.

***Acknowledgments*:** This work was accomplished by the financial support of Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014-2020) and co-financed by the European Union through the European Structural and Investment funds.

References

1. Coscia, F., A. Taler-Verčič, V. T. Chang, L. Sinn, F. J. O'Reilly, T. Izoré, M. Renko, I. Berger, J. Rappsilber, D. Turk, J. Löwe. The Structure of Human Thyroglobulin. – Nature, Vol. **578**, 2020, Issue 7796, pp. 627-630.
2. Chistiakov, D. A. Immunogenetics of Hashimoto's Thyroiditis. – Journal of Autoimmune Diseases, Vol. **2**, 2005, Issue 1, pp. 1-21.
3. Wang, C., L. M. Crapo. The Epidemiology of Thyroid Disease and Implications for Screening. – Endocrinology and Metabolism Clinics of North America, Vol. **26**, 1997, Issue 1, pp. 189-218.
4. Rose, N. R., R. Bonita, C. L. Burek. Iodine: An Environmental Trigger of Thyroiditis. – Autoimmunity Reviews, Vol. **1**, 2002, Issue 1-2, pp. 97-103.
5. Köhling, H. L., S. F. Plummer, J. R. Marchesi, K. S. DavIdge, M. Ludgate. The Microbiota and Autoimmunity: Their Role in Thyroid Autoimmune Diseases. – Clinical Immunology, Vol. **183**, 2017, Issue 10, pp. 63-74.
6. Imaizumi, M., A. Pritsker, M. Kita, L. Ahmad, P. Unger, T. Davies. Pregnancy and Murine Thyroiditis: Thyroglobulin Immunization Leads to Fetal Loss in Specific Allogeneic Pregnancies. – Endocrinology, Vol. **142**, 2001, Issue 2, pp. 823-829.
7. Patronov, A., I. Doytchinova. T-Cell Epitope Vaccine Design by Immunoinformatics. – Open Biology, Vol. **3**, 2013, Article 120139.
8. Sasazuki, T., H. Inoko, S. Morishima, Y. Morishima. Gene Map of the HLA Region, Graves' Disease and Hashimoto Thyroiditis, and Hematopoietic Stem Cell Transplantation. – Advances in Immunology, Vol. **129**, 2016, pp. 175-249.
9. Hadithi, M., H. de Boer, J. W. R. Meijer, F. Willekens, J. A. Kerckhaert, R. Heijmans, A. S. Peña, C. D. A. Stehouwer, C. J. J. Mulder. Coeliac Disease in Dutch Patients with Hashimoto's Thyroiditis and Vice Versa. – World Journal of Gastroenterology, Vol. **13**, 2007, Issue 11, pp. 1715-1722.
10. Yasunaga, M. Antibody Therapeutics and Immunoregulation in Cancer and Autoimmune Disease. – Seminars in Cancer Biology, Vol. **64**, 2020, Issue 1, pp. 1-12.
11. Manoylov, I. K., G. V. Boneva, I. A. Doytchinova, N. M. Mihaylova, A. I. Tchorbanov. Protein-Engineered Molecules Carrying GAD65 Epitopes and Targeting CD35 Selectively Down-Modulate Disease-Associated Human B Lymphocytes. – Clinical & Experimental Immunology, Vol. **197**, 2019, Issue 3, pp. 329-340.
12. Manoylov, I. K., G. V. Boneva, I. A. Doytchinova, N. M. Mihaylova, A. I. Tchorbanov. Suppression of Disease-Associated B Lymphocytes by GAD65 Epitope-Carrying Protein-Engineered Molecules in a Streptozotocin-Induced Mouse Model of Diabetes. – Monoclonal Antibodies in Immunodiagnosis and Immunotherapy, Vol. **38**, 2019, Issue 5, pp. 201-208.
13. The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. – Nucleic Acids Research, Vol. **47**, 2019, Issue D1, pp. D506-515.
14. Dimitrov, I., P. Garnev, D. R. Flower, I. Doytchinova. EpiTOP – A Proteochemometric Tool for MHC Class II Binding Prediction. – Bioinformatics, Vol. **26**, 2010, Issue 16, pp. 2066-2068.
15. Lapinsh, M., P. Prusis, A. Gutcaits, T. Lundstedt, J. E. S. Wikberg. Development of Proteo-Chemometrics: A Novel Technology for the Analysis of Drug-Receptor Interactions. – Biochimica and Biophysica Acta, Vol. **1525**, 2001, Issue 1-2, pp. 180-190.

16. G o n z a l e z-G a l a r z a, F. F., L. Y. T a k e s h i t a, E. J. S a n t o s, F. K e m p s o n, M. H. M a i a, A. L. S i l v a, A. L. S i l v a, G. S. G h a t t a o r a y a, A. A l f i r e v i c, A. R. J o n e s, D. M i d d l e t o n. Allele Frequency Net 2015 Update: New Features for HLA Epitopes, KIR and Disease and HLA Adverse Drug Reaction Associations. – Nucleic Acid Research, Vol. **28**, 2015, Database Issue, pp. D784-788.
17. A t a n a s o v a, M., I. D i m i t r o v, D. R. F l o w e r, I. D o y t c h i n o v a. EpiDOCK: A Molecular Docking-Based Tool for MHC Class II Binding Prediction. – Protein Engineering Design and Selection, Vol. **26**, 2013, Issue 10, pp. 631-634.
18. L i, C. W., R. O s m a n, F. M e n c o n i, E. S. C o n c e p c i o n, Y. T o m e r. Flexible Peptide Recognition by HLA-DR Triggers Specific Autoimmune T-Cell Responses in Autoimmune Thyroiditis and Diabetes. – Journal of Autoimmunity, Vol. **76**, 2017, Issue 1, pp. 1-9.

Received: 15.09.2020; Second Version: 30.10.2020; Accepted: 4.11.2020