Comparative Bioinformatic Analysis of Human and Swine MHC Sequences: Implications for Xenotransplantation

Michael Olp[1], Matthew Cusick[1], Sam Ho[2,3]

1. Pathology, University of Michigan Medicine, Ann Arbor, MI, United States.
2. Gift of Hope Organ & Tissue Donor Network, Itasca, IL, United States.
3. Michigan State University, East Lansing, MI, United States.

Abstract

Human and swine Major Histocompatibility Complex (MHC) proteins—HLA and SLA, respectively—play pivotal roles in immune recognition and transplant compatibility. We conducted a comprehensive bioinformatic comparison of HLA and SLA protein sequences to evaluate their similarity and assess implications for pig-to-human xenotransplantation. Class I and II MHC sequences from both species were aligned and analyzed for homology, revealing moderate sequence identity (~ 50% − 80%) with conserved overall structure but notable divergent residues. A ProtBert transformer model was fine-tuned to classify sequences by species and locus, clustering most SLA alleles with their closest HLA counterparts. Attention- weight analysis of the model highlighted key polymorphic residues in peptide-binding regions that differ between HLA and SLA. Structural homology modeling confirmed that while the swine MHC proteins adopt the canonical MHC fold, the species-specific residues map to surfaces involved in T-cell recognition. These findings provide insight into which aspects of SLA are human-like versus distinctly "xeno," informing strategies to mitigate immune rejection in xenotransplantation. Our study demonstrates the utility of computational methods in predicting cross-species immunological compatibility, laying groundwork for targeted genetic modifications and further experimental validation.

Introduction

Major Histocompatibility Complex (MHC) molecules are highly polymorphic cell-surface glycoproteins essential for adaptive immunity, mediating antigen presentation to T lymphocytes and essential for both tissue compatibility and immune recognition. In humans, MHC class I (HLA-A, -B, -C) and class II (HLA-DR, -DQ, -DP) proteins present peptides to CD8+ and CD4+ T cells, respectively, and their extraordinary allelic diversity—focused in peptide-binding regions—makes these proteins both vital for immune defense and the primary barriers to transplantation due to potential alloreactivity. Swine leukocyte antigens (SLA), the porcine counterparts, perform analogous functions but have diverged substantially from human HLA since the species' split ~80 million years ago, though notable structural and sequence homology persists. **This similarity poses a paradox for xenotransplantation**, as shared MHC motifs could support functional compatibility but also provoke cross-species immune attack due to pre-existing or elicited antibodies and T cells. While advances in genetic engineering have addressed some rejection mechanisms, such as removing pig-specific glycans, the molecular divergence between SLA and HLA remains a primary source of T cell and antibody-mediated rejection. In this study, we comprehensively compare HLA and SLA using sequence alignments, evolutionary analyses, and advanced machine learning approaches (e.g., ProtBert models), integrating sequence identity, divergence, classifier mapping, and homology modeling to identify molecular determinants of xenogeneic immune responses. Our analysis aims to elucidate the degree of immunological similarity between SLA and HLA, thereby informing strategies to engineer pig donors for improved human graft acceptance.

Methods and Materials

Representative HLA and SLA protein sequences (class I: HLA-A, -B, -C; SLA-1, -2, -3, etc.; class II: HLA-DRB, -DQB, -DPB, HLA-DOβ, and SLA orthologs) were collected from IMGT/HLA and IPD-MHC databases. Sequences were aligned using MAFFT and manually curated. Sequence similarity was quantified via DIAMOND alignments, with identities summarized by locus and visualized by hierarchical clustering. Evolutionary divergence between HLA and SLA was assessed by calculating Jensen–Shannon divergence (JSD) for each alignment position. For functional insights and classification, a ProtBert transformer model was fine-tuned to distinguish MHC loci and then used to map SLA sequences into the HLA landscape, with attention scores pinpointing discriminant residues. Structural homology models were generated with ColabFold/AlphaFold2, using PDB structures as references where available, and PyMOL was used for structural alignment and visualization. High-attention and high-JSD residues were mapped to structures to highlight key immunological differences.

Results

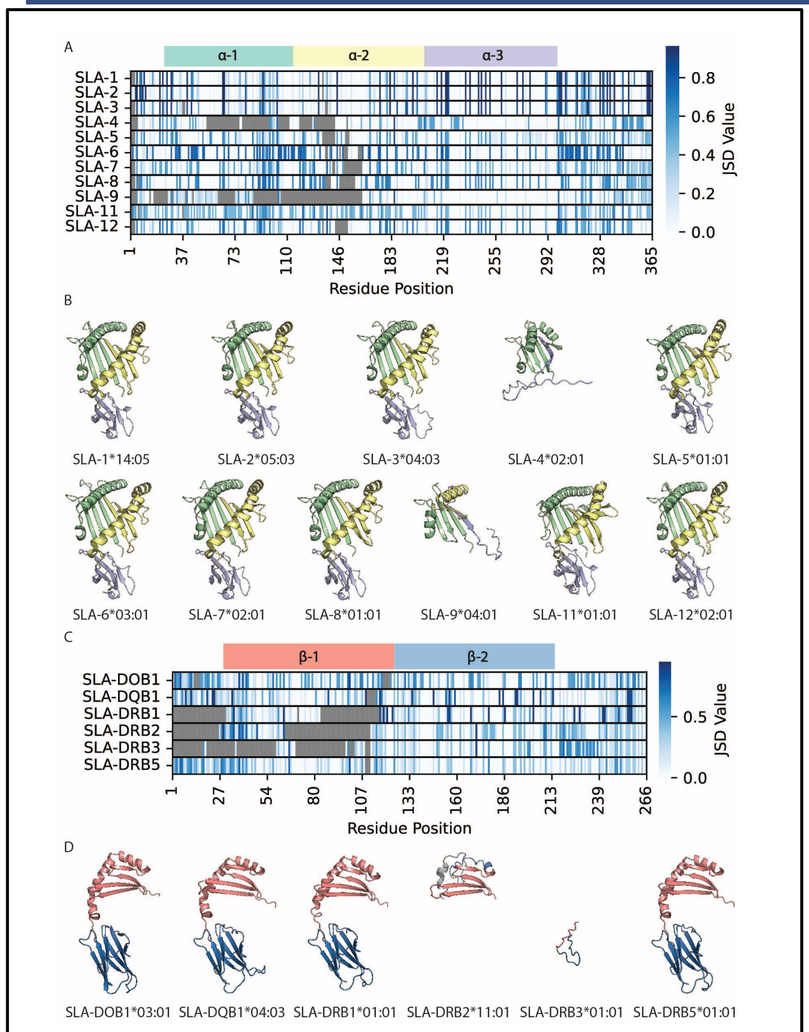

Figure 1. Divergence and structural comparison of HLA vs. SLA. (A) Jensen–Shannon divergence (JSD) heatmap for class I MHC: each row represents a pig SLA class I locus compared to the set of human HLA class I (A, B, C) sequences. Blue intensity reflects the magnitude of divergence at each aligned position (darker = higher JSD), and gray columns indicate positions with gaps in the SLA alignment. Key regions of the class I heavy chain (α1, α2, α3 domains) are annotated along the x-axis. (B) Ribbon structure illustrations of representative SLA class I protein models colored according to domain. (C) JSD heatmap for class II MHC (comparing SLA class II vs. HLA class II sequences in the β1 domain of DR/DQ). (B) Ribbon structure illustrations of representative SLA class II protein sequences colored according to domain.

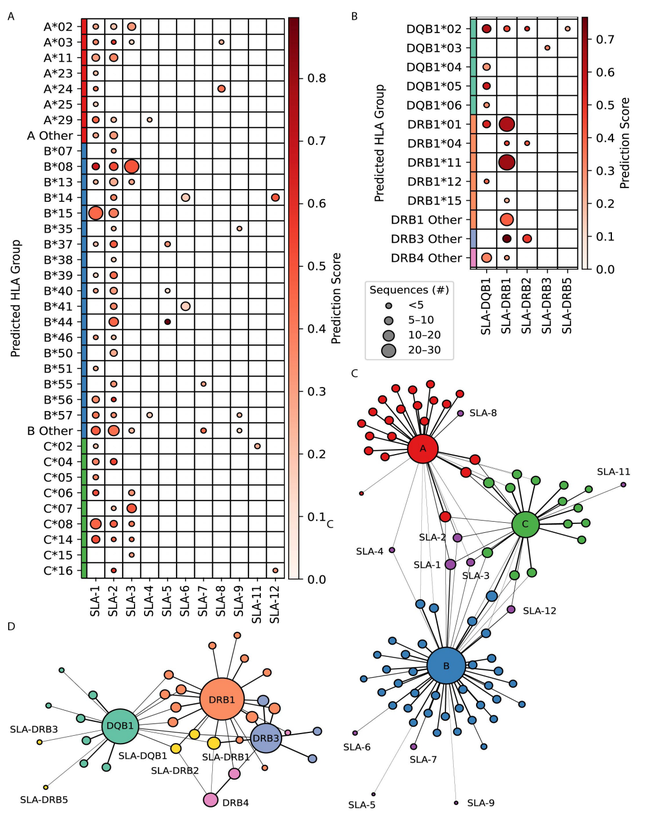

Figure 2. Mapping of SLA sequences onto HLA categories via ProtBert classification. (A) Predicted HLA classifications of SLA class I alleles using the ProtBert model. Each cell represents a predicted classification score, with the size of circles indicating the number of sequences assigned to that classification. (B) SLA class II classification predictions, following the same approach as in (A). (C) Network representation of SLA class I sequences, showing their connectivity to HLA-A, -B, and -C groups based on classification probabilities. The size of each node reflects its relative classification confidence, and edges indicate assignment strength. (D) Network representation of SLA class II sequences, depicting their relationships to HLA-DRB, -DQB, and -DRB3/4/5 groups. Color coding differentiates HLA groups, and node sizes correspond to prediction confidence.

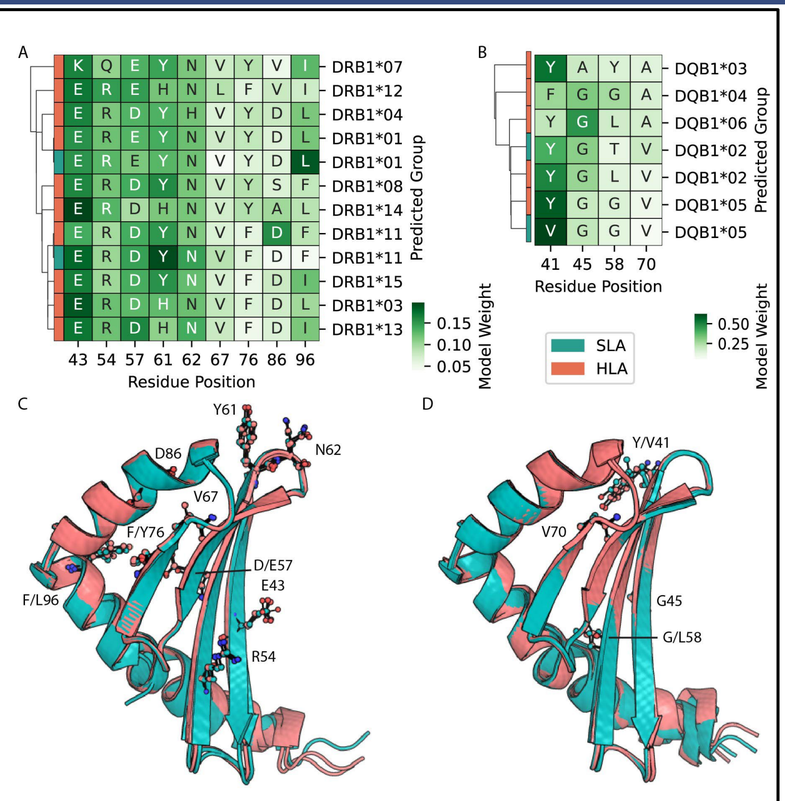

Figure 3. Key residues and structures for class II MHC. (A) Attention clustermap for HLA-DR vs SLA-DRB sequences (similar format to Figure 4A). (B) Attention clustermap for HLA-DQ vs SLA-DQB sequences. Polymorphic positions in the β1 domain (e.g., β-chain positions 26, 30, 37, 57, 67) show high attention differentiation. (C) Structural overlay of a human HLA-DR1 (orange) with pig SLA-DR (teal), highlighting high-attention residue sites as spheres. (D) Overlay of HLA-DQ (orange) with SLA-DQ (teal) with key differing residues marked.

Key Findings

1. Sequence Homology
- SLA and HLA share **~50–80% sequence identity**.
- **Class I:** SLA-1, -2, -3 share ~73–76% identity with HLA-A/B; SLA-6 and SLA-11 are more divergent (~47–60%).
- **Class II:** SLA-DRB1*01:01 shows ~78% identity with HLA-DRB1*15:01; SLA-DQB1 ~77% with HLA-DQB1*06:02.
- **Conservation:** Structural framework preserved, but ~25–30% of residues differ—especially in peptide-binding regions.

2. Divergence
- Conserved regions: α3 domain (class I), β2 domain (class II).
- Divergent regions: **peptide-binding grooves and T-cell receptor contact sites**.
- **Suggests SLA can function structurally as MHC but appears foreign to human immunity.**

3. Machine Learning Mapping
- ProtBert classified SLA into HLA categories with high accuracy.
- SLA-1/2/3 map variably across HLA-A, -B, -C.
- SLA-DRB1 aligns with HLA-DR, SLA-DQB1 with HLA-DQ.
- Indicates **functional parallels but incomplete overlap**.

4. Structural & Residue-Level Insights
- Structural folds nearly identical between SLA and HLA.
- Attention and divergence analyses pinpointed **key polymorphic residues** in peptide-binding grooves (e.g., SLA class I residue 156, β-chain residues 57, 70–74 in class II).
- These are predicted **immunogenic hotspots** likely to trigger human T- and B-cell responses.

Conclusions

Similarity: SLA ≈ 70% identical to HLA, structurally conserved.

Difference: Divergence concentrated in immune-sensitive regions (peptide-binding/TCR contact).

Takeaway: SLA is human-like in architecture but immunologically distinct. Computational mapping provides a roadmap for engineering less immunogenic pigs to improve xenograft survival and potential diagnostic tools to assess the human immune response to xenotransplants.

Contacts

Michael D. Olp, MD, PhD
House Officer III
Department of Pathology
University of Michigan
Email: olpmi@med.umich.edu

Matthew F. Cusick, PhD, F(ACHI)
Associate Professor of Pathology
Director, Histocompatibility Laboratory
Michigan Medicine Pathology
Email: mfcusick@med.umich.edu
(734) 764-0519 | C: (801) 652-3146

Sam Ho, PhD, F(ACHI), FADLM
Director, Histocompatibility & Infectious Disease Testing Laboratory
Gift of Hope Organ & Tissue Donor Network
Email: sho@giftofhope.org
O: (630) 758-2660 | F: (630) 758-2760 | C: (734) 604-6064