**Targeting the SPP1-CD44 Axis in Pediatric High-Grade Glioma through Integrated Single-Cell and Structural Bioinformatics Approaches**

Shiwani Limbu1, Ambuj Kumar2\*

1Quantitative and System Biology Graduate Program, University of California, Merced, USA.

2Center of Childhood Cancer and Blood Diseases, Abigail Wexner Research Institute at Nationwide Children’s Hospital, Columbus, Ohio, USA

\*Corresponding Author: ambuj.kumar@nationwidechildrens.org

**Abstract**

Pediatric high-grade glioma (pHGG) is a highly aggressive brain tumor characterized by transcriptional plasticity and an immunosuppressive microenvironment. Single-cell RNA-seq analysis revealed diverse malignant and immune cell populations, with tumor-associated macrophages (TAMs) emerging as the primary source of SPP1 (osteopontin), a glycoprotein that suppresses T cell activation through CD44 binding. Cell-cell communication analysis identified the SPP1-CD44 axis as a dominant immunosuppressive pathway in the tumor microenvironment. Despite extensive transcription factor screening, no strong regulators of SPP1 were identified, suggesting regulation occurs via alternative mechanisms. To assess structural features of SPP1, replica exchange molecular dynamics simulations were performed, revealing that the CD44-binding domain is conformationally stable. Phosphorylation at Ser169, a conserved site, further stabilized this region, suggesting a potential mechanism for enhanced CD44 interaction. To disrupt this axis, among 2,500 variants of anti-SPP1 23C3 antibody, a lead candidate with improved predicted affinity and minimal sequence divergence was identified. Together, this study integrates transcriptomic and structural bioinformatics approaches to target SPP1-CD44 axis in pHGG.

1. **Introduction**

Pediatric high-grade glioma (pHGG) is an aggressive and highly heterogenous brain tumor, originating from the glial cells in brain [1]. Single-cell transcriptomic studies have shown that malignant pHGG cells recapitulate neural lineage programs, falling into four main cell states: astrocyte-like (AC-like), oligodendrocyte progenitor–like (OPC-like), neural progenitor–like (NPC-like), and mesenchymal-like (MES-like) [1,2]. These states correspond to earlier bulk transcriptional subtypes (classical, proneural, mesenchymal) but can coexist within one tumor [3]. For example, AC-like and MES-like populations often dominate “classical” and “mesenchymal” GBMs respectively, whereas OPC/NPC-like cells are enriched in “proneural” tumors [2]. Spatial transcriptomics confirms that these states localize to distinct niches – e.g. MES-like cells cluster around hypoxic, necrotic regions, while OPC/NPC-like cells populate infiltrative edges [3]. Importantly, GBM cells can transition between states (plasticity), which complicates therapies targeting any single subtype​ [2]. These transitions also create intermediate cell states adding another layer of complexity.

pHGG is characterized by its extensive immunosuppressive microenvironment [4]. Here, tumor-associated macrophages (TAMs) dominate the immune infiltrate​, often outnumbering T cells and other lymphocytes [5]. These TAMs adopts immunosuppressive phenotypes that hinder T cell activation and cytotoxic function in pHGG [6,7]. One key mediator in this crosstalk is Secreted Phosphoprotein 1 (SPP1), also known as osteopontin (OPN), which is abundantly produced by TAMs [8]. ​SPP1 is a glycoprotein that binds the CD44 receptor on T cells, acting as an **immune checkpoint** that downgrades T cell activation [9,10] in both adult and pediatric gliomas. TAMs with high SPP1 expression is linked to impaired T cell responses and poor patient outcomes​ [10]. It has been shown that when T cells are stimulated in culture in the presence of recombinant SPP1, their proliferation and activation are markedly inhibited​ [11]. SPP1 causes a dose-dependent reduction in T cell division [11]​. It also suppresses the production of key cytokines like interferon-gamma (IFN-γ) by activated T cells​ [11]. Notably, the presence of SPP1 leads to fewer T cells expressing early activation markers such as CD69 and IL-2 receptor α (CD25) after stimulation [11]​. Lack of interferon signaling has shown to suppress CAR T cell activity by creating a SPP1 dependent immunosuppressive tumor microenvironment [12]. Adding anti-SPP1 antibody i.e. bioXcell Clone 100D3 and clone MPIIIB10 showed reversal of immunosuppression TME and enhanced CAR T cell antitumor responses [12]. This suggests that SPP1-CD44 interaction is key to modulating immunosuppressive TME, whose blockade could enhance T cell activity and improve responses to checkpoint inhibitors [13].

Beyond immunosuppression, SPP1-CD44 signaling in glioblastoma also promotes tumor aggressiveness by promoting stem cell–like traits and radiation resistance in glioma cells within the perivascular niche [14]​. In addition, elevated SPP1 in the tumor microenvironment has been correlated with more extensive TAM infiltration in cancer [15]​, suggesting SPP1 may also act as a chemoattractant to recruit or retain macrophages. This dual role makes the SPP1-CD44 axis a compelling therapeutic target in pHGG. In this study we aim to combine single cell RNA sequencing data analysis and structural bioinformatics methods to identify approaches to block SPP1 role forming immunosuppressive microenvironment in pHGG.

1. **Materia and Methods**
   1. **Data collection**

Publicly available single-cell RNA-seq data for pediatric high-grade glioma were downloaded from the Single-Cell Pediatric Cancer Atlas Portal (ID: SCPCP000001). Only tumors obtained from patients ≤18 years of age were retained (13,663 glioblastoma cells and 1,061 other pHGG cells). Downstream analysis of raw count data is performed using Seurat v5.0 [16].

* 1. **Single-cell preprocessing and cell type annotation**

Cells with less than 200 expressed genes, or more than 5000 expressed genes, or with percentage of mitochondrial gene expressed more than 20 were filtered out. Top 2,000 variable genes per sample were identified for downstream analysis. Cell-cycle effects were calculated using Seurat CellCycleScoring function which implements scoring strategy described in Tirosh et. Al (2016) [17], and subsequently regressed out during data scaling using Seurat ScaleData function. Sample integration was performed using Harmony, based on the top 50 principal components (PCs). Clustering of integrated cells was then performed using 30 PCs. Differentially expressed genes for each cluster were obtained using Seurat FindAllMarkers function. Canonical marker genes, together with gene set enrichment analysis (GSEA) conducted using gseGO (ClusterProfiler v4.6), was used to assign 11 cell types to all cells: T cells, NK cells, naïve B cells, microglia-derived TAM (MGD TAM), MGD transitioning macrophage (M1/M2), bone-marrow-derived (BMD) macrophage, MES-APC-like, MES-AC-like, MES-AC-like-cycling, OPC-like, and undetermined cells.

* 1. **Copy Number Variation (CNV) Inference**

InferCNV tool [18] was applied to estimate large-scale chromosomal copy number variations (CNVs) from single-cell transcriptomic data to distinguish malignant from non-malignant cell populations based on genomic instability patterns. NK cells, being non-malignant immune cells with stable genomes, were used as the reference population. The expression data was processed using default inferCNV parameters, with cutoff value 0.1. CNV profiles were then visualized to confirm tumor-associated genomic aberrations across identified cell types.

* 1. **Cell-cell communication analysis**

CellChat [19] was used to estimate cell-cell communication and their associated ligand-receptor pairs. Pathway-level communication was inferred by summarizing ligand-receptor interaction probabilities for each signaling pathway using computeCommunProbPathway(). Finally, the overall communication network was aggregated using aggregateNet() by counting the number of interactions and summarizing their communication probabilities.

* 1. **Gene-regulatory network inference**

Transcriptome-wide TF-gene target relationships were reconstructed using pySCENIC [20,21]. Homo sapiens 1390 curated TFs were used to create TF-gene pair adjacency matrix. hg38 Refseq\_r80 SCENIC+ mc\_v10\_clust databases (<https://resources.aertslab.org/cistarget/databases/homo_sapiens/hg38/refseq_r80/mc_v10_clust/gene_based/>) [RRID:SCR\_024808] is then used with pySCENIC ctx tool to compute the regulon enrichment matrix. The pySCENIC aucell tool was then used to generate a loom file containing regulon enrichment scores and Area Under the Curve (AUC) values. TF activity with AUC > 0.5 and Spearman correlation > 0.5 with SPP1 expression in MGD-TAM, MGD-macrophage, or MD-macrophage populations were considered strong candidate regulators of SPP1 in macrophage subsets.

* 1. **Molecular modelling and Solute tempered replica exchange molecular-dynamics (st-REMD) simulation of SPP1 protein**

Amino acid sequence of SPP1 protein was obtained from UniProt (P10451) [22]. Initial 3D protein structure of SPP1 was modelled using I-TASSER-MTD [23]. The full-length human SPP1 protein model was subjected to solute tempered replica exchange molecular dynamics (st-REMD) simulations at 283.15 K, 303.15 K, 333.15 K, and 353.15 K (four replicas, 200 ns each), using GROMACS 2023.2 [24,25] and the CHARMM36 force field [26]. Systems were solvated in a rectangular box with TIP3P water molecules at 10 Å marginal radius and neutralized by added potassium ions (K+) and chloride (Cl-) ions. All forcefield parameters and solvated protein-solvent systems for st-REMD simulation was generated using CHARMM-GUI webserver [27]. Initial energy minimization was performed using the steepest descent algorithm for 5,000 steps or until a convergence criterion of 1000 kJ/mol/nm was reached. To maintain structural integrity during minimization, position restraints were applied with force constants of 400 kJ/mol/nm² on backbone atoms and 40 kJ/mol/nm² on side chains. Electrostatics were treated with Particle-Mesh Ewald (PME) [28]. Van der Waals interactions were computed using a force-switch modifier between 1.0-1.2 nm. Systems were then equilibrated using a 4 fs timestep under NVT ensemble conditions for 10 ns, with position restraints maintained on heavy atoms. Initial velocities were generated from a Maxwell distribution corresponding to the reference temperature. Electrostatics and van der Waals settings were identical to the minimization step, and hydrogen bonds were constrained using LINCS.

Four chains with solute temperature set at 283.15K, 303.15K, 333.15K, and 353.15K respectively were initiated with solvent temperature maintained at 313.15K across all chains, where each chain was allowed to switch every 5 ns. Pressure was maintained at 1 bar using isotropic C-rescale barostat [29]. Electrostatics were treated using Particle Mesh Ewald (PME) method [28]. Conformational stability was evaluated using per-residue RMSF across all chains. An energy landscape heatmap was generated based on root mean square deviation (RMSD) and radius of gyration (Rg) values across all chains. A representative protein conformation was selected from a global energy minimum on the energy landscape for downstream analyses.

* 1. **Phosphorylation site prediction**

NetPhos 3.1 [30] was used to predict phosphorylation residue site on SPP1 protein. Residue with score >0.8 was picked as a likely phosphorylation residue site for downstream analysis.

* 1. **Conventional MD simulation of unmodified and phosphorylated SPP1**

The relaxed SPP1 structure obtained from st-REMD was used as the starting point for conventional molecular dynamics (MD) simulations. Phosphorylation was introduced at residue S169 (pSer169), based on prediction results, and force field parameters for both the unmodified and pSer169 SPP1 structures were generated using the CHARMM-GUI webserver. Energy minimization and equilibration steps were performed under the same conditions as those used in the st-REMD protocol. Production MD simulations were carried out for both, the unmodified and pSer169 SPP1, at 303.15 K using the NPT ensemble. For each form, three independent MD replicates were conducted to ensure reproducibility. Quadratic mean of rmsf values across three chains were calculated for each residue in unmodified and pSer169 SPP1 structure. To identify SPP1 protein domains affected by pSer169.

* 1. **Antibody docking and in-silico affinity maturation**

Complementarity-determining regions (CDRs) of parental anti-SPP1 antibody 23C3 [31] were identified using PyIgClassify2 [32]**.** 23C3 CDR region was docked to the relaxed unmodified SPP1-CD44 binding domain (residue 121-140) [33] with Haddock [34]. 2,500 23C3 antibody variants were generated with RosettaAntibodyDesign tool [35]. For every variant, full heavy and light chains sequence embeddings were obtained using ESM-2 8M model [36]. CDR-specific embeddings (H1-H4, L1-L4) were extracted from their corresponding chain full sequence embedding. Cosine distance between variant CDR sequence embedding and 23C3 CDR sequence embedding was calculated and combined with Rosetta dG\_separated (SPP1-antibody binding free energy) to pick top antibody hits. Variants were visualized with UMAP generated using cosine distance obtained from embeddings. Variant with minimal dG\_separated and minimal embedding distance from 23C3 was selected as the top hit variant. Mutations were mapped onto the protein complex using PyMOL [37].

1. **Results and discussion**
   1. **Single‐cell landscape of pediatric HGG reveals a dominant SPP1‐expressing TAM population**

To define the cellular composition of the pediatric HGG microenvironment, unsupervised clustering and UMAP projection was applied on single cell RNA-seq data, leading to identified ten major cell populations (Fig. 1a), including malignant glial‐lineage states (MES-like, MES-AC-like, MES-AC-like Cycling, MES-APC-like, OPC-like), lymphoid cells (T cells, NK cells, Naïve B), and myeloid cells segregating into two macrophage subsets (MGD Macrophage, MGD TAM) plus a Microglia‐derived TAM cluster (MD Macrophage). A small “Undetermined” cluster did not express known canonical markers and may likely represents rare stromal elements.

* 1. **Genomic validation of malignant versus immune clusters by inferCNV**

InferCNV to detect large-scale chromosomal copy‐number alterations at single‐cell resolution, using T cells, NK cells, and Naïve B cells as a diploid reference. The resulting CNV heatmap (Fig. 1b) revealed that all malignant glial clusters display characteristic aberration-such as chromosome 7 gain [38], as well as chromosome 1, 6, and 13 loss [39,40], while immune clusters show flat profiles consistent with a normal karyotype.

A screenshot of a computer screen

AI-generated content may be incorrect.

***Fig 1.*** *Single cell RNA-seq data analysis results. a) UMAP showing all 11 cell types identified using canonical maker gene expression profile, b) InferCNV results showing gain (red) and loss (loss) of copy number variants across all tumor cells. c) Canonical marker dotplot of commonly observed cell types in pHGG, d) GSEA results showing activated pathways across all cell types.*

* 1. **Gene set enrichment analysis highlights functionally distinct pathway programs in immune and tumor cell types**

Single cell clusters were annotated with cell type label using their corresponding canonical gene expression markers (Fig 1c). Gene set enrichment analysis (GSEA) results show strong enrichment for the T cell receptor complex, MHC class II protein complex, and immunological synapse in T cells (Fig. 1d), consistent with an antigen-experienced, though likely restrained, T cell state within the glioma TME. Microglia-derived TAMs (MGD TAMs) showed enrichment of ribosome pathways (Fig. 1d), indicating high protein synthesis activity, which is often seen in metabolically active cells. This suggests that the MGD TAMs are actively producing proteins, possibly for cytokine secretion, antigen presentation, or immune modulation. In addition, enrichment for primary lysosome and lytic vacuole pathways suggests active engagement in phagocytosis, autophagy, or endocytosis (Fig. 1d). Furthermore, co-enrichment of MHC class II protein complex indicates that MGD TAMs are likely communicating with CD4⁺ T cells within the TME.

Homeostatic or transitioning MGD macrophages displayed an active immune modulation enrichment profile. Specifically, pathways such as tertiary granule membrane, azurophil granule membrane, and ficolin-1 rich granule membrane suggest these macrophages are trafficking immune granules and may be skewed toward an M2-like immunosuppressive phenotype. Simultaneously, the upregulation of Golgi stack, Golgi cisterna membrane, and Golgi apparatus subcompartment points to heightened protein processing activity [41], indicating functional transition in the macrophage sub population towards M2. Moreover, ficolin-enriched granule pathways imply a potential role in complement activation via the lectin pathway [42], which has dual effects in cancer, where it can either promote immune evasion through anaphylatoxins [43] or mediate tumor cell lysis [44].

MES-APC-like cells displayed a strong oxidative and translational metabolic program, suggestive of aggressive, invasive behavior (Fig 1d). Enrichment of mitochondrial inner membrane, NADH dehydrogenase complex, respiratory chain complex I, cytochrome complex, and proton-transporting ATP synthase complex indicates high oxidative phosphorylation (OXPHOS) activity [45], supporting the energy demands of invasion and proliferation. Additionally, pathways such as ribosome, mitochondrial ribosome, ribonucleoprotein complex, and rough endoplasmic reticulum indicate high translation activity. Furthermore, enrichment of spliceosomal complexes-including U2- and U12-type spliceosome, pre-catalytic spliceosome, and tri-snRNP complex-suggests that alternative splicing (AS) is actively being used in transitionary MES phenotype [46]. Ben Mrid et al., (2025) recently showed that AS is extensively rewired across glioma subtypes and associates with MES transition [47]. This raises important questions about whether splicing modulation may serve as a regulatory mechanism for glioma cell plasticity and a potential therapeutic target.

MES-AC-like cells represent a particularly intriguing transitioning phenotype. Despite lacking canonical markers of neurons or their progenitors, this cluster displays co-expression of MES-like and AC-like gene signatures while simultaneously activating multiple neuronal pathways (Fig 1d). This suggests a state of neuronal mimicry, where tumor cells may partially acquire neuron-like features, potentially facilitating immune evasion or integration into the neural niche. Another group of MES-AC like cells shows high expression of cycling cell markers (Fig 1c) as well as signatures of active mitosis (Fig 1d). Enrichment of condensed chromosome, mitotic spindle, spindle midzone, cleavage furrow, and centriole point to cells in active cell division, most likely during S/G2/M phases of the cycle. Supporting this, activation of the kinesin complex, microtubule cytoskeleton, and microtubule organizing center indicates that these cells are undergoing chromosome segregation and cytokinesis. Furthermore, enrichment of nuclear replication fork, heterochromatin, and DNA repair complex pathways underscores ongoing DNA replication and repair, likely reflecting the heightened stress and genomic instability associated with rapid tumor cell proliferation. Collectively, gene enrichment signatures across all cancer cell type indicates that TME is actively under transition state promoting immunosuppressive phenotypes.

* 1. **Cell communication analysis pinpoints SPP1-CD44 as a key immunosuppressive axis**

To interrogate intercellular communication, we applied a Cellchat to all clusters (Fig. 2a-d). Among outgoing signals, SPP1 ranked within the top ligands secreted by MGD TAMs (Fig. 2a), while its receptor CD44 was highly expressed on T cells (Fig. 1c). In addition, Fig 2b shows that SPP1 is the top incoming signal in T cells. Dot‐plot analysis of communication probabilities (Fig. 2c) confirmed that the MGD TAM - T cell axis via SPP1-CD44 is the strongest single ligand–receptor interaction in the microenvironment, exceeding even canonical cytokine pathways such as CCL3-CCR1 or IL1B-IL1R1. Network topology mapping (Fig. 2d) revealed that MGD TAMs and MGD transitioning macrophages are the dominant senders of SPP1 signals, with negligible autocrine SPP1-CD44 loops in tumor cells. Violin plots of SPP1 transcript abundance (Fig. 2e) show that MGD TAMs and MGD transitioning macrophage cells express SPP1 at levels 3-10-fold higher than other immune and cancer cells. Together, these data demonstrate that a MGD TAM and MGD transitioning macrophage cells are principal source of SPP1 in pHGG, highlighting SPP1 as a potential therapeutic target.

A screenshot of a computer screen

AI-generated content may be incorrect.

***Fig 2.*** *Cellchat cell-cell communication analysis results. a) Outgoing signal heatmap. The intensity of color represents higher communication probability. b) incoming signal heatmap. The intensity of color represents higher communication probability. c) All outgoing signaling network from MGD TAMs. e) Network topology heatmap of SPP1. f) Violin plots of SPP1 transcript abundance across all clusters.*

* 1. **Transcription factors regulating SPP1**

pySCENIC was applied to identify transcription factors (TFs) that might drive SPP1 expression in tumor‐associated macrophages (TAMs). Among six high‐confidence TFs (THRB, TEAD1, RFX3, NFIB, MAFB, KLF12) predicted by pySCENIC, none exhibited strong regulon activity (AUC < 0.5), and correlation with SPP1 was uniformly weak (|r| < 0.2) in both MGD TAMs and MGD transitioning Macrophages (Fig 3a,b). MAFB stands out as highly expressed in MGD Macrophage (> 60% of cells) with high average expression, and moderately expressed in MGD TAM (Fig 3b,c,d), yet low regulon activity indicates that none of the predicted TF’s act as a dominant transcriptional driver of SPP1 in TAMs, limiting the feasibility of directly targeting these TFs to suppress SPP1 production. Therefore, alternative strategies, such as intervention at the level of upstream signaling pathways or epigenetic modulators could prove more effective.

A collage of graphs and charts

AI-generated content may be incorrect.

***Fig 3****. pySCENIC results highlight that no one TF dominates SPP1 gene expression regulation in MGD TAM. a) TF Regulon activity profile of SPP1 transcription factors, here size of the dot shows regulon activity and color shows gene expression correlation of the corresponding TF with SPP1 gene expression. b) TF gene expression dot plot. The size of the dot represents number of cells expressing corresponding TF in each cell type population, and the color represents average expression of the gene within the cell type population, c) MAFB gene expression feature plot showing expression of this gene across all cell types. d) MAFB violin plot showing expression of this gene across all cell types.*

* 1. **Candidate epitope regions within disordered SPP1**

SPP1 is a highly disordered protein, which enables it to undergo conformational transitional associated with its diverse range of interactions in TME. Despite its intrinsic disorder, SPP1 exhibits segments of relative structural stability that has shown to serve as viable antibody epitopes [31]. Recent work showed that using an anti-SPP1 antibody (bioXcell Clone: 100D3) and (clone MPIIIB10) in combination with mIL13Rα2 CAR T cell therapy shows enhanced CAR T cell antitumor response [12]. Antibody 2K1, or its chimeric form C2K1, recognizes the equivalent epitope region of human SPP1 (162SVVYGLR168) [48]. Another anti-SPP1 antibody, 23C3, binds to residue segment 42-48 [31]. Therefore, antibody-dependent targeting of the CD44-binding domain of SPP1 holds therapeutic promise.

Due to the disordered nature of SPP1 protein, we first applied replica‐exchange MD simulation at four temperatures (283 K, 303 K, 333 K and 353 K) to identify relative thermodynamic stability of different regions of the protein. Residue-wise RMSF profiles revealed extensive fluctuations across the protein, reflecting temperature-dependent modulation of its dynamic regions (Fig. 4a). Antibody binding regions 2K1, C2K1 (residue 162-168), and 23C3 (residue 43-48), consistently showed significantly lower fluctuation across replicas (Fig 4a) compared to the highly dynamic remainder of the protein. Importantly, CD44‐binding domain (residue 121-140) displayed comparable stability to these known candidate epitope loops, suggesting it may be conformationally accessible for antibody binding. To capture a representative SPP1 conformation, all MD frames from all replicas were projected onto a two‐dimensional free‐energy surface defined by RMSD and radius of gyration (Rg) (Fig. 4b). The global free‐energy minimum corresponded to a compact ensemble centered at RMSD ≈ 1.9 nm and Rg ≈ 4.09 nm. The centroid structure was extracted from this basin (Fig. 4c) for downstream analyses.

A screenshot of a computer screen

AI-generated content may be incorrect.

***Fig 4.*** *SPP1 protein structure and sequence analysis results. a) Root mean squared fluctuation of SPP1 protein residues under replica exchange molecular dynamics simulation starting at temperatures 283.15K (black), 303.15K (red), 333.15K (green), and 353.15K (blue). Residue range 42-48 is 23C3 antibody binding motif of SPP1, residue range 162-168 is 2K1 and C2K1 binding motif of SPP1, and residue range 121-140 is CD44 binding motif of SPP1, b) Energy landscape-based conformation sampling from four replicate exchange runs. Here y-axis represents radius of gyration values across all 4 chains and x axis represents RMSD values across all 4 chains. c) Most stable SPP1 conformation across all 4 replica exchange trajectories. CD44 binding motif (residue 121-140) is shown in orange, 23C3 binding motif is shown in blue, and 2K1 and C2K1 binding motif is shown in cyan. and known SPP1 antibody binding interface shown in blue, and cyan, d) SPP1 protein residue phosphorylation score heatmap. Arrow highlights the top computationally predicted likely phosphorylation residue position 169 with score > 0.8, e) Sequence alignment of SPP1 across mammals. Orange bar highlights SPP1-CD44 binding motif, blue bar highlights 23C3 binding motif, cyan bar highlight 2K1 and C2K1 binding motif, and red arrow represent phosphorylation residue position 169 on SPP1 protein sequence. f) Root mean squared fluctuation quadratic mean of normal SPP1 (black) and residue 169 phosphorylated SPP1 (red), showing increase in stability in SPP1-CD44 binding interface upon phosphorylation.*

* 1. **Phosphorylation at Ser169 may allosterically stabilize the CD44‐binding motif**

Given the disorder nature of SPP1 protein, post‐translational modifications could induce local folding and create stable binding surfaces. NetPhos 3.1 predicted a high‐confidence phosphorylation site at Ser169 (PKC motif; score > 0.8) (Fig. 4d). Multi‐species alignment (Fig. 4e) confirmed that Ser169, as well as the known epitope regions 43–48, 162–168, and the subset (residue 123–131) of CD44‐binding domain, are highly conserved, indicating functional significance of their corresponding residues.

To evaluate the impact of residue 169 phosphorylation, three independent 200‐ns MD replicates each of unmodified and Ser169‐phosphorylated SPP1 were conducted. Quadratic mean of RMSF values of all three chains for each residue position was calculated to explore difference between RMSF patterns in unmodified and Ser169‐phosphorylated SPP1. Result indicates a marked reduction in fluctuation specifically in residues 125–135, a region that overlaps with the core CD44‐binding domain (Fig. 4f). This allosteric dampening suggests that Ser169 phosphorylation may play a role in mediating SPP1-CD44 interaction, and thereby it can also promote a more ordered, antibody‐accessible conformation of the CD44‐binding loop. In near future, we aim to extend MD simulations of unmodified and Ser169‐phosphorylated SPP1 into microsecond time scale to validate the impact of residue 169 phosphorylation on SPP1-CDD4 binding interface.

* 1. **Antibody‐optimization landscape and selection of a lead anti-SPP1 variant**

Starting from the parental 23C3 monoclonal antibody, in-silico library of 2,500 CDR-mutated variants were assessed for i) predicted binding free energy to unmodified SPP1 (dG\_separated) and ii) amino acid sequence/biophysical similarity to 23C3 within the eight CDRs. Heavy- and light-chain ESM-2 embeddings, restricted to the H1-H4 and L1-L4 indices, were projected with UMAP (Fig. 5a). The resulting manifold shows two distinct sequence divergence path with wide range of binding energy values. The color scale encodes for corresponding dG\_separated values. Few variants with high relative SPP1 binding stability are shown as red dots. To integrate the two optimization criteria selected earlier, we plotted dG\_separated against the cosine-esm2 embedding distance from parental 23C3 (Fig. 5b). The global minimum-representing the most stable binder with minimal sequence embedding drift-lies near the origin of the density cloud (black arrow) (Fig. 5b). This variant combines the lowest computed binding energy with one of the smallest embedding distances (< 0.26), indicating relatively modest CDR remodeling was sufficient to achieve the desired affinity gain. Structural analysis of this variant (Fig. 5c) reveals seven mutations in the heavy chain (S25D, T28D, N30K, I31R, N35M, T59L, T60V; red spheres) and ten in the light chain (R24A, A25C, E27D, N28D, I29V, Y30W, S31K, L33F, Q70T, Q89V; blue spheres). Because this variant satisfies both high stability and least sequence divergence, it is prioritized as the lead candidate.

A diagram of a blue and green color

AI-generated content may be incorrect.

***Fig 5****. Antibody directed evolution results obtained from RosettaAntibodyDesign tool. a) UMAP showing 2500 new antibodies shown as colored dots and the parental 23C3 antibody shown as blast star. Color of the dots highlight energy required to break interaction between SPP1 and the corresponding antibody variant. UMAP is calculated using cosine distance between amino acid sequence esm2 embeddings of CDR regions of two antibody variants. b) Hex plot showing 2500 antibody variants. Here energy required to break SPP1 binding with the corresponding antibody variant is shown as x axis, and cosine distance of amino acid sequence esm2 embeddings of CDR regions of an antibody variant with parental 23C3 amino acid sequence esm2 embeddings of CDR regions. c) Most stable antibody protein structure (grey) bonded to SPP1 protein structure (wheat). Blue spheres are the residue changes in this top hit antibody as compared to 23C3 amino acid sequence.*

Despite the encouraging in-silico affinity profile of our 23C3 variant, several caveats remain. First, all predictions were derived from static RosettaAntibodyDesign scoring and short-timescale MD snapshots, which do not account for motions associated with conformational changes that may contribute to antibody-SPP1 binding. Our 200-ns trajectories are likely insufficient to accurately capture slow, allosteric rearrangements in highly disordered SPP1 protein. To address this, we will extend simulations to 10 micro-second timescale to confirm that PKC-mediated phosphorylation of Ser169 indeed stabilizes the CD44-binding domain and quantify how this post-translational modification modulates the free-energy landscape of the antibody–antigen complex. Parallel, long-timescale simulations of both unmodified and Ser169-phosphorylated SPP1 docked to the antibody will allow direct estimation of binding affinities and conformational entropies under physiologically relevant fluctuations. Furthermore, prior to experimental exploration, we aim to screen surface residues of the new variant against comprehensive publicly available immunogenicity libraries to eliminate potential class I or class II T-cell epitopes that could trigger immune responses. Collectively, these follow-up studies will provide the necessary immunogenicity, kinetic, and thermodynamic validation to advance our computationally optimized anti-SPP1 antibody toward experimental tests.

**Code Availability:**All code associated with this work is available on Github (<https://github.com/Sashoss/Pediatric_Immuno-Oncology>)

**References**

1. …, Koptyra MP, Vanguri RS, McGrory S, Resnick AC, ... A longitudinal single-cell and spatial multiomic atlas of pediatric high-grade glioma. bioRxiv. 2024;03: 583588. doi:10.1101/2024.03.06.583588

2. De Silva MI, Stringer BW, Bardy C. Neuronal and tumourigenic boundaries of glioblastoma plasticity. Trends in Cancer. 2023;9: 223–236. doi:10.1016/j.trecan.2022.10.010

3. Eberhart CG, Bar EE. Spatial enrichment of cellular states in glioblastoma. Acta Neuropathol. 2020;140: 85–87. doi:10.1007/s00401-020-02165-3

4. Ross JL, Velazquez Vega J, Plant A, MacDonald TJ, Becher OJ, Hambardzumyan D. Tumour immune landscape of paediatric high-grade gliomas. Brain. 2021;144: 2594–2609. doi:10.1093/brain/awab155

5. Xuan W, Lesniak MS, James CD, Heimberger AB, Chen P. Context-Dependent Glioblastoma–Macrophage/Microglia Symbiosis and Associated Mechanisms. Trends Immunol. 2021;42: 280–292. doi:10.1016/j.it.2021.02.004

6. Haydar D, Ibañez-Vega J, Krenciute G. T-Cell Immunotherapy for Pediatric High-Grade Gliomas: New Insights to Overcoming Therapeutic Challenges. Front Oncol. 2021;11. doi:10.3389/fonc.2021.718030

7. Frederico SC, Sharma N, Darling C, Taori S, Dubinsky AC, Zhang X, et al. Myeloid cells as potential targets for immunotherapy in pediatric gliomas. Front Pediatr. 2024;12. doi:10.3389/fped.2024.1346493

8. Wei J, Marisetty A, Schrand B, Gabrusiewicz K, Hashimoto Y, Ott M, et al. Osteopontin mediates glioblastoma-associated macrophage infiltration and is a potential therapeutic target. J Clin Invest. 2019;129. doi:10.1172/JCI121266

9. van den Broek, Thijs J.M.; Hoogendijk, Raoull; Kranendonk, Mariëtte E.G.; Lammers, Julie A.S.; Krishnamoorthy, Akshaya L.; van Ineveld, Ravian L.; Molleson, Milo; Tsvetkov, Vasily O.; Ringnalda, Femke C.A.; van de Wetering, Marc; Su, Yan; Bianco, John I.; A. Single-cell spatial analysis of pediatric high-grade glioma reveals a novel population of SPP1+/GPNMB+ myeloid cells with immunosuppressive and tumor-promoting capabilities. bioRxiv Prepr Serv Biol. 2025;03: 643953. doi:https://doi.org/10.1101/2025.03.18.643953

10. Tang W, Lo CWS, Ma W, Chu ATW, Tong AHY, Chung BHY. Revealing the role of SPP1+ macrophages in glioma prognosis and therapeutic targeting by investigating tumor-associated macrophage landscape in grade 2 and 3 gliomas. Cell Biosci. 2024;14. doi:10.1186/s13578-024-01218-4

11. Klement JD, Paschall A V., Redd PS, Ibrahim ML, Lu C, Yang D, et al. An osteopontin/CD44 immune checkpoint controls CD8+ T cell activation and tumor immune evasion. J Clin Invest. 2018;128: 5549–5560. doi:10.1172/JCI123360

12. Gholamin S, Natri HM, Zhao Y, Xu S, Aftabizadeh M, Comin-Anduix B, et al. Overcoming myeloid-driven resistance to CAR T therapy by targeting SPP1. bioRxiv Prepr Serv Biol. 2025. doi:10.1101/2025.04.01.646202

13. Shurin MR. Osteopontin controls immunosuppression in the tumor microenvironment. J Clin Invest. 2018;128: 5209–5212. doi:10.1172/JCI124918

14. Pietras A, Katz AM, Ekström EJ, Wee B, Halliday JJ, Pitter KL, et al. Osteopontin-CD44 signaling in the glioma perivascular niche enhances cancer stem cell phenotypes and promotes aggressive tumor growth. Cell Stem Cell. 2014;14: 357–369. doi:10.1016/j.stem.2014.01.005

15. Huang Z, Li Y, Liu Q, Chen X, Lin W, Wu W, et al. SPP1-mediated M2 macrophage polarization shapes the tumor microenvironment and enhances prognosis and immunotherapy guidance in nasopharyngeal carcinoma. Int Immunopharmacol. 2025;147. doi:10.1016/j.intimp.2024.113944

16. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42: 293–304. doi:10.1038/s41587-023-01767-y

17. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (80- ). 2016;352: 189–196. doi:10.1126/science.aad0501

18. Tickle T and TI and GC and BM and HB. inferCNV of the Trinity CTAT Project. Klarman Cell Obs Broad Inst MIT Harvard, Cambridge, MA, USA. 2019.

19. Jin S, Plikus M V., Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. bioRxiv. 2023; 2023.11.05.565674. Available: https://www.biorxiv.org/content/10.1101/2023.11.05.565674v1%0Ahttps://www.biorxiv.org/content/10.1101/2023.11.05.565674v1.abstract

20. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15: 2247–2276. doi:10.1038/s41596-020-0336-2

21. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: Single-cell regulatory network inference and clustering. Nat Methods. 2017;14: 1083–1086. doi:10.1038/nmeth.4463

22. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Res. 2025;55: D609–D617. doi:10.1093/nar/gkae1010

23. Zhou X, Zheng W, Li Y, Pearce R, Zhang C, Bell EW, et al. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. Nat Protoc. 2022;17: 2326–2353. doi:10.1038/s41596-022-00728-0

24. Bekker H, Berendsen H, Dijkstra E, Achterop S, Van Drunen R, Van der Spoel D, et al. Gromacs: A parallel computer for molecular dynamics simulations. Phys Comput. 1993;92: 252–256.

25. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008. doi:10.1021/ct700301q

26. Huang J, Mackerell AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. J Comput Chem. 2013;34: 2135–2145. doi:10.1002/jcc.23354

27. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem. 2008;29: 1859–1865. doi:10.1002/jcc.20945

28. Cheatham TE, Miller JL, Fox T, Darden TA, Kollman PA. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. Journal of the American Chemical Society. 1995. doi:10.1021/ja00119a045

29. Bernetti M, Bussi G. Pressure control using stochastic cell rescaling. J Chem Phys. 2020;153. doi:10.1063/5.0020514

30. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol. 1999. doi:10.1006/jmbi.1999.3310

31. Du J, Hou S, Zhong C, Lai Z, Yang H, Dai J, et al. Molecular Basis of Recognition of Human Osteopontin by 23C3, a Potential Therapeutic Antibody for Treatment of Rheumatoid Arthritis. J Mol Biol. 2008;382: 835–842. doi:10.1016/j.jmb.2008.07.075

32. Kelow, Simon; Faezov, Bulat; Xu, Qifang; Parker, Mitchell; Adolf-Bryfogle, Jared; Dunbrack Jr. RL. A penultimate classification of canonical antibody CDR conformations. bioRxiv Prepr Serv Biol. 2022. doi:10.1101/2022.10.12.511988

33. Lin YH, Yang-Yen HF. The Osteopontin-CD44 Survival Signal Involves Activation of the Phosphatidylinositol 3-Kinase/Akt Signaling Pathway. J Biol Chem. 2001;276: 46024–46030. doi:10.1074/jbc.M105132200

34. Cyril Dominguez, Rolf Boelens, Alexandre M. J. J. Bonvin. HADDOCK:  A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc. 2003;125: 1731–1737.

35. Adolf-Bryfogle J, Kalyuzhniy O, Kubitz M, Weitzner BD, Hu X, Adachi Y, et al. RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. PLoS Comput Biol. 2018;14. doi:10.1371/journal.pcbi.1006112

36. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science (80- ). 2023;379: 1123–1130. doi:10.1126/science.ade2574

37. Schrödinger L. The PyMOL Molecular Graphics. Version 252. Available: https://pymol.org/2/support.html%0Ahttps://scholar.google.com/scholar?hl=en&as\_sdt=0%2C5&q=The+PyMOL+Molecular+Graphics+System%2C+Version+1.74.4+Schrodinger%2C+LLC.+https%3A%2F%2Fpymol.org%2F+%3B+Accessed+10+February+2020.&btnG=%0Ahttps://pymol.org/2/supp

38. Goodarzi AA, Noon AT, Deckbar D, Ziv Y, Shiloh Y, Löbrich M, et al. Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. J Cell Biol. 2014;5: 22. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2000477&tool=pmcentrez&rendertype=abstract

39. Giunti L, Pantaleo M, Sardi I, Provenzano A, Magi A, Cardellicchio S, et al. Genome-wide copy number analysis in pediatric glioblastoma multiforme. Am J Cancer Res. 2014;4: 293–303.

40. Qu HQ, Jacob K, Fatet S, Ge B, Barnett D, Delattre O, et al. Genome-wide profiling using single-nucleotide polymorphism arrays identifies novel chromosomal imbalances in pediatric glioblastomas. Neuro Oncol. 2010;12: 153–163. doi:10.1093/neuonc/nop001

41. Day KJ, Staehelin LA, Glick BS. A three-stage model of Golgi structure and function. Histochem Cell Biol. 2013;140: 239–249. doi:10.1007/s00418-013-1128-3

42. Matsushita M, Fujita T. Ficolins and the lectin complement pathway. Immunol Rev. 2001;180: 78–85. doi:10.1034/j.1600-065X.2001.1800107.x

43. Sayegh ET, Bloch O, Parsa AT. Complement anaphylatoxins as immune regulators in cancer. Cancer Med. 2014;3: 747–758. doi:10.1002/cam4.241

44. Lu P, Ma Y, Wei S, Liang X. The dual role of complement in cancers, from destroying tumors to promoting tumor development. Cytokine. 2021;143. doi:10.1016/j.cyto.2021.155522

45. Sharma L, Lu J, Bai Y. Mitochondrial Respiratory Complex I: Structure, Function and Implication in Human Diseases. Curr Med Chem. 2009;16: 1266–1277. doi:10.2174/092986709787846578

46. Turunen JJ, Niemelä EH, Verma B, Frilander MJ. The significant other: Splicing by the minor spliceosome. Wiley Interdiscip Rev RNA. 2013;4: 61–76. doi:10.1002/wrna.1141

47. Ben Mrid R, El Guendouzi S, Mineo M, El Fatimy R. The emerging roles of aberrant alternative splicing in glioma. Cell Death Discov. 2025;11. doi:10.1038/s41420-025-02323-0

48. Yamamoto N, Nakashima T, Torikai M, Naruse T, Morimoto J, Kon S, et al. Successful treatment of collagen-induced arthritis in non-human primates by chimeric anti-osteopontin antibody. Int Immunopharmacol. 2007;7: 1460–1470. doi:10.1016/j.intimp.2007.06.009