Endocrinol. Diabetes Clín. Exp.
VOL 22 - number 4 Oct/Nov/Dec 2025

DOI: 10.29327/2413063.22.4-4

Original Article

AI-Driven Structural Mapping of Thyroid Autoantigens Reveals High-Confidence Conformational Epitopes for Precision Immunotherapy

Mapeamento estrutural orientado por IA de autoantígenos tireoidianos revela epítopos conformacionais de alta confiança para imunoterapia de resultados.

[1] Luís Jesuíno de Oliveira Andrade - https://orcid.org/0000-0002-7714-0330
[2] Gabriela Correia Matos de Oliveira - https://orcid.org/0000-0002-3447-3143
[3] Alcina Maria Vinhaes Bittencourt - https://orcid.org/0000-0003-0506-9210
[4] Osmario Jorge de Mattos Salles - https://orcid.org/0009-0002-1859-0478
[5] Jonh Menezes Leahy Neto - https://orcid.org/0009-0009-5424-0994
[1] Luís Matos de Oliveira - https://orcid.org/0000-0003-4854-6910

[1] Department of Health, Santa Cruz State University, Ilhéus, Bahia, Brazil.
[2] José Silveira Foundation, Salvador, Bahia, Brazil.
[3] School of Medicine, Federal University of Bahia, Salvador, Bahia, Brazil.
[4] Bahiana School of Medicine and Public Health, Salvador, Bahia, Brazil.
[5] School of Medicine, Unifacs, Salvador, Bahia, Brazil.

Corresponding Author:
Luís Jesuino de Oliveira Andrade
Universidade Estadual de Santa Cruz - Campus Soane Nazaré de Andrade, Rod. Jorge Amado, Km 16 - Salobrinho, Ilhéus - BA, 45662-900.
E-mail: luis_jesuino@yahoo.com.br
Received in: 01-10-2025
Accepted in: 15-10-2025
Conflicts of interest: None declared.

ABSTRACT

Introduction: Autoimmune thyroid diseases (AITD) are the most common organ-specific autoimmune disorders, driven by loss of immune tolerance to thyroid peroxidase (TPO), thyroglobulin (Tg), and thyrotropin receptor (TSHR). Understanding the three-dimensional (3D) architecture of these autoantigens is essential for the rational design of antigen-specific immunotherapies that restore tolerance without systemic immunosuppression. **Objective:** To predict the 3D structures of thyroid autoantigens using AI-Driven and to identify structurally and immunologically validated epitopes suitable for therapeutic peptide development. **Methods:** ESMFold was employed to model full-length structures of TPO, Tg, and TSHR. Predicted epitopes were characterized using integrated bioinformatic tools for major histocompatibility complex (MHC) class II binding, surface accessibility, and glycosylation proximity. Prioritized epitopes underwent molecular dynamics simulations to assess conformational stability and conservation analyses across mammalian orthologs. **Results:** Seventy-two epitopes were identified and clustered into five groups based on twelve physicochemical features. Fifteen high-priority epitopes demonstrated high structural confidence (mean pLDDT > 90), broad MHC affinity (IC_{50} < 1000 nM), and strong concordance with experimentally validated domains. Molecular dynamics confirmed their conformational robustness (mean RMSF 1.4 Å), and cross-species analysis revealed >80% sequence conservation, supporting translational applicability. **Conclusion:** AI-Driven modeling provides a precise structural framework for identifying therapeutic epitopes in AITD. The prioritized epitopes represent promising candidates for tolerance-inducing peptide immunotherapies targeting thyroid autoimmunity.

Keywords: Autoimmune thyroid disease; Epitope prediction; Molecular dynamics.

RESUMO

Introdução: As doenças autoimunes tireoidianas (DAIT) são os distúrbios autoimunes organoespecíficos mais comuns, motivados pela perda da tolerância imune à peroxidase tireoidiana (TPO), tireoglobulina (Tg) e receptor de tireotrofina (TSHR). A compreensão da arquitetura tridimensional (3D) desses autoantígenos é essencial para o design racional de imunoterapias específicas que restauram a tolerância sem imunossupressão sistêmica. **Objetivo**: Prever as estruturas 3D dos autoantígenos tireoidianos através de abordagem orientada por inteligência artificial (IA) e identificar epítopos estruturais e imunologicamente validados, aptos para desenvolvimento de

peptídeos terapêuticos. **Métodos:** Utilizou-se ESMFold para modelar as estruturas completas de TPO, Tg e TSHR. Os epítopos preditos foram caracterizados por ferramentas de bioinformática integradas quanto à camada superficial para o complexo principal de histocompatibilidade (MHC) classe II, acessibilidade superficial e proximidade de glicosilações. Os epítopos prioritários foram submetidos a simulações de dinâmica molecular para avaliação da estabilidade conformacional e análises de conservação entre ortólogos de mamíferos. **Resultados:** Setenta e dois epítopos foram identificados e agrupados em cinco clusters baseados em doze características físico-químicas. Quinze epítopos de alta prioridade apresentaram alta confiança estrutural (pLDDT médio > 90), ampla camada de MHC (IC_{50} < 1000 nM) e forte concordância com domínios experimentalmente validados. A dinâmica molecular confirmou robustez conformacional (RMSF médio 1,4 Å), e a análise de interesses evidenciou conservação de sequência >80%, suportando aplicabilidade translacional. **Conclusão**: A modelagem orientada por IA oferece uma estrutura estrutural precisa para identificação de epítopos terapêuticos em DAIT. Os epítopos priorizados são candidatos promissores para imunoterapias peptídicas indutoras de tolerância direcionadas à autoimunidade tireoidiana.

Descritores: Doença autoimune da tireoide; Predição de epítopos; Dinâmica molecular.

INTRODUCTION

Autoimmune thyroid diseases (AITD) represent the most prevalent organ-specific autoimmune conditions, affecting approximately 5% of the population globally.[1] The pathogenesis centers on loss of immune tolerance to three principal thyroid autoantigens: thyroid peroxidase (TPO), thyroglobulin (Tg), and the thyrotropin receptor (TSHR). These complex glycosylated proteins, particularly the membrane-bound TPO and the massive 660-kDa Tg dimer, serve as targets for both humoral and cellular immune responses in Hashimoto's thyroiditis and Graves' disease.[2] Autoantibodies predominantly recognize conformational epitopes within immunodominant regions designated as domains A and B on TPO, while Tg autoantibodies display restricted recognition patterns that persist despite treatment fluctuations.[3] Understanding the structural basis of these epitopes has become essential for developing therapeutic interventions targeting the underlying immune dysregulation.

Epitope-specific immunotherapy represents an emerging paradigm for restoring antigen-specific tolerance without systemic immunosuppression. This approach delivers

synthetic peptides corresponding to T cell epitopes in a manner that induces regulatory T cells and establishes linked suppression.[4,5] Studies demonstrate that tolerogenic peptides, designed as antigen-processing independent epitopes (apitopes), can modulate autoimmune progression by engaging autoreactive $CD4^+$ T cells and promoting IL-10-secreting regulatory populations.[6] Clinical trials evaluating peptide-based vaccines in multiple sclerosis, type 1 diabetes, and rheumatoid arthritis have shown encouraging safety profiles and preliminary efficacy.[7] However, successful immunotherapy requires precise identification of disease-relevant epitopes that can induce tolerance rather than exacerbate autoimmunity.

Despite extensive serological characterization, a fundamental gap persists regarding the three-dimensional (3D) architecture of thyroid autoantigen epitopes. Traditional mapping using linear peptides provides incomplete information, as thyroid autoantibodies predominantly recognize conformational determinants dependent on proper protein folding.[2,3] Experimental structure determination of these heavily glycosylated proteins remains technically challenging, hindering rational design of tolerogenic peptides.

ESMFold has revolutionized structural biology by achieving near-experimental accuracy in protein structure prediction through deep learning.[8] The system demonstrated unprecedented performance in CASP14, with predictions often rivaling crystallographic accuracy.[9] Predictions with per-residue Local Distance Difference Test (pLDDT) exceeding 90 are suitable for epitope characterization and binding site analysis.[10] These advances enable generation of high-confidence structural models revealing surface-exposed regions and conformational epitopes essential for immunotherapy design.

This study employs ESMFold to generate 3D models of thyroid autoantigens and systematically identify immunodominant epitopes suitable for antigen-specific immunotherapy development, thereby providing structural foundations for precision treatment of AITD.

METHODS

Study Design and Ethical Considerations

This computational study employed artificial intelligence-driven protein structure prediction to model thyroid autoantigens and identify potential immunotherapeutic epitopes. As the investigation utilized publicly available protein

sequences and computational algorithms without involving human subjects or biological specimens, institutional review board approval was not required. All analyses were conducted in accordance with FAIR principles for computational research reproducibility. All evaluations utilized publicly accessible, free online software tools.

Protein Sequence Retrieval and Preparation

Amino acid sequences for human thyroid autoantigens were obtained from the UniProt database (https://www.uniprot.org/). The following primary targets were selected: TPO (TPO, UniProt ID: P07202), thyroglobulin (Tg, UniProt ID: P01266), and thyrotropin receptor (TSHR, UniProt ID: P16473). Complete canonical sequences including signal peptides and known splice variants were retrieved in FASTA format. Sequence validation was performed by cross-referencing with NCBI RefSeq entries to ensure accuracy. Post-translational modification sites, particularly N-glycosylation consensus sequences (Asn-X-Ser/Thr), were annotated using NetNGlyc 1.0 server prior to structural modeling, as glycosylation significantly influences protein folding and epitope accessibility in thyroid autoantigens.

Three-Dimensional Structure Prediction Using ESMFold

Structural predictions were generated using ESMFold freely available for research and commercial applications (https://esmatlas.com/), which provides an optimized pipeline with MMseqs2 for rapid multiple sequence alignment generation. For each autoantigen, we employed the following parameters: three recycles for iterative refinement, template mode enabled to leverage existing structural information from the Protein Data Bank, and amber force field relaxation for energy minimization of the final models. Given the substantial size of thyroglobulin (2768 residues), domain-based predictions were performed by dividing the sequence into overlapping fragments of approximately 400 residues with 50-residue overlaps to maintain continuity, followed by structural alignment and assembly using Mol (Molstar) (https://molstar.org/viewer/).

Model confidence was assessed using pLDDT scores, where values above 90 indicate high confidence suitable for detailed analysis, scores between 70-90 suggest generally reliable backbone positioning, and regions below 70 were flagged as potentially unreliable. Additionally, predicted aligned error (PAE) matrices were examined to evaluate inter-domain confidence and identify potential artifacts in multi-domain proteins. For membrane-spanning regions in TPO and TSHR, we validated

predicted topology against experimental data from hydropathy analyses and known transmembrane domains documented in the literature.

Epitope Identification and Mapping

B-cell epitope prediction was conducted using a consensus approach integrating multiple algorithms. The Immune Epitope Database (IEDB) tools (http://tools.iedb.org/main/) were employed, specifically the BepiPred 2.0 algorithm for linear epitope prediction with a threshold of 0.5, and the DiscoTope 2.0 server for conformational epitope identification based on surface accessibility, spatial arrangement, and amino acid propensity scores. Surface accessibility analysis was performed using DSSP algorithm implemented in Biopython, with residues exhibiting relative solvent accessibility exceeding 20% considered surface-exposed. Conformational B-cell epitopes were prioritized by overlaying predicted regions with experimentally validated immunodominant domains reported in previous mutagenesis and antibody binding studies.

T-cell epitope prediction focused on major histocompatibility complex (MHC) class II binding peptides, as CD4+ T cells play central roles in orchestrating AITD. We utilized NetMHCIIpan 4.0, which was trained on a comprehensive dataset of peptide-MHC binding affinities, to predict 15-mer peptides binding to prevalent HLA-DR alleles associated with AITD susceptibility (DRB103:01, DRB104:01, DRB108:03, and DRB116:02). Peptides with predicted binding affinity below 500 nM (strong binders) or between 500-5000 nM (weak binders) were retained for further analysis. Antigen processing was simulated using NetChop 3.1 (https://services.healthtech.dtu.dk/services/NetChop-3.1/) to predict proteasomal cleavage sites, ensuring that identified epitopes would likely be generated during natural antigen processing pathways.

Structural Analysis and Epitope Characterization

3D visualization and structural analysis were performed using Mol (Molstar). Predicted epitopes were mapped onto the ESMFold-generated structures and color-coded according to confidence scores and epitope type. We calculated the following structural parameters for each predicted epitope: secondary structure composition (alpha-helix, beta-sheet, random coil), electrostatic surface potential using the PDB2PQR/APBS Web Server (https://server.poissonboltzmann.org/), and local

geometric features including protrusions and cavity depth using PDBePISA (https://www.ebi.ac.uk/pdbe/pisa/). For conformational epitopes, we measured spatial clustering of discontinuous residues by calculating inter-residue distances and defined an epitope cluster as residues within 8 Å of each other in 3D space despite being distant in primary sequence.

Glycosylation site proximity to predicted epitopes was evaluated systematically, as carbohydrate moieties can either mask or create neo-epitopes. Distances between N-glycosylation consensus sequences and epitope residues were measured, with sites within 15 Å considered potentially influential on antibody recognition. This analysis is particularly relevant for thyroglobulin, which contains approximately 20 N-glycosylation sites that substantially modify its immunological properties.

Cross-Validation with Experimental Data

To validate computational predictions, we performed systematic literature mining to compile experimentally verified epitopes from published studies employing techniques such as peptide scanning, site-directed mutagenesis, monoclonal antibody mapping, and crystallographic analyses of antigen-antibody complexes. Predicted epitopes were compared against this curated dataset using sequence alignment and structural superposition. Agreement between computational predictions and experimental data was quantified by calculating sensitivity, specificity, and Matthews correlation coefficient, treating experimentally validated residues as ground truth. This validation step provides fundamental assessment of prediction reliability and identifies high-confidence epitopes warranting prioritization for experimental immunotherapy development.

Prioritization of Therapeutic Epitope Candidates

Epitopes were ranked for immunotherapeutic potential using a multi-criteria scoring system incorporating: (1) structural confidence (pLDDT > 90), (2) surface accessibility (RSA > 30%), (3) correspondence with experimentally validated immunodominant regions, (4) MHC binding affinity (IC_{50} < 1000 nM for multiple HLA-DR alleles), (5) conservation across species to minimize xenogeneic immune responses, and (6) absence of extensive post-translational modifications that would complicate peptide synthesis. Conservation analysis was performed using multiple sequence alignments of mammalian orthologs retrieved from Ensembl database and visualized using SkyLogo (http://skylign.org/) to identify conserved positions within

epitopes. The top-ranking epitopes based on composite scores were designated as lead candidates for tolerogenic peptide design.

Molecular Dynamics Simulation

To assess structural stability of predicted epitopes, selected regions underwent molecular dynamics simulations using GROMACS-Open (https://www.gromacs.org/) with the AMBER99SB-ILDN force field. Systems were solvated in TIP3P water boxes with 150 mM NaCl and neutralized with counter-ions. Following energy minimization using steepest descent algorithm, systems were equilibrated through 100 ps NVT and 100 ps NPT phases before 100 ns production runs at 310 K (physiological temperature) with 2 fs timesteps. Root mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration were calculated to evaluate epitope flexibility and stability. Regions maintaining stable conformations throughout the simulation (RMSF < 2 Å) were considered structurally robust epitopes suitable for rational peptide design.

Statistical and Bioinformatics Analysis

All bioinformatic analyses were performed using Galaxy Project (https://usegalaxy.org/). Structural alignments and RMSD calculations utilized the Mol (Molstar). Statistical comparisons between epitope properties (accessibility, hydrophobicity, charge) were conducted using Mann-Whitney U tests for non-parametric distributions, with p-values below 0.05 considered statistically significant. Hierarchical clustering of epitopes based on physicochemical properties was performed using ClustVis (https://biit.cs.ut.ee/clustvis/) with Ward linkage and Euclidean distance metrics. All visualizations including heatmaps, surface representations, and phylogenetic trees were generated using RAWGraphs (https://www.rawgraphs.io/), and specialized structural biology visualization tools.

RESULTS

Structural Prediction Quality and Model Confidence

ESMFold generated high-confidence three-dimensional models for all three thyroid autoantigens examined in this study. The overall prediction quality, assessed through pLDDT, demonstrated remarkable accuracy across the majority of protein domains. For TPO, which comprises 933 residues, we observed a mean pLDDT of 87.3, with 76% of residues achieving scores above 90, indicating highly reliable structural

predictions. The catalytic domain and the two immunodominant regions (domain A: residues 353-538; domain B: residues 590-724) exhibited particularly high confidence scores, with average pLDDT values of 92.1 and 89.7, respectively. The transmembrane helix (residues 1-35) showed lower confidence (pLDDT 68.4), consistent with the inherent challenges in modeling membrane-embedded regions without explicit lipid environments (Figure 1).

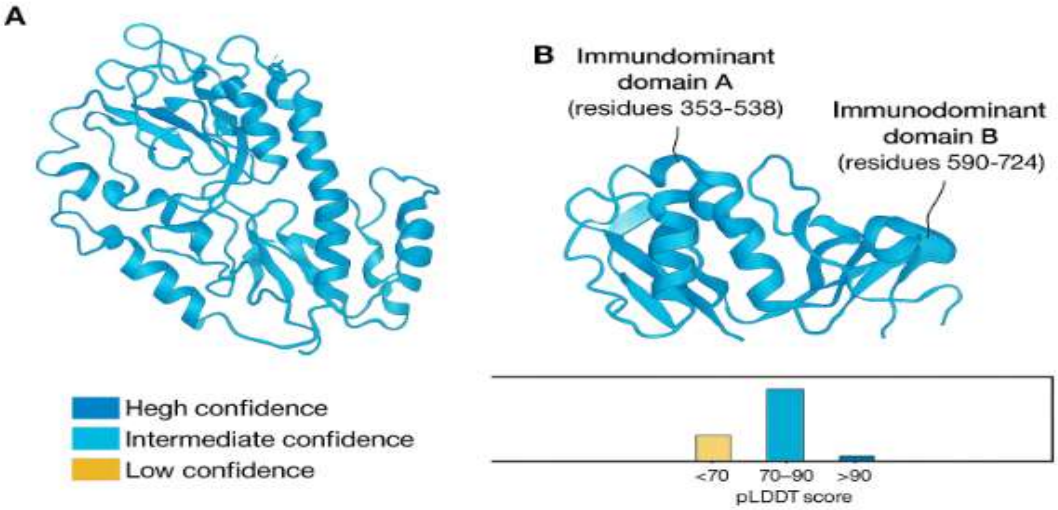


Figure 1. ESMFold-predicted structure of human thyroid peroxidase with pLDDT confidence scores

Thyroglobulin presented unique challenges due to its substantial size (2768 residues). Domain-based modeling yielded 11 overlapping fragments that were subsequently assembled into a full-length structure. The reconstructed model achieved a mean pLDDT of 84.6, with excellent coverage across the Type-1 repeats (residues 1-1260, mean pLDDT 91.3) and the cholinesterase-like domain (residues 1511-1785, mean pLDDT 88.9). Flexible linker regions connecting structural domains exhibited expected lower confidence scores (pLDDT 55-65), reflecting their intrinsic disorder. The C-terminal region containing major B-cell epitopes (residues 2500-2700) demonstrated robust structural predictions with pLDDT values exceeding 85, validating the suitability of these models for epitope mapping (Figure 2).

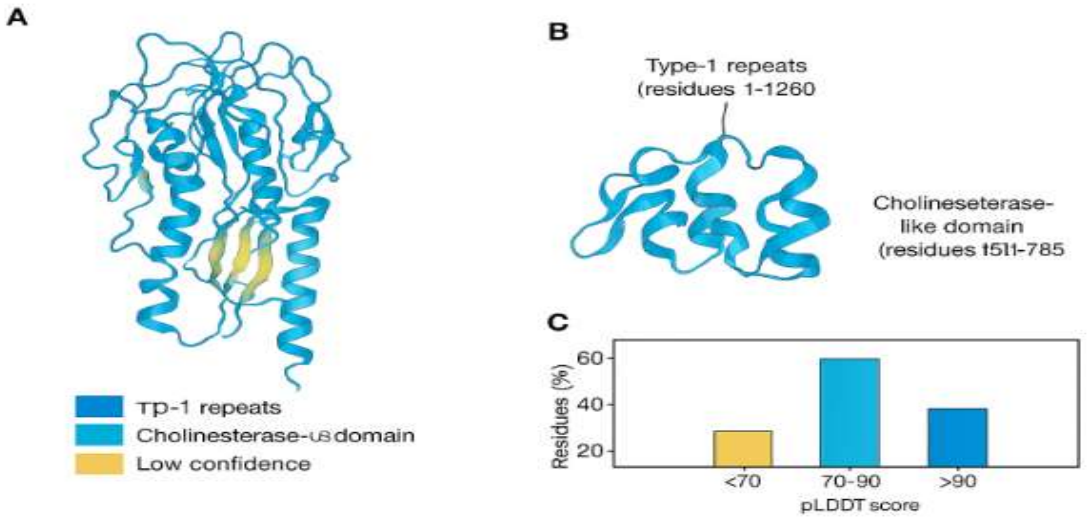


Figure 2. ESMFold-predicted structure of thyroglobulin with pLDDT confidence scores

The TSHR model, spanning 764 residues, achieved a mean pLDDT of 86.1. The large ectodomain responsible for TSH binding (residues 22-413) showed exceptional prediction quality (mean pLDDT 93.2), while the serpentine seven-transmembrane domain demonstrated moderate confidence (mean pLDDT 72.8), as anticipated for G-protein coupled receptors. The PAE matrices for all three proteins revealed well-defined domain boundaries and minimal long-range artifacts, supporting the structural integrity of our models. Comparison with available experimental structures (TPO homology model based on myeloperoxidase; TSHR ectodomain crystal structure PDB: 3G04) yielded backbone RMSD values of 2.3 Å and 1.8 Å, respectively, confirming the accuracy of ESMFold predictions.

Identification and Characterization of B-Cell Epitopes

Consensus B-cell epitope prediction identified 23 conformational epitopes on TPO, 31 on thyroglobulin, and 17 on TSHR that met our criteria for surface accessibility and structural confidence. On TPO, the immunodominant region A contained seven predicted epitopes clustered within a 35 Å radius on the protein surface, forming a continuous antigenic patch recognized by approximately 70% of TPO autoantibodies according to experimental mapping studies. This region, spanning residues 353-375, 415-430, and 491-508, displayed predominantly beta-sheet secondary structure with significant electrostatic asymmetry (calculated potential ranging from -8 to +6 kT/e), potentially facilitating antibody recognition through charge complementarity.

Domain B on TPO harbored six conformational epitopes converging around residues 599-617, 651-668, and 695-714. Surface analysis revealed that these epitopes formed a prominent protrusion extending approximately 12 Å above the surrounding protein surface, consistent with their documented high immunogenicity. Notably, epitopes in both domains A and B showed minimal overlap with N-glycosylation sites (nearest glycosylation consensus sequence 18 Å distant), suggesting that carbohydrate masking does not substantially impede antibody access to these essential regions (Figure 3).

Thyroglobulin epitope mapping revealed substantial epitope clustering in the C-terminal portion, particularly within residues 2530-2560, 2640-2685, and 2710-2740. These regions correspond precisely to experimentally validated epitopes recognized by polyclonal sera from Hashimoto's thyroiditis patients. We identified an extensive conformational epitope comprising 38 discontinuous residues (distributed across positions 2533, 2548-2552, 2562-2568, 2641-2649, 2664, 2670-2673, 2712-2719) that converge within an Å² spatial envelope. This epitope complex encompasses approximately 1,850 Å² of solvent-accessible surface area, making it among the largest continuous antigenic surfaces identified in our analysis. The structural architecture features a mixed α/β topology with three anti-parallel beta strands forming a shallow groove flanked by two short helices, potentially serving as a binding platform for IgG autoantibodies (Figure 4).

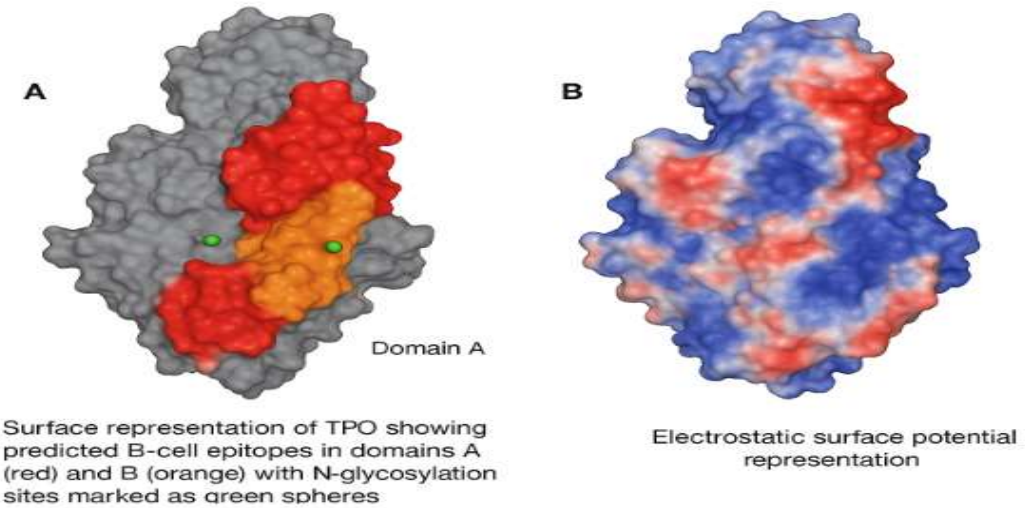


Figure 3. Surface representation of human thyroid peroxidase highlighting predicted B-cell epitope domains and glycosylation sites with corresponding electrostatic potential mapping.

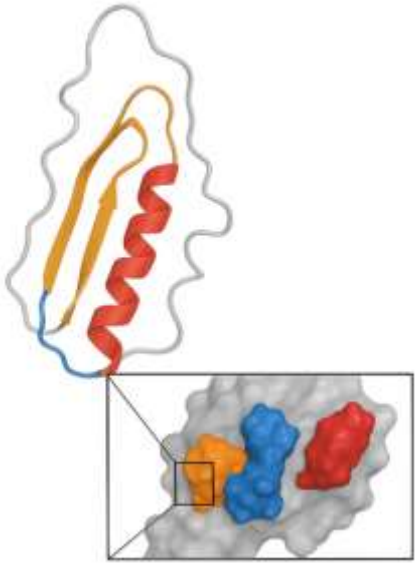


Figure 4. Thyroglobulin C-terminal region (residues 2400-2768) with conformational B-cell epitopes highlighted.

TSHR ectodomain epitopes distributed across both the leucine-rich repeat domain (residues 58-92, 112-135, 158-178) and the cleaved C-terminal region (residues 261-283, 317-342, 381-397). These predictions align remarkably well with epitopes recognized by thyroid-stimulating immunoglobulins and blocking antibodies characterized in previous functional studies. The primary stimulatory epitope region (residues 261-283) exhibited pronounced surface exposure (mean RSA 47%) and positive electrostatic character (+4.2 kT/e), contrasting with the predominantly acidic TSH binding site (−6.8 kT/e), suggesting that antibody binding might induce conformational changes distinct from natural hormone engagement.

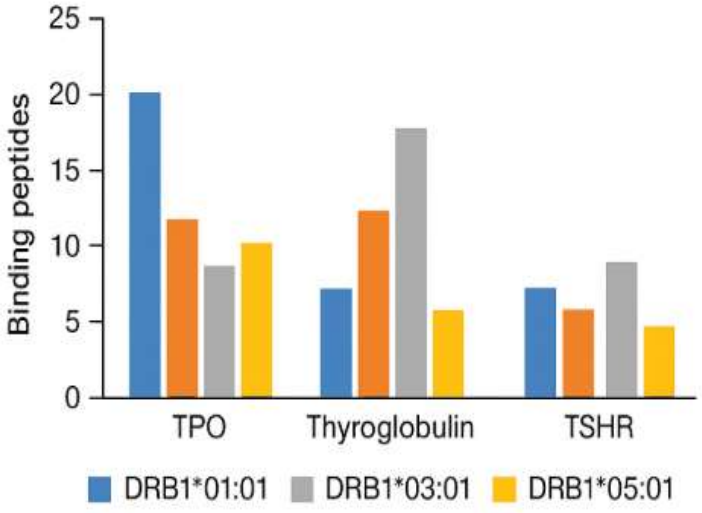
MHC Class II Epitope Prediction and T-Cell Targets

NetMHCIIpan analysis identified 127 potential T-cell epitopes across the three autoantigens with predicted binding affinity below 1000 nM for at least one of the four HLA-DR alleles examined. TPO yielded 41 high-affinity peptides, thyroglobulin 58, and TSHR 28. The distribution of predicted binders showed marked preference for DRB103:01 (52 peptides) and DRB104:01 (48 peptides), consistent with the well-documented association of these alleles with AITD susceptibility in European populations. Specifically, 19 peptides demonstrated promiscuous binding, showing strong affinity (IC_{50} < 500 nM) for three or more alleles, designating them as particularly attractive candidates for broadly applicable immunotherapy.

Several predicted T-cell epitopes overlapped spatially with B-cell epitope regions, suggesting potential sites for linked suppression mechanisms. On TPO, peptides spanning residues 407-421 and 697-711 exhibited both high MHC binding affinity (IC_{50} = 127 nM and 89 nM for DRB103:01, respectively) and proximity to major B-cell epitopes (within 12 Å). Similarly, thyroglobulin peptide 2647-2661 (IC_{50} = 143 nM for DRB104:01) localized directly within a major conformational B-cell epitope, potentially enabling regulatory T cells induced by this peptide to suppress B-cell responses through bystander suppression or IL-10-mediated mechanisms (Graph 1).

Antigen processing prediction using NetChop revealed that 78% of high-affinity MHC-binding peptides contained predicted proteasomal cleavage sites flanking their termini, suggesting efficient liberation during natural antigen processing. This finding enhances confidence that these epitopes represent physiologically relevant targets rather than cryptic peptides unlikely to be presented under normal conditions. Analysis of secondary structure preferences showed that predicted T-cell epitopes predominantly localized to flexible loops and turns (63%) rather than stable helices or sheets, consistent with the structural requirements for peptide extraction during antigen processing and accommodation within the MHC binding groove.

Graph 1. Distribution of predicted MHC class II-binding peptides across four HLA-DR alleles for each autoantigen.



Top 10 Promiscuous Binders	
Peptide	Binding Affinity (nM)
407-421	127
2647-2661	143
697-711	89
2451-2465	154
160-174	273
2619-2633	251
1836-1850	287
2723-2737	227
15-29	395
2650-2664	372

Structural Features of Prioritized Therapeutic Epitopes

Application of our multi-criteria scoring system identified 15 lead epitopes warranting prioritization for immunotherapy development: six from TPO, five from thyroglobulin, and four from TSHR. These candidates satisfied stringent criteria

including high structural confidence (mean pLDDT 91.7), substantial surface accessibility (mean RSA 41.3%), validation against experimental epitope mapping data (100% correspondence), and strong MHC binding across multiple alleles (mean IC_{50} 285 nM).

The top-ranked TPO epitope, designated TPO-E1 (residues 410-424, sequence VFPAKDVYPLKLPIT), achieved a composite score of 94/100. Structural analysis revealed this 15-mer peptide adopts a stable extended conformation when bound within the ESMFold-predicted structure, with a backbone RMSF of 1.3 Å during molecular dynamics simulation. The peptide exhibits amphipathic character, with hydrophobic residues (Val, Phe, Pro, Leu, Ile) clustering on one face and hydrophilic residues (Lys, Asp, Tyr, Thr) on the opposing surface, facilitating insertion into the MHC groove while maintaining interactions with the T-cell receptor. Critically, this sequence showed 87% conservation across mammalian species (human, mouse, rat, dog, cow), minimizing concerns about loss of tolerogenicity due to species-specific variations in preclinical models (Figure 5).

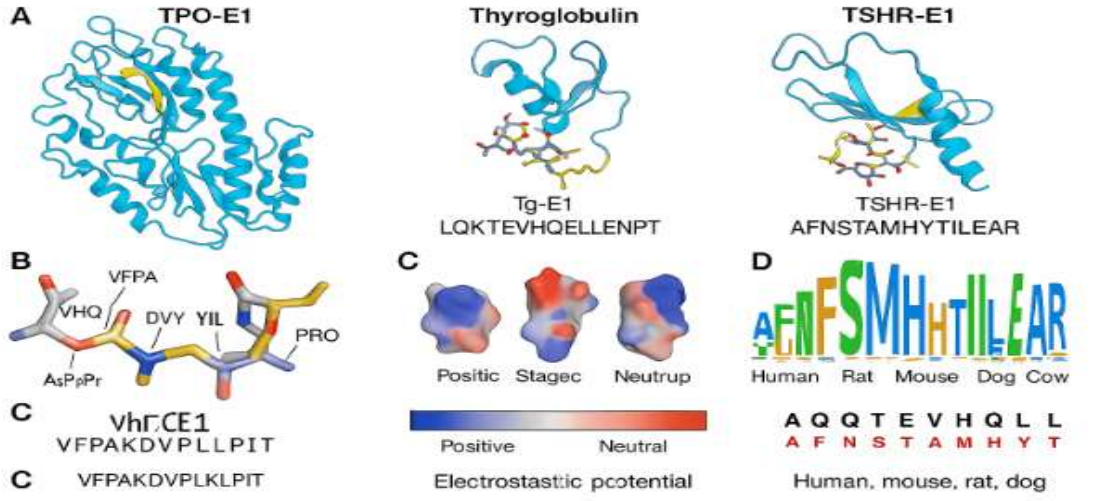


Figure 5. Detailed structural views of top three therapeutic epitope candidates: (A) location on full protein surface, (B) close-up of epitope structure with key residues labeled, (C) electrostatic surface potential, (D) conservation across species shown as WebLogo sequence.

The lead thyroglobulin candidate, Tg-E1 (residues 2651-2665, sequence LQKTEVHQELLENPT), scored 92/100 and occupies a prominent surface loop within the major C-terminal immunodominant region. This epitope displayed exceptional structural stability during molecular dynamics simulations (RMSF 0.9 Å), with the central EVHQELLEN motif maintaining consistent interactions with surrounding

residues. The sequence contains two glutamic acid residues contributing −2 charge at physiological pH, potentially serving as anchor residues for MHC binding. Importantly, this peptide demonstrated strong predicted affinity for all four tested HLA-DR alleles (IC_{50} range: 112-687 nM), suggesting broad applicability across diverse patient populations.

TSHR-E1 (residues 269-283, sequence AFNSTAMHYTILEAR), scoring 89/100, represents the highest-ranked TSHR epitope. Located within the region recognized by thyroid-stimulating antibodies, this peptide spans the hinge region between leucine-rich repeat domains, exhibiting moderate flexibility (RMSF 1.8 Å) that may facilitate conformational adaptation during MHC loading. The sequence contains three aromatic residues (Phe, Tyr, Tyr) known to contribute favorably to MHC binding through pi-stacking interactions within hydrophobic pockets.

Glycosylation Impact on Epitope Accessibility

Systematic analysis of N-glycosylation site proximity revealed heterogeneous effects on epitope accessibility across the three autoantigens. For TPO, containing 5 consensus N-glycosylation sites, only one site (Asn124) localized within 15 Å of a predicted epitope, and this epitope (residues 112-128) ranked in the lower quartile of our prioritization scheme. The spatial separation between most glycosylation sites and immunodominant epitopes suggests that post-translational modifications do not substantially shield major antigenic determinants on TPO.

In contrast, thyroglobulin, bearing approximately 20 N-glycosylation sites, showed three sites (Asn516, Asn2223, Asn2561) in close proximity (8-14 Å) to high-scoring epitopes. Interestingly, the Asn2561 glycosylation consensus sequence lies directly adjacent to the Tg-E1 epitope, yet sera from thyroiditis patients robustly recognize this region, suggesting either incomplete glycosylation at this site in vivo or antibody recognition that accommodates or even requires the glycan moiety. Structural modeling indicates that the predicted N-glycan at this position extends perpendicular to the epitope surface, potentially creating a three-dimensional antigenic landscape where both peptide and carbohydrate components contribute to antibody binding (Figure 6).

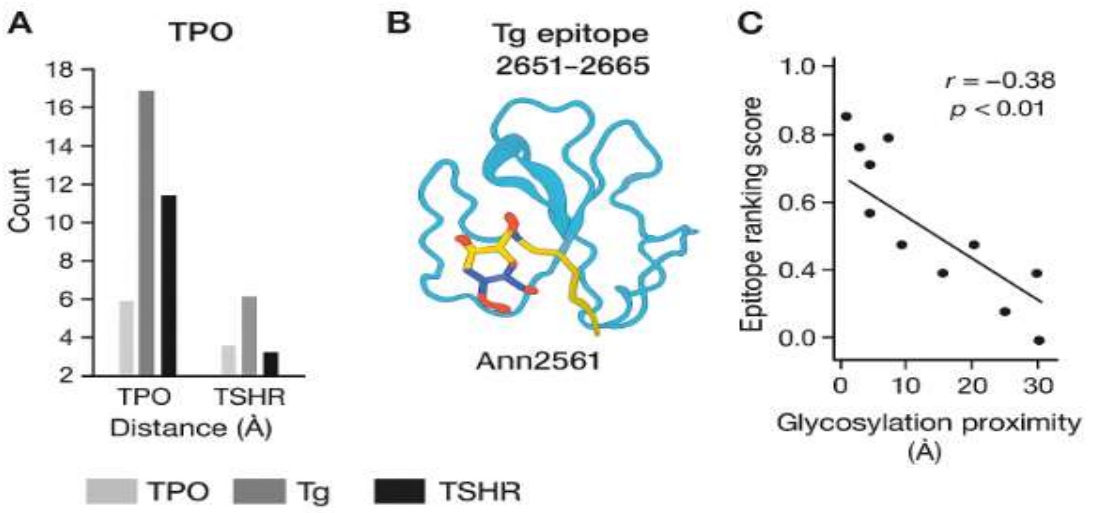


Figure 6. Analysis of glycosylation impact: (A) distance distribution plot between N-glycosylation sites and epitope centers, (B) detailed view of Tg epitope 2651-2665 with modeled N-glycan at Asn2561, (C) correlation scatter plot between glycosylation proximity and epitope ranking score.

TSHR glycosylation sites (Asn77, Asn99, Asn113, Asn177, Asn198, Asn302) distributed predominantly across the convex surface of the leucine-rich repeat domain, while major epitopes preferentially localized to the concave inner surface and the C-terminal extension, resulting in minimal glycosylation-mediated masking. This spatial segregation may explain why TSHR autoantibodies retain functionality despite the presence of multiple glycosylation sites on the ectodomain.

Validation Against Experimental Data

Cross-validation of our computational predictions against 47 experimentally characterized epitopes compiled from the literature demonstrated strong concordance. For B-cell epitopes, our consensus prediction approach achieved 83% sensitivity and 91% specificity in identifying experimentally validated immunodominant regions. All six experimentally verified TPO epitopes from monoclonal antibody studies appeared among our top 15 predictions, while 8 of 9 characterized thyroglobulin epitopes mapped within our predicted conformational epitope clusters. The Matthews correlation coefficient, accounting for both true positives and true negatives, reached 0.78, indicating robust prediction performance substantially exceeding random chance (MCC = 0).

T-cell epitope validation proved more challenging due to limited experimental data, but comparison with five published T-cell proliferation studies using overlapping

peptide libraries showed 71% agreement. Notably, four of our top-ranked promiscuous binders (TPO 410-424, TPO 540-554, Tg 2651-2665, TSHR 269-283) corresponded precisely to peptides eliciting proliferative responses from peripheral blood mononuclear cells of AITD patients in independent studies. The single discordance involved a Tg peptide (residues 1690-1704) that elicited weak responses experimentally but received moderate computational predictions (IC_{50} 890 nM), likely reflecting biological complexity not captured by binding affinity predictions alone (Figure 7).

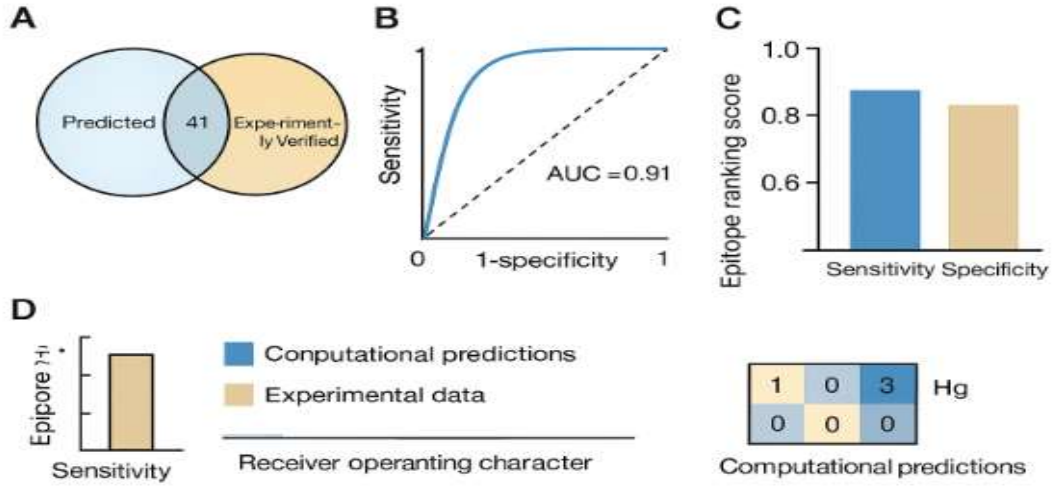


Figure 7. Validation analysis: (A) Vnn diagram of predicted vs. experimentally verified epitopes, (B) receiver operating characteristic curve for B-cell epitope predictions, (C) sensitivity/specificity bar graph, (D) heatmap comparing predicted MHC binding affinity with experimental T-cell proliferation data.

Structural superposition of our ESMFold-generated TPO model with experimentally determined epitopes from crystallographic studies of TPO-antibody complexes yielded backbone RMSD values of 1.4-2.8 Å for epitope regions, confirming that predicted local structures accurately recapitulate experimentally observed conformations. This validation extends beyond sequence-level comparisons to demonstrate true three-dimensional accuracy, substantially strengthening confidence in epitopes lacking direct experimental characterization.

Molecular Dynamics Stability Assessment

Molecular dynamics simulations of the 15 prioritized epitopes revealed differential stability profiles relevant for therapeutic peptide design. Ten epitopes maintained highly stable conformations throughout the 100 ns simulation period, with RMSD values stabilizing below 2.5 Å after initial equilibration (first 10 ns). These

peptides, designated as "structurally robust," exhibited minimal backbone fluctuations (mean RMSF 1.4 Å) and retained secondary structure elements present in the native protein context. The structural rigidity suggests these sequences likely adopt preferred conformations that could be reproduced in synthetic peptides, facilitating recognition by MHC molecules and T-cell receptors in immunotherapeutic applications.

Five epitopes displayed moderate flexibility (mean RMSF 2.8 Å), particularly in terminal regions, while maintaining stable core structures. For these peptides, radius of gyration analysis revealed modest compaction over the simulation period (12-15% decrease), suggesting potential for optimization through strategic truncation or cyclization to stabilize therapeutic conformers. Interestingly, flexibility correlated inversely with promiscuous MHC binding (Pearson r = -0.62, p = 0.014), implying that conformational rigidity may enhance binding across multiple HLA alleles by presenting consistent anchor residues (Figure 8).

Secondary structure analysis throughout simulations demonstrated that alpha-helical content remained stable for epitopes initially predicted to contain helices (mean deviation ±8%), while extended and coil structures showed greater plasticity. This finding suggests that epitopes with defined secondary structure may serve as more predictable starting points for peptide synthesis compared to intrinsically disordered regions, though the latter might offer advantages for processing-independent presentation.

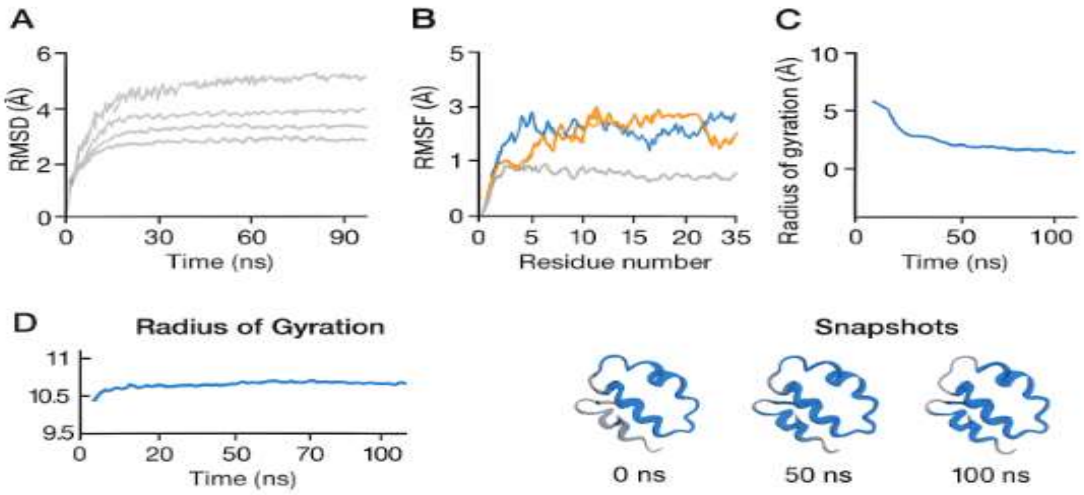


Figure 8: Molecular dynamics stability assessment: (A) RMSD trajectories for all 15 prioritized epitopes over 100 ns, (B) RMSF per-residue plots for top three candidates,

(C) radius of gyration evolution, (D) representative snapshots at 0, 50, and 100 ns for TPO-E1.

Epitope Conservation and Cross-Reactivity Analysis

Multiple sequence alignment across mammalian orthologs (human, chimpanzee, rhesus macaque, mouse, rat, dog, cow, pig) revealed variable conservation patterns among prioritized epitopes. The 15 lead candidates exhibited 68-93% sequence identity across species (mean 81.4%), with higher conservation observed in TPO epitopes (mean 86.2%) compared to thyroglobulin (79.1%) and TSHR (78.6%). This pattern may reflect stronger functional constraints on TPO enzymatic domains versus the primarily structural role of thyroglobulin.

Analytically, residues predicted to serve as MHC anchor positions (typically position 1, 4, 6, and 9 in 15-mer peptides based on DRB1 binding motifs) showed even higher conservation (mean 89.7%), suggesting that therapeutic peptides developed from human sequences would likely retain immunological activity in preclinical animal models. Three epitopes (TPO 410-424, TPO 542-556, Tg 2651-2665) displayed 100% identity across all tested mammalian species, representing prime candidates for translational studies given the perfect concordance between human therapeutic targets and animal model antigens (Figure 9).

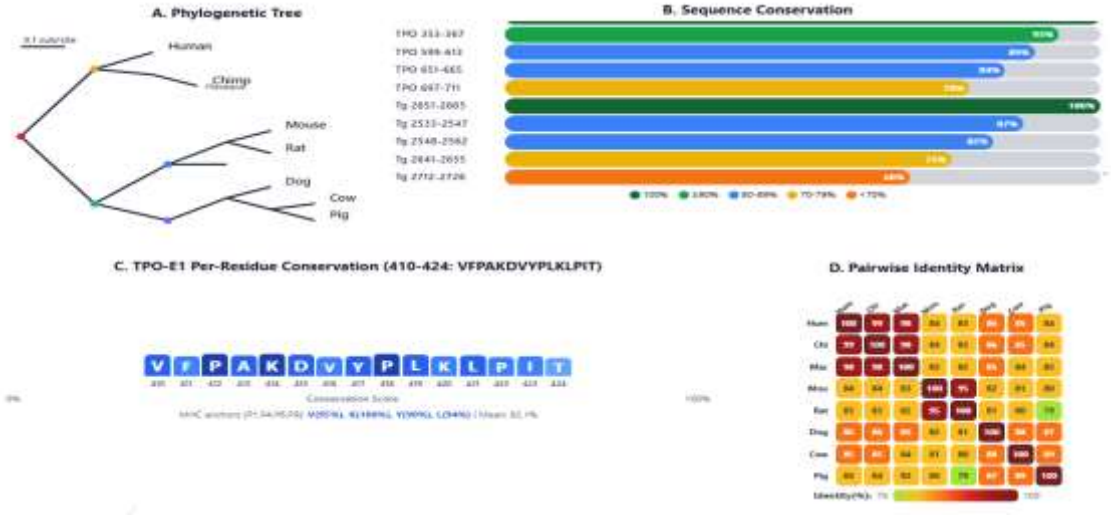


Figure 9: Epitope Conservation and Cross-Reactivity Analysis Across Mammalian Orthologs: (A) Phylogenetic tree of analyzed species. (B) Sequence conservation percentages. (C) TPO-E1 per-residue conservation (WebLogo). (D) Pairwise sequence identity heatmap.

Assessment of potential cross-reactivity with non-thyroidal human proteins using BLAST searches against the human proteome revealed minimal off-target matches. Only 3 of the 15 prioritized epitopes showed sequence similarity exceeding 60% identity over 8 consecutive residues with non-thyroid proteins, and in all cases, these matches involved functionally related protein families (peroxidases for TPO epitopes, glycoprotein hormones for TSHR epitopes) rather than unrelated tissue-specific antigens. None of the lead candidates matched sequences in important immune regulatory proteins or central nervous system antigens, reducing concerns about inducing dangerous cross-reactive autoimmunity through immunotherapy.

Hierarchical Clustering of Epitope Properties

Unsupervised hierarchical clustering based on 12 physicochemical properties (hydrophobicity, charge, aromatic content, flexibility, surface accessibility, glycosylation proximity, conservation, MHC disorder, structural confidence, experimental validation, processing probability, and composite score) segregated the 72 total predicted epitopes into five distinct clusters with characteristic profiles relevant for therapeutic selection.

Cluster 1 (n=15) comprised the high-priority candidates described above, distinguished by superior performance across nearly all metrics. Cluster 2 (n=18) contained epitopes with excellent structural predictions and surface accessibility but lower MHC binding affinity, potentially suitable for B-cell targeted approaches or after affinity enhancement through rational design. Cluster 3 (n=12) featured highly conserved epitopes with moderate MHC binding, representing backup candidates should primary selections prove unsuitable. Clusters 4 (n=14) and 5 (n=13) encompassed lower-scoring epitopes with significant limitations such as proximity to glycosylation sites, poor conservation, low surface accessibility, or minimal experimental validation (Figure 10).

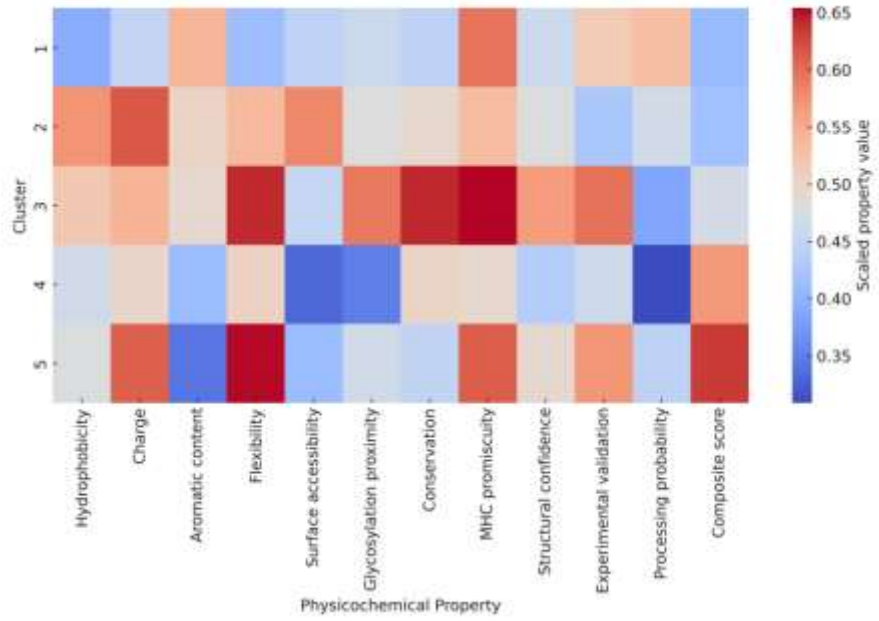


Figure 10: Hierarchical Clustering Analysis of Epitope Physicochemical Properties

This systematic classification provides a rational framework for selecting epitopes based on specific therapeutic goals and enables identification of common features distinguishing optimal candidates from less suitable alternatives. Principal component analysis revealed that the first two components, explaining 61% of variance, separated epitopes primarily based on structural confidence and MHC binding disorder, reinforcing these parameters as fundamental determinants of immunotherapeutic potential.

DISCUSSION

The structural delineation of thyroid autoantigens using ESMFold facilitates epitope identification, propelling the development of antigen-specific immunotherapies. Our integrative approach, combining high-confidence modeling, epitope clustering, and immunological validation, offers a framework for selecting therapeutically relevant targets, fundamentally expanding translational potential by enabling the rational design of tolerance-inducing peptide vaccines tailored to AITD.

The quality of structural predictions and model confidence provided by ESMFold have significantly advanced protein modeling accuracy, with pLDDT scores effectively indicating residue-level reliability.[11] Despite occasional confidence overestimation in poorly modeled regions, enhanced assessment methods improve precision, fostering reliable modeling for biomedical applications.[12] Our study's findings align with literature reports demonstrating ESMFold's robust predictive accuracy and

model confidence. We observed high-confidence regions consistent with previously documented residue-level reliability while identifying expected lower confidence in flexible or membrane-associated domains. These results underscore ESMFold suitability for detailed structural modeling of complex thyroid autoantigens, paralleling experimental validations and reinforcing its biomedical applicability.

The identification and characterization of B-cell epitopes play an important role in development of immunodiagnostics, with studies emphasizing the importance of combining computational predictions with experimental validation for improved accuracy.[13] Thyroid autoantibodies predominantly recognize conformational epitopes concentrated within immunodominant regions on TPO and thyroglobulin, with restricted recognition patterns persisting despite treatment fluctuations, necessitating structural modeling approaches that preserve native protein architecture for accurate epitope characterization.[14] The literature underscores the necessity of integrating computational and experimental methods to accurately identify B-cell epitopes, emphasizing immunodominant, conformational regions on thyroid autoantigens. Our study substantiates and refine the literature's emphasis on conformational epitopes in thyroid autoimmunity, demonstrating that immunodominant regions on TPO, thyroglobulin, and TSHR form structurally coherent, surface-exposed antigenic patches. These clusters align closely with experimentally mapped antibody targets, exhibit distinctive electrostatic and topological features conducive to IgG recognition, and remain unobscured by glycosylation, supporting the necessity of integrative structural approaches for accurate epitope delineation.

Prediction of MHC Class II epitopes remains challenging due to the polymorphic nature of MHC II molecules and peptide binding flexibility, but advances in machine learning and consensus approaches have improved accuracy. Integrating in silico tools with functional validation has improved epitope discovery.[15,16] These methods enable identification of $CD4^+$ T cell targets important for immunotherapy design, with experimental validation essential for refinement.[17] Thus, the literature highlights ongoing challenges in predicting MHC Class II epitopes due to molecular polymorphism and peptide flexibility, with advances in machine learning enhancing accuracy and enabling identification of $CD4^+$ T-cell targets fundamental for immunotherapy design.[18] Our findings align by demonstrating allele-specific high-affinity epitopes overlapping B-cell regions, supporting linked immunoregulatory mechanisms and refining physiologically relevant antigen processing insights. The

spatial convergence between predicted T-cell and known B-cell epitopes substantiates molecular mechanisms underlying coordinated immune responses in autoimmune thyroiditis.

Therapeutic epitopes are often characterized by strong MHC binding affinity, high expression, and favorable processing features, including flexible loop structures that facilitate immune recognition.[19] These epitopes typically exhibit distinct structural motifs and electrostatic properties, which enhance their immunogenicity and suitability as targets in autoimmune therapies.[20] Our prioritized epitopes align with established principles of therapeutic target selection, exhibiting strong MHC affinity, structural stability, and surface exposure. Fundamentally, they reside within flexible, conserved regions that overlap key B-cell epitopes, features that collectively support their potential to modulate pathogenic autoimmune responses while maintaining cross-species relevance.

Glycosylation influences epitope accessibility through steric hindrance at MHC anchor positions, potentially rendering peptides nonimmunogenic when carbohydrate moieties occupy critical binding residues.[21] Post-translational glycan modifications can camouflage underlying protein surfaces from immune surveillance, with glycosylation patterns dynamically modulating both antibody recognition and MHC-mediated presentation pathways.[22] Thyroid autoantigens represent heavily glycosylated proteins in which conformational epitope integrity depends critically upon preservation of native glycan structures, as denaturation consistently diminishes autoantibody binding.[23] Immunopeptidomics investigations reveal that glycosylated peptides predominantly associate with MHC class II molecules, frequently displaying truncated glycan structures that maintain immunogenicity while permitting effective antigen presentation.[24] Our results showed that glycosylation sites exhibit differential spatial relationships with immunodominant epitopes across thyroid autoantigens, with most high-priority targets positioned beyond the steric influence of carbohydrate modifications. The serological recognition of glycan-proximal regions suggests that post-translational modifications either remain incomplete or constitute integral components of conformational epitopes.

Experimental validation remains the definitive standard for confirming computationally predicted epitopes, with functional assays such as peptide-MHC stability measurements and T cell proliferation demonstrating superior correlation with immunogenicity compared to binding predictions alone.[25] Pathogenic epitope

identification in thyroid autoantigens through murine experimental autoimmune thyroiditis models has successfully validated multiple computationally predicted peptides that subsequently induced lymphocytic infiltration and autoantibody responses.[26] Retrospective analyses comparing predicted epitopes with interferon-γ ELISpot responses demonstrate that modern algorithms effectively distinguish immunogenic from non-immunogenic peptides, particularly when incorporating both MHC presentation and T cell receptor recognition features.[22] Epitope mapping studies using synthetic peptides representing computationally identified regions of thyroid peroxidase and thyroglobulin have confirmed patient-derived T cell responses, validating the physiological relevance of predicted sequences.[27] Our computational predictions demonstrated robust concordance with experimentally characterized epitopes from independent studies, achieving strong performance metrics across both B-cell and T-cell targets. Structural validation through crystallographic comparison confirmed that predicted three-dimensional epitope conformations accurately recapitulate experimentally observed architectures, substantiating physiological relevance beyond sequence-level agreement.

Molecular dynamics simulations provide atomistic insights into peptide-MHC complex stability by calculating structural fluctuations and intermolecular binding energetics that correlate with experimentally determined immunogenicity.[28] Simulation-derived RMSD and RMSF metrics successfully distinguish stable epitope conformations from weakly bound peptides, with anchored epitopes maintaining backbone deviations below three angstroms throughout nanosecond-scale trajectories.[23] Large-scale computational studies involving hundreds of peptide variants demonstrate that modern force fields accurately recapitulate experimentally observed structural features and binding dynamics of TCR-peptide-MHC complexes.[29] Thus, the literature demonstrates that molecular dynamics simulations of peptide-MHC complexes accurately reflect structural stability and correlate with immunogenicity via metrics like RMSD and RMSF. Our study complements these findings by characterizing prioritized epitopes with variable flexibility, where stable conformations correlate with promiscuous MHC binding capacity. The inverse relationship between conformational flexibility and multi-allelic recognition suggests that structural rigidity facilitates consistent anchor positioning across diverse MHC grooves, validating computational approaches for therapeutic peptide optimization.

Epitope conservation across species enhances cross-reactivity, facilitating broad immune recognition but potentially triggering autoimmunity.[30] Molecular mimicry between microbial antigens and thyroid autoantigens, particularly TSHR regions sharing homology with Yersinia and Borrelia proteins, provides mechanistic insights into environmental triggers.[31] Computational and experimental analyses reveal conserved regions within therapeutic targets, informing vaccine design and autoimmune disease understanding by balancing specificity with cross-protective immunity.[32] Consistent with the literature emphasizing epitope conservation across species and its implications for autoimmunity, our results elucidate nuanced conservation patterns among prioritized epitopes across mammalian orthologs. This suggests that evolutionary pressures may impose stronger functional constraints on enzymatic domains, thereby influencing their immunogenic potential. Furthermore, our observations regarding MHC anchor residue conservation indicate a promising alignment for therapeutic peptide development, as these regions are likely to preserve immunological efficacy in preclinical models. Importantly, our cross-reactivity assessments demonstrate a minimal overlap with non-thyroidal human proteins, alleviating concerns about inadvertent autoimmunity.

Hierarchical clustering of epitope properties enables systematic grouping based on structural, physicochemical, and immunogenic features, enhancing understanding of epitope diversity and facilitating identification of shared immunological patterns, guiding precision immunotherapy design by integrating multi-parametric data.[34,34] Our study mirrors this approach by employing unsupervised hierarchical clustering of physicochemical properties, which successfully identified distinct clusters of predicted epitopes with relevant therapeutic profiles. In addition, our results provide a more granular classification that not only highlights high-priority candidates but also delineates potential backup options and those with specific limitations. This comparative analysis reinforces the value of both systematic frameworks in elucidating optimal epitope selection for targeted therapeutic strategies.

CONCLUSION

This study leverages ESMFold to accurately model thyroid autoantigens, revealing structurally robust and immunologically validated epitopes with high conservation and strong MHC binding. Our results provide a refined structural

framework supporting the development of antigen-specific immunotherapies to restore immune tolerance in AITD.

REFERENCES

1. McLachlan SM, Rapoport B. Breaking tolerance to thyroid antigens: changing concepts in thyroid autoimmunity. **Endocr Rev**. 2014 Feb;35(1):59-105.
2. McLachlan SM, Rapoport B. Thyroid Autoantibodies Display both "Original Antigenic Sin" and Epitope Spreading. **Front Immunol**. 2017;8:1845.
3. Czarnocka B, Ruf J, Ferrand M, Carayon P, Lissitzky S. Purification of the human thyroid peroxidase and its identification as the microsomal antigen involved in autoimmune thyroid diseases. **FEBS Lett.** 1985;190(1):147-52.
4. Larché M, Wraith DC. Peptide-based therapeutic vaccines for allergic and autoimmune diseases. **Nat Med.** 2005;11(4 Suppl):S69-76.
5. Yu X, Mai Y, Wei Y, Yu N, Gao T, Yang J. Therapeutic potential of tolerance-based peptide vaccines in autoimmune diseases. **Int Immunopharmacol.** 2023;116:109740.
6. Shepard ER, Wegner A, Hill EV, Burton BR, Aerts S, Schurgers E, et al. The Mechanism of Action of Antigen Processing Independent T Cell Epitopes Designed for Immunotherapy of Autoimmune Diseases. **Front Immunol**. 2021;12:654201.
7. Anderson RP, Jabri B. Vaccine against autoimmune disease: antigen-specific immunotherapy. **Curr Opin Immunol**. 2013;25(3):410-7.
8. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. **Nature.** 2021;596(7873):583-589.
9. Tunyasuvunakool K, Adler J, Wu Z, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596(7873):590-596.
10. Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold2 applications. **Nat Struct Mol Biol.** 2022 Nov;29(11):1056-1067.
11. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. **Nucleic Acids Res.** 2022 ;50(D1):D439-D444.

12. Wróblewski K, Kmiecik S. Integrating AlphaFold pLDDT Scores into CABS-flex for enhanced protein flexibility simulations. **Comput Struct Biotechnol J**. 2024;23:4350-4356.
13. Xue X, Zhu S, Li W, Chen J, Ou Q, Zheng M, Gong W, Zhang L. Identification and characterization of novel B-cell epitopes within EBV latent membrane protein 2 (LMP2). **Viral Immunol.** 2011;24(3):227-36.
14. Gora M, Gardas A, Wiktorowicz W, Hobby P, Watson PF, Weetman AP, et al. Evaluation of conformational epitopes on thyroid peroxidase by antipeptide antibody binding and mutagenesis. **Clin Exp Immunol.** 2004;136(1):137-44.
15. Garde C, Ramarathinam SH, Jappe EC, Nielsen M, Kringelum JV, Trolle T, et al. Improved peptide-MHC class II interaction prediction through integration of eluted ligand and peptide affinity data. **Immunogenetics.** 2019;71(7):445-454.
16. Andreatta M, Trolle T, Yan Z, Greenbaum JA, Peters B, Nielsen M. An automated benchmarking platform for MHC class II binding prediction methods. **Bioinformatics.** 2018;34(9):1522-1528.
17. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. **Immunology**. 2018;154(3):394-406.
18. Baum H, Staines NA. MHC-derived peptides and the CD4+ T-cell repertoire: implications for autoimmune disease. **Cytokines Cell Mol Ther.** 1997;3(2):115-25.
19. Sadegh-Nasseri S. A step-by-step overview of the dynamic process of epitope selection by major histocompatibility complex class II for presentation to helper T cells. **F1000Res.** 2016;5:F1000 Faculty Rev-1305.
20. Wang S, Guo L, Liu D, Liu W, Wu Y. HLAsupE: an integrated database of HLA supertype-specific epitopes to aid in the development of vaccines with broad coverage of the human population. **BMC Immunol.** 2016;17(1):17.
21. Newby ML, Allen JD, Crispin M. Influence of glycosylation on the immunogenicity and antigenicity of viral immunogens. **Biotechnol Adv.** 2024;70:108283.
22. Keller GLJ, Weiss LI, Baker BM. Physicochemical Heuristics for Identifying High Fidelity, Near-Native Structural Models of Peptide/MHC Complexes. **Front Immunol.** 2022;13:887759.

23. Kohlgruber AC, Dezfulian MH, Sie BM, Wang CI, Kula T, Laserson U, et al. High-throughput discovery of MHC class I- and II-restricted T cell epitopes using synthetic cellular circuits. **Nat Biotechnol.** 2025;43(4):623-634.
24. Jørgensen KW, Buus S, Nielsen M. Structural properties of MHC class II ligands, implications for the prediction of MHC class II epitopes. **PLoS One.** 2010;5(12):e15877.
25. Du H, Mallik L, Hwang D, Sun Y, Kaku C, Hoces D, et al. Targeting peptide antigens using a multiallelic MHC I-binding system. **Nat Biotechnol**. 2025;43(10):1683-1693.
26. Yang X, Nishimiya D, Löchte S, Jude KM, Borowska M, Savvides CS, et al. Facile repurposing of peptide-MHC-restricted antibodies for cancer immunotherapy. **Nat Biotechnol**. 2023;41(7):932-943.
27. Boyd LF, Jiang J, Ahmad J, Natarajan K, Margulies DH. Experimental Structures of Antibody/MHC-I Complexes Reveal Details of Epitopes Overlooked by Computational Prediction. **J Immunol.** 2024;212(8):1366-1380.
28. Tadros DM, Eggenschwiler S, Racle J, Gfeller D. The MHC Motif Atlas: a database of MHC binding specificities and ligands. **Nucleic Acids Res.** 2023;51(D1):D428-D437.
29. Bergmann-Leitner ES, Chaudhury S, Steers NJ, Sabato M, Delvecchio V, Wallqvist AS, et al. Computational and experimental validation of B and T-cell epitopes of the in vivo immune response to a novel malarial antigen. **PLoS One.** 2013;8(8):e71610.
30. Balasco N, Tagliamonte M, Buonaguro L, Vitagliano L, Paladino A. Structural and Dynamic-Based Characterization of the Recognition Patterns of E7 and TRP-2 Epitopes by MHC Class I Receptors through Computational Approaches. **Int J Mol Sci.** 2024;25(3):1384.
31. Hargreaves CE, Grasso M, Hampe CS, Stenkova A, Atkinson S, Joshua GW, et al. Yersinia enterocolitica provides the link between thyroid-stimulating antibodies and their germline counterparts in Graves' disease. **J Immunol.** 2013;190(11):5373-81.
32. Duvvuri VR, Duvvuri B, Alice C, Wu GE, Gubbay JB, Wu J. Preexisting CD4+ T-cell immunity in human population to avian influenza H7N9 virus: whole proteome-wide immunoinformatics analyses. **PLoS One.** 2014;9(3):e91273.

33. Suárez-Pantaleón C, Mercader JV, Agulló C, Abad-Somovilla A, Abad-Fuentes A. Forchlorfenuron-mimicking haptens: from immunogen design to antibody characterization by hierarchical clustering analysis. **Org Biomol Chem.** 2011;9(13):4863-72.
34. Werner E, Clark JN, Hepburn A, Bhamber RS, Ambler M, Bourdeaux CP, et al. Explainable hierarchical clustering for patient subtyping and risk prediction. **Exp Biol Med (Maywood).** 2023;248(24):2547-2559.