**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. 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) ​[cell.com](https://www.cell.com/trends/cancer/fulltext/S2405-8033(22)00233-3#:~:text=expand%2C%20they%20gain%20stochastic%20mutations,of%20equal%20proportions%20of%20two)​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC10942465/#:~:text=Neftel%20et%20al,like%20state%20which%20expressed). These states correspond to earlier bulk transcriptional subtypes (classical, proneural, mesenchymal) but can coexist within one tumor​[cell.com](https://www.cell.com/trends/cancer/fulltext/S2405-8033(22)00233-3#:~:text=varying%20levels%20of%20fitness,to%20proliferative%20and%20multipotent%20to)​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC7362280/#:~:text=genetic%20and%20epigenetic%20heterogeneity%2C%20and,in%20the%20IvyGAP%20GBM%20database). 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​[cell.com](https://www.cell.com/trends/cancer/fulltext/S2405-8033(22)00233-3#:~:text=expand%2C%20they%20gain%20stochastic%20mutations,of%20equal%20proportions%20of%20two)​[cell.com](https://www.cell.com/trends/cancer/fulltext/S2405-8033(22)00233-3#:~:text=However%2C%20individual%20GBM%20tumours%20are,A). 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​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC7362280/#:~:text=%28https%3A%2F%2Fglioblastoma,specific%20anatomical%20locations%20within%20GBM)​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC7362280/#:~:text=regions). Importantly, GBM cells can transition between states (plasticity), which complicates therapies targeting any single subtype​[cell.com](https://www.cell.com/trends/cancer/fulltext/S2405-8033(22)00233-3#:~:text=Such%20cellular%20plasticity%20is%20particularly,rise%20to%20aggressive%20tumour%20recurrence)​[cell.com](https://www.cell.com/trends/cancer/fulltext/S2405-8033(22)00233-3#:~:text=phenotypes%20and%20varying%20fitness%20to,complete%20loss%20of%20the%20aberration). These transitions also creates intermediate cell states adding another layer of complexity.

pHGG is characterized by its extensive immunosuppressive microenvironment. Here, tumor-associated macrophages (TAMs) dominate the immune infiltrate​, often outnumbering T cells and other lymphocytes. These TAMs adopts immunosuppressive phenotypes that hinder T cell activation and cytotoxic function in pHGG. One key mediator in this crosstalk is Secreted Phosphoprotein 1 (SPP1), also known as osteopontin (OPN), which is abundantly produced by TAMs in pHGG. ​Osteopontin is a glycoprotein that binds the CD44 receptor on T cells, acting as an **“immune checkpoint”** that blunts T cell activation (​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC10956315/#:~:text=To%20elucidate%20the%20mechanisms%20of,Additionally)​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC10956315/#:~:text=,PMC%20free%20article)) in both adult and pediatric gliomas. TAMs with high SPP1 expression is linked to impaired T cell responses and poor patient outcomes​ ([pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC10956315/#:~:text=SPP1%20demonstrated%20a%20significant%20association,dataset%2C%20two%20distinct%20molecular%20subtypes)​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC10956315/#:~:text=consist%20of%20brain,2h)). The SPP1–CD44 interaction on T cells inhibits their proliferation and promotes early exhaustion, allowing the tumor to evade immune attack (​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC10956315/#:~:text=To%20elucidate%20the%20mechanisms%20of,Additionally)​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC6264653/#:~:text=authors%20demonstrate%20that%20downregulation%20of,to%20immune%20checkpoint%20inhibitor%20therapy)). It has been shown that when CD3<sup>+ </sup>T cells are stimulated in culture (with anti-CD3/anti-CD28) in the presence of recombinant osteopontin, their proliferation and activation are markedly inhibited​. Osteopontin causes a dose-dependent reduction in T cell division (as shown by CFSE dilution assays)​. It also suppresses the production of key cytokines like interferon-gamma (IFN-γ) by activated T cells​. Notably, the presence of osteopontin leads to fewer T cells expressing early activation markers such as CD69 and IL-2 receptor α (CD25) after stimulation​. This suggests that OPN–CD44 engagement raises the threshold for T cell activation, keeping T cells in a less responsive state. In these assays, osteopontin even reduced the induction of PD-1 on T cells​, indicating that the T cells were not progressing to a fully activated (or exhausted) phenotype but rather being held in check at an early stage of activation. Notably, disrupting this axis is an emerging strategy in immunotherapy - **osteopontin is now recognized as an immune-suppressive pathway** whose blockade could reinvigorate T cells and improve responses to checkpoint inhibitors (​[pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC6264653/#:~:text=authors%20demonstrate%20that%20downregulation%20of,to%20immune%20checkpoint%20inhibitor%20therapy)).

Beyond immunosuppression, osteopontin/CD44 signaling in glioblastoma also promotes tumor aggressiveness: it maintains stem cell–like traits and radiation resistance in glioma cells within the perivascular niche​ ([pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC3999042/#:~:text=radiation%20resistance,like)). In addition, elevated SPP1 in the tumor microenvironment has been correlated with more extensive TAM infiltration in both adult and pediatric gliomas​, 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 glioblastoma. 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). Raw count were processed in Seurat v5.0.

* 1. **Single cell type identification**

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 scored and regressed out. Sample integration was conducted using 50 principal components (PCs) and Harmony. Clustering of integrated cells was performed using 30-PCs. Differential expression was performed with Seurat FindAllMarkers tool. Canonical marker genes, together with gene set enrichment analysis (GSEA) conducted using gseGO (ClusterProfiler v4.6), was used to assign 11 cell types: 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 inference**

Gene-level counts were exported and analyzed with inferCNV. NK cells were used as a reference.

* 1. **Cell communication analysis**

CellChat 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. Homo sapiens 1390 curated TFs (<https://github.com/Ambuj-UF/Pediatric_glioblastoma_singelcell/blob/main/Notebook/Step9_Pyscenic/in/hs_hgnc_curated_tfs.txt>) 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/>) is then used with pyscenic ctx to create regulon enrichment matrix. pyscenic aucell is used to create loom file with regulon enrichment results and AUC values. Here, TF regulons with AUC > 0.5 and spearman correlation > 0.5 with SPP1 in MGD-TAM, MGD-macrophage, or MD-macrophage will be considered as a strong regulator of SPP1 in macrophage cells.

* 1. **Solute tempered replica exchange molecular-dynamics (st-REMD) simulation**

A full-length human SPP1 model was subjected to replica exchange molecular dynamics (REMD) simulations at 283.15 K, 303.15 K, 333.15 K, and 353.15 K (four replicas, 200 ns each), using GROMACS 2023 and the CHARMM36 force field. Systems were solvated in a rectangular box with TIP3P water molecules at 10 Å marginal radius. Systems were neutralized by added potasium ions (K+) and chloride (Cl-) ions. Simulation forcefield parameters were generated using CHARMM-GUI webserver. 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) and a real-space cutoff of 1.2 nm. 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 (τ = 5.0 ps, compressibility = 4.5e-5 bar⁻¹). Electrostatics were treated using PME with a 1.2 nm cutoff. 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 was used to predict likely phosphorylation sites on SPP1 protein. Site with score >0.8 is picked for downstream analysis.

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

The relaxed SPP1 structure obtained from replica exchange molecular dynamics (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 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. Except for the change in temperature and occasional variations in chain configuration, all other simulation parameters were identical to those used in the st-REMD setup. Quadratic mean of rmsf values across three chains were calculated for each residue in unmodified and pSer169 SPP1 structure.

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

Complementarity-determining regions (CDRs) of parental anti-SPP1 antibody 23C3 were identified using **PyIgClassify2.** 23C3 CDR region was docked to the relaxed unmodified SPP1 CD44 binding domain (residue 121-140) with Haddock. 2,500 23C3 antibody variants were generated with Rosett antibody design tool (antibody\_designer.linuxgccrelease). For every variant, full heavy and light chains sequence embeddings were obtained using ESM-2 8M model. CDR-specific embeddings (H1–H4, L1–L4) were extracted from their corresponding chain full sequence embedding. Cosine distances between variant CDR sequence and 23C3 CDR was calculated and combined with Rosetta dG\_separated. 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 topic hit variant. Mutations were mapped onto the complex with PyMOL v3.0.

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, we performed single‐cell RNA sequencing on freshly resected pHGG specimens. Unsupervised clustering and UMAP projection identified eleven 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 likely represents doublets or rare stromal elements.

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

We next applied 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 alterations—such as chromosome 7 gain and chromosome 10 loss—while immune clusters show flat profiles consistent with a normal karyotype. Among tumor states, the MES-like and MES-AC-like Cycling populations exhibited the most pronounced copy gains (for example, focal amplification on 7p), underscoring their heightened genomic instability relative to OPC- and APC-like cells.

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

Each single cell clusters were annotated with their cell type using their canonical gene expression markers (Fig 1c). To dissect the biological roles of individual cell populations within the pediatric high-grade glioma (pHGG) tumor microenvironment (TME), we performed gene set enrichment analysis (GSEA) and visualized normalized enrichment scores (NES) as a heatmap (Fig. 1d). T cells exhibited strong enrichment for the T cell receptor complex, MHC class II protein complex, and immunological synapse, consistent with an antigen-experienced, though likely restrained, T cell state within the glioma TME. Microglia-derived TAMs (MGD TAMs) were functionally distinct, marked by a suite of enrichment signatures consistent with a metabolically active and immune-modulatory phenotype. Enrichment of ribosome pathways indicates high protein synthesis activity, which is often seen in activated or 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. 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 myeloid derived macrophages displayed an enrichment profile indicative of active immune modulation. 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, upregulation of Golgi stack, Golgi cisterna membrane, and Golgi apparatus subcompartment points to heightened protein processing activity, a hallmark of macrophages in functional transition. Notably, ficolin-enriched granule pathways imply a potential role in complement activation via the lectin pathway, which has dual effects in cancer: either promoting immune evasion through anaphylatoxins or mediating tumor cell lysis via membrane attack complex formation.

MES-APC-like cells displayed a strong oxidative and translational metabolic program, suggestive of aggressive, invasive behavior. Enrichment of mitochondrial inner membrane, NADH dehydrogenase complex, respiratory chain complex I, cytochrome complex, and proton-transporting ATP synthase complex collectively points to high oxidative phosphorylation (OXPHOS) activity, supporting the energy demands of invasion and proliferation. Additionally, pathways such as ribosome, mitochondrial ribosome, ribonucleoprotein complex, and rough endoplasmic reticulum position these cells as translationally hyperactive. Further enrichment of spliceosomal complexes-including U2- and U12-type spliceosome, pre-catalytic spliceosome, and tri-snRNP complex-suggests that alternative splicing is actively being used in transitionary MES phenotype. 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 transitional 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. This suggests a state of neuronal mimicry, wherein tumor cells partially acquire neuron-like features, potentially facilitating immune evasion or integration into the neural niche. MES-AC-like cycling cells were found exhibiting hallmark signatures of active mitosis. 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 implicates these cells in chromosome segregation and cytokinesis. Further 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. Gene enrichment signatures indicates that Tumor Microenvironment is actively under transition and actively trying to evade immune response.

* 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,b,c,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 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 M1/M2 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 transitioning macrophage cells are principal source of SPP1 in pHGG and that SPP1-CD44 engagement is poised to modulate T cell function in situ.

* 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 Macrophages. MAFB stands out as high expressed in MGD Macrophage (> 60% of cells) with high average expression, and moderately expressed in MGD TAM, 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 regulons to suppress SPP1 production. Therefore, alternative strategies, such as intervention at the level of upstream signaling pathways or epigenetic modulators could prove more effective.

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

SPP1 is a highly a 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 (<https://pmc.ncbi.nlm.nih.gov/articles/PMC2793339/>). Another mAb, 2K1 or its chimeric form C2K1, recognizes the equivalent epitope region of human osteopontin (162SVVYGLR168) and is shown to ameliorate collagen-induced arthritis (CIA) in cynomolgus monkey. Murine anti-osteopontin mAb, namely, 23C3, was designed which is of great potential in treatment of RA (<https://pmc.ncbi.nlm.nih.gov/articles/PMC2793339/>). Recent work showed that using an anti-SPP1 antibody (bioXcell Clone: 100D3) and (clone MPIIIB10) in combination with mIL13Rα2 CAR T cell therapy showed enhanced CAR T cell antitumor response (<https://www.biorxiv.org/content/10.1101/2025.04.01.646202v1.full>).

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 (Fig. 4a). Two stretches-residues 2K1, C2K1 (residue 162–168), and 23C3 (residue 43–48) binding regions consistently showed significantly lower fluctuation across replicas (Fig 4a) compared to the highly dynamic remainder of the protein. Importantly, the established CD44‐binding region (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 (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 the from this basin (Fig. 4c) for downstream analyses.

* 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‐interacting segment, 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 and 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 clone**

Starting from the parental 23C3 monoclonal antibody, in-silico library of 2,500 CDR-mutated variants with RosettaAntibodyDesign and assessed each candidate for (i) predicted binding free energy to unmodified SPP1 (Rosetta 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 is organized into discrete clusters that reflect sequence divergence; the color scale encodes the corresponding dG\_separated values. Few variants with high relative SPP1 binding stability are shown as red dots. To integrate the two optimization criteria, we plotted dG\_separated against the cosine-embedding distance from parental 23C3 (Fig. 5b). The global minimum-representing the most stable binder with minimal sequence drift-lies near the origin of the density cloud (black arrow). This variant combines the lowest computed binding energy with one of the smallest embedding distances (< 0.26), indicating that only 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/biophysical difference criteria, it is prioritized as the lead candidate.

Despite the encouraging in-silico affinity profile of our redesigned 23C3 variant, several caveats remain. First, all predictions were derived from static Rosetta scoring and short-timescale MD snapshots, which does not account for human adaptive immune recognition against its surface residues. Therefore, 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. Second, our 200-ns trajectories are likely insufficient to accurately capture slow, allosteric rearrangements in highly disordered SPP1. To address this, we will extend simulations to few 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. 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.