Recognition of the glycan headgroup of stage-specific embryonic antigen-4 by a cancer-targeting antibody

Caroline Soliman[a], Jia Xin Chua[b,c], Mireille Vankemmelbeke[b,c], Richard S. McIntosh[b], Andrew J. Guy[a], Ian Spendlove[b], Lindy G. Durrant[b,c], Paul A. Ramsland[a,d,e,1]

[a]School of Science, RMIT University, Melbourne, Victoria, Australia.
[b]Academic Department of Clinical Oncology, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, City Hospital Campus, Nottingham, United Kingdom.
[c]Scancell Limited, Academic Department of Clinical Oncology, University of Nottingham, City Hospital Campus, Nottingham, United Kingdom.
[d]Department of Immunology, Central Clinical School, Monash University, Melbourne, Victoria, Australia.
[e]Department of Surgery Austin Health, University of Melbourne, Melbourne, Victoria, Australia.

Running title: *Antibody recognition of SSEA-4*

[1]Corresponding author: Dr Paul A. Ramsland, *Vice Chancellors Principal Research Fellow*, School of Science, RMIT University, Plenty Road, Bundoora, Victoria 3083, Australia. Phone +61 3 9925 7024, email paul.ramsland@rmit.edu.au

Keywords: antibody-based tumor targeting, carbohydrate-binding antibody, glycolipid antigen, stage-specific embryonic antigen, anti-cancer antibody

ABSTRACT

Cancer remains a leading cause of morbidity and mortality worldwide, requiring ongoing development of targeted therapeutics such as monoclonal antibodies. Carbohydrates on embryonic cells are often highly expressed in cancer and are therefore attractive targets for antibodies. Stage-specific embryonic antigen-4 (SSEA-4) is one such glycolipid target expressed in many cancers, including breast and ovarian carcinomas. Here we define the structural basis for recognition of SSEA-4 by a novel monospecific chimeric antibody (ch28/11). Five X-ray structures of ch28/11 Fab complexes with the SSEA-4 glycan headgroup displayed highly similar 3D structures indicating a stable binding mode. By adopting a horseshoe-shaped conformation in a deep groove, the glycan headgroup likely sits flat against the membrane to allow the antibody to interact with SSEA-4 on cancer cells. Sialic acid of SSEA-4 plays a dominant role in dictating the exquisite specificity of the ch28/11 antibody. This is further supported by molecular dynamics simulations, which show the stability of SSEA-4 compared to SSEA-3. These high-resolution views of how a glycolipid interacts with an antibody should advance a new class of cancer-targeting immunotherapy.

INTRODUCTION

It has long been known that glycosylation of malignant cells differs significantly to that of healthy cells (1). Aberrant glycosylation in cancer (associated with several glycan epitopes) has been associated with tumor progression and metastasis, and consequently, glycans are potential targets for therapeutic antibodies (2-5). Specifically, glycosphingolipids (GSLs) are a promising class of glycan antigens for antibody targeting in cancer. Acting as structural components of cell membranes GSLs are composed of one or more monosaccharides attached to either a sphingoid or a ceramide lipid. GSLs have important roles in immune cell function, membrane signaling, cell adhesion, apoptosis and cell differentiation (6-8).

Furthermore, GSLs are often overexpressed in various types of human malignancies, including GD2 and GD3 in melanoma (9,10) and Globo-H in breast and ovarian cancers (11).

Stage-specific embryonic antigens (SSEAs) are a family of glycoconjugate antigens, consisting of SSEA-1 (also known as CD15 or Lewis X), and two related GSLs SSEA-3 (also known as Gb5Cer) and SSEA-4 (also known as sialyl-Gb5Cer) (12). Both SSEA-3 and SSEA-4 are known cell surface markers of human embryonic and pluripotent stem cells. While their biological function is relatively unclear, they are known to be overexpressed in many cancers, including breast and ovarian cancers and may be very relevant to cancer stem cells (11,13-17). SSEA-3 and SSEA-4 share a common core glycan structure (Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ1), with SSEA-4 containing a terminal sialic acid (Neu5Acα2-3Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ1) and both glycan headgroups are linked to a ceramide lipid. SSEA-4 is synthesized from an SSEA-3 precursor through the action of β-galactoside α2,3-sialyltransferase 2 (ST3GAL2). Overexpression of ST3GAL2 has been implicated in poor outcomes for various cancers, including breast and ovarian cancer, which suggests a potential role of SSEA-4 in the development or maintenance of a tumor environment (16,18). SSEA-4 has also been shown to be expressed on chemotherapy-resistant and cancer stem cell populations of breast and ovarian tumors (16,17). As such, SSEA-4 presents an attractive target for antibody-based cancer therapeutics.

Currently, the only commercially available monoclonal antibodies (mAbs) against SSEA-3 or SSEA-4 are MC631, a rat IgM anti-SSEA-3 mAb, and MC813-70, a mouse IgG3 anti-SSEA-4 mAb. However, both antibodies demonstrate some cross-reactivity, with MC813 cross-reacting with multiple sugars including SSEA-3 and Forssman (19). The potential cytotoxicity of MC631 has not yet been reported, but MC813-70 has been shown to induce complement-dependent cytotoxicity of highly expressing SSEA-4 glioblastoma multiforme cell lines *in vitro* and suppress tumor growth *in vivo* (19,20). However, without further engineering, these rodent mAbs are not suitable for human immunotherapy due to the human anti-rodent immune response that limits *in vivo* efficacy and safety.

In response to the over-expression of SSEA-3 and SSEA-4 on the surface of cancer cells, a panel of anti-SSEA mAbs were developed. The lead mAb was a mouse IgG3 known as FG28/11, which has direct cytolytic activity against SSEA-4 positive cells. Following characterization, FG28/11 was chimerized as a mouse-human IgG1 mAb and renamed to ch28/11 (Chua et al., unpublished data). Ch28/11 is currently undergoing preclinical development against a variety of SSEA-4 positive tumor types and its characterization will be reported elsewhere.

In this study, we report the crystal structure of ch28/11 Fab bound to the SSEA-4 glycan headgroup. From three crystal forms, five independent ch28/11 Fab:SSEA-4 complexes were determined at resolutions between 1.5 and 2.7 Å. All ch28/11 Fab:SSEA-4 complexes displayed near-identical three-dimensional (3D) structures for the glycan-antibody interface. Our findings from crystallography and molecular dynamics simulations explain the basis for ch28/11 mAb specificity for SSEA-4 and identify a critical role for the terminal sialic acid, which is not present in SSEA-3, in antibody recognition.

RESULTS

The FG28/11 antibody is highly specific for SSEA-4. ELISAs were used to reveal strong binding of FG28/11 to SSEA-4 and a complete absence of cross-reactivity with closely related glycans SSEA-3, Globo-H and Forssman. All of the glycan antigens contain a GalNAcβ1-3Galα1-4Galβ1-4Glc core with varying terminal sugar residues (sialic acid in SSEA-4) (Fig. 1). For comparison, anti-SSEA-3 antibody MC631, anti-SSEA-4 antibody MC813-70 and anti-Forssman antibody M1/87 were also tested. MC631 bound well to SSEA-3 and SSEA-4 with some binding to Globo-H, and MC813 bound well to SSEA-3, SSEA-4 and Forssman. As expected, MC1/87 only bound to Forssman and the negative control SLe^x binding antibody did not interact with any of the four glycans (Fig. 1A). Therefore, FG28/11 was shown to be monospecific for SSEA-4, while MC631 and MC813 both displayed cross-reactivity to related glycans.

The FG28/11 mAb was further assessed by glycan array, where binding was tested against 585 mammalian glycans. FG28/11 only bound to SSEA-4 with no cross-reactivity with any of the other glycans tested, whereas MC631 bound to SSEA-4, SSEA-3 and Galβ1-3GalNAcβ1-3Gal (Fig. S1). In addition, tumor cell binding was characterized by flow cytometry of a panel of cancer cell lines as well as immunostaining of tumors. FG28/11 showed strong binding to one ovarian cancer cell line (SKOV3) and weak to moderate binding to two other ovarian cancer cell lines (OVCAR5 and IGROV1) as well as two breast cancer cell lines (T47D and MCF7), but was negative for binding to colorectal cancer cell lines (COLO205 and HCT15) (Fig. 1B and Fig. S2). This was confirmed by immunostaining, where FG28/11 stained 28% (223/798) of breast and 14% (41/289) of ovarian tumors (Fig. 1C).

Production, crystallization, and structure determination of ch28/11 Fab:SSEA-4 complexes. Fab was produced by papain digestion of ch28/11 IgG and protein A purification, and the purity and uniformity confirmed by Coomassie-stained SDS-PAGE and size-distribution analysis of dynamic light scattering (DLS) data (Fig. 2A & B). As expected, the purified Fab was a disulfide-linked dimer of 50 kDa consisting of paired heavy (H) and light (L) chains of approximately 25 kDa sizes. By DLS, the Fab sample consisted of a single population with an average hydrodynamic diameter (D_H) of 6.6 nm (compared to IgG with an average D_H of 13.8 nm). Thus, ch28/11 Fab was determined to be suitable for structural studies.

To elucidate the structural basis for recognition of SSEA-4 by the ch28/11 mAb, crystal structures of the Fab:glycan headgroup complex were determined. Co-crystals of ch28/11 Fab and SSEA-4 hexasaccharide were obtained in three different conditions, resulting in three crystal forms, with the highest resolution structure determined at 1.5 Å (R_{work}/R_{free} = 0.172/0.197) from a tetragonal $P4_12_12$ crystal (Fig. 2C). The monoclinic $P2_1$ crystal structure was refined at 1.9 Å resolution (R_{work}/R_{free} = 0.176/0.230). Both the tetragonal and monoclinic crystals had a single ch28/11 Fab:SSEA-4 complex in the asymmetric unit. The hexagonal $P6_2$ crystal structure was refined at 2.7 Å resolution (R_{work}/R_{free} = 0.169/0.246) with three ch28/11 Fab:SSEA-4 complexes in the asymmetric unit (identified as complex 1, 2 and 3). X-ray diffraction data collection and crystallographic refinement statistics are presented in Table 1.

The electron density maps allowed fitting of the L and H polypeptide chains for each Fab structure, except for one disordered loop in the H chain constant domain, which is distant from the antigen-binding site and often missing from Fab structures (21). The missing CH1 residues were 133-137 for the 2 higher-resolution structures, 132-137 for complex 1 and 2, and 133-134 for complex 3 from the lowest resolution structure (sequential numbering). In all structures, there is strong electron density corresponding to binding site residues and five of the six sugar residues of SSEA-4 (slightly weaker electron densities were associated with the terminal glucose), but the entire hexasaccharide was fitted into electron density maps (Fig. 2D & Fig. S3).

Groove-type recognition of the SSEA-4 glycan by the ch28/11 mAb. The 3D structures for each of the five versions of ch28/11 Fab:SSEA-4 complexes were very similar, with SSEA-4 positioned to fill a deep groove-shaped binding site (Fig. 3). The Fab has a quaternary structure typical of most antibodies with the binding site formed by six complementarity-determining regions (CDRs), three from the L chain (identified as L1, L2, and L3) and three from the H chain (identified as H1, H2, and H3). The Fv portions (comprised of VL and VH; sequences presented in Fig. S4) were shown to be near-identical in 3D structure, with most of the minor variations occurring in loops surrounding the binding groove (CDRs and some framework region loops). The constant regions (comprised of CL and CH1) also showed some differences (Fig. 3A & Fig. 3B). A surface view of the 1.5 Å resolution structure illustrates that the center of the binding site is occupied by the H3 loop, with the SSEA-4 hexasaccharide bent around this loop in a horseshoe-like conformation (Fig. 3C).

Antibody bound conformation of the SSEA-4 hexasaccharide. In the ch28/11 binding site, the SSEA-4 hexasaccharide is surrounded by all six CDRs (L1-3 and H1-3) and snugly fills the entire binding groove (Fig. 3 & Fig. S5). This snug fit

corresponds to a stable glycan conformation where the galactose (GAL2, GLA4, and GAL5), N-acetyl-galactosamine (NGA3) and terminal sialic acid (SIA1) residues all display very little movement between structures (Fig. 4; see figure for monosaccharide codes). In contrast, the terminal glucose (BGC6) that would normally be attached to the ceramide lipid, exhibits more flexibility in its position (Fig. 4A & B), which corresponds with it being the most solvent-exposed portion of the SSEA-4 glycan (Fig. 3). Further examining the conformations of the five SSEA-4 hexasaccharide structures in a Ramachandran-like plot showed clustering of the *phi* and *psi* angles of the various glycosidic linkages (Fig. 4C). The same glycosidic conformations were observed for SIA1-GAL2 and GLA4-GAL5 as demonstrated by similar *phi* and *psi* angles. GAL2-NGA3 and NGA3-GLA4 glycosidic linkages have similar *psi* but different *phi* angles indicating different conformations for these linkages. In line with its position in the binding site and flexibility, a distinct cluster with a larger variation of *phi* and *psi* angles occurred for the GAL5-BGC6 glycosidic linkage (Fig. 4C).

Molecular details of ch28/11 Fab recognition of the glycan headgroup of SSEA-4. In each of the five structures, the Fab participates in numerous hydrogen-bond and van der Waals interactions with each of the six sugar residues of SSEA-4 (Fig. 5 & Fig. S6). All antibody residues involved in binding SSEA-4 are located within the L and H chain CDRs, except for one residue (His 33L), which forms a single hydrogen-bond with the glycan ligand in each structure (see Fig. S4 for glycan contact residues and numbering schemes). A total of 791.5 Å^2 of the SSEA-4 glycan headgroup is buried in the ch28/11 binding groove. Contributions to the interaction of ch28/11 mAb and glycan for the constituent monosaccharides are: SIA1, 202.0 Å^2; GAL2, 117.5 Å^2; NGA3, 167.6 Å^2; GLA4, 141.5 Å^2; GAL5, 118.7 Å^2; BCG6, 44.2 Å^2.

Within the expansive ch28/11 binding groove, contacts are made with each residue of SSEA-4. Sialic acid (SIA1) dominates the interaction with the antibody by participating in more contacts compared to any other residue in the glycan chain. It is secured in place by hydrogen bonds between the nitrogen on the N-acetyl moiety and Ser-31H, the O8 hydroxyl and Gly-33H, and from the carboxylate group to Gly-53H and Asp-54H. The N-acetyl-galactosamine (NGA3) of SSEA-4 is involved in hydrogen bonds from the O4 and O6 hydroxyls to Tyr-93L and Ala-90L, respectively. The α-galactose (GLA4) engages in one hydrogen bond from the O4 hydroxyl to Ala-90L and two hydrogen bonds from the O6 hydroxyl to His-33L and Gly-101H. The next galactose in the chain (GAL5) is involved in three hydrogen bonds from the O2 and O3 hydroxyls to Asp-49L. In contrast, galactose at position 2 (GAL2) and glucose at position 6 (BGC6) do not engage in direct hydrogen bonding with the ch28/11 mAb. SSEA-4 is further stabilized by numerous van der Waals interactions between binding site residues and each monosaccharide in the glycan chain (Fig. 5).

The ch28/11:SSEA-4 interface (at 1.5 Å resolution) has a well-defined solvent structure with 23 ordered water molecules surrounding the SSEA-4 glycan. Sialic acid (SIA1) again dominates the interactions with 8 ordered waters involved in stabilising its conformation in the binding groove. Other monosaccharide units of SSEA-4 are involved in fewer water-mediated interactions: GAL2, 5 waters; NGA3, 4 waters; GLA4, 4 waters; GAL5, 3 waters; BCG6, 3 waters.

A solvent-accessible surface representation shows SSEA-4 snugly fitted into the binding groove of the Fab, folding around the centrally located H3 loop in a horseshoe-shaped conformation, with the N-acetyl-galactosamine hydrophobic face stacking against the Tyr-102H side chain to slot into its position in the deepest part of the binding groove (Fig. 6A & B). Thus, the ch28/11 binding groove is totally occupied by the SSEA-4 glycan headgroup and the ordered water molecules to present an almost continuous convex surface at the end of the Fab. The absence of protruding CDR loops would permit the intact ch28/11 mAb to bind closely to the cell membrane where the SSEA-4 hexasaccharide is directly anchored by its ceramide tail. This could be relevant for the direct cytotoxic activity of the ch28/11 antibody.

Structure of the unliganded ch28/11 Fab reveals conformational changes are involved in recognition of SSEA-4. To further examine the SSEA-4 binding mechanism, the crystal structure of the unliganded ch28/11 Fab was determined. X-ray diffraction data was obtained

from a monoclinic $P2_1$ crystal at 2.5 Å resolution (R_{work}/R_{free} = 0.164/0.239; Table 1). There were two Fabs in the asymmetric unit, both with clear electron density maps in the binding site (Fig. S3). Differences in 3D structure of the binding sites are mainly due to crystal packing, where the variable domains of Fab 2 are packed against the light chain of the constant domain of Fab 1. To compare free with bound structures, the Fv regions from the unliganded Fabs were superimposed on the 1.5 Å resolution ch28/11 Fab:SSEA-4 structure. For the most part, there is little variation in binding site residues, with the exception of CDR3 of the heavy chain. In the free Fab structures, this loop appears to collapse into the binding groove, with the Tyr-102H side chain in particular shifting inward to fill part of the groove (Fig. 7 and Fig. S5). A multitude of small conformational differences occur in the all H chain CDRs around the site of interaction with sialic acid of SSEA-4. Overall these results suggest SSEA-4 recognition by the ch28/11 mAb represents an induced fit binding mechanism.

Structural basis for ch28/11 specificity for SSEA-4. Molecular dynamics (MD) simulations were performed to examine why the ch28/11 mAb does not cross-react with SSEA-3, which compared to SSEA-4 is only missing the terminal sialic acid residue. Carbohydrate flexibility can be assessed by examining root-mean-square fluctuation (RMSF) of carbohydrate residues, which is a time-averaged measure of atom deviation from a reference position. RMSF analysis of carbohydrate residues, averaged over 1.5 μs of MD simulations, show that the central galactose and N-acetyl-galactosamine residues (numbered 3-5) are highly stable in the binding site for both SSEA-3 and SSEA-4. Additionally, in both glycans the terminal glucose (BCG6) was mobile in the binding site, although more so for SSEA-3. For SSEA-3, the galactose at position 2 (GAL2) displayed high RMSF values, while the same residue in SSEA-4 was highly stable across the MD simulations (Fig. 8A & B). This is further evident when examining the *phi* (ϕ) vs. *psi* (ψ) plots of each glycosidic linkage. The linkage between galactose (GAL2) and N-acetyl-galactosamine (NGA3) is more flexible in SSEA-3 displaying a range of conformations when compared to SSEA-4. It should also be noted that the linkage between galactose (GAL5) and the terminal glucose (BGC6) is also more flexible in SSEA-3 when compared to SSEA-4 (Fig. 8C & D). When overlaying the 2 main ligand conformations of SSEA-3 taken from MD simulations, there is good alignment of the three central carbohydrate residues, and the largest difference in conformation occurs with the GAL2 and BCG6 monosaccharides (Fig. 8E). Therefore, it can be concluded that the ch28/11 binding groove is more stable in the presence of SSEA-4 due to the presence of the terminal sialic acid residue absent in SSEA-3. Furthermore, the sialic acid (SIA1) of SSEA-4 was highly stable during MD simulations (Fig. 8B) as is the SIA1-GAL2 glycosidic linkage (Fig. 8D), which supports its critical role in binding and specificity for SSEA-4 by the ch28/11 mAb.

DISCUSSION

Traditionally cancer therapies have been broad acting and non-specific, such as surgery, chemotherapy, and radiotherapy. However, there has been an increasing trend towards targeted therapeutics with greater specificity and reduced off-target effects. These include several glycoprotein-specific antibody-based therapies, such as trastuzumab (Herceptin) for metastatic HER2-overexpressing breast cancers (22) and cetuximab (Erbitux) for HER1-positive colorectal carcinomas (23). Dinutuximab (Unituxin) was the first carbohydrate-binding antibody approved as a treatment for GD2-positive pediatric neuroblastomas (24). Another class of successful cancer therapeutic antibody, generically called checkpoint inhibitors, work by activating the immune system of patients. Examples include ipilimumab (Yervoy) that targets CTLA-4 for the treatment of advanced melanoma and renal carcinomas (25) and atezolizumab (Tecentriq) that binds PD-L1 and was first introduced to treat bladder and urinary tract cancers (26,27). Thus, most antibody-based therapies either target specific signaling pathways or cancer-specific antigens, allowing immune recognition of cancer cells. Despite their growing use, most approved therapeutic antibodies are only indicated for specific and limited types of malignancies. Although there are many antibody-based therapies in development, there is a distinct need for continued identification of new mAbs to complement existing therapies (5,28,29).

Antibodies targeting glycans and particularly glycolipid antigens have great potential for cancer immunotherapy but have to date progressed slowly mainly due to challenges with development, specificity and efficacy (3,4). In addition, for glycan antibodies to have meaningful prognostic value, monospecificity is crucial. Collectively, our structural studies with the newly developed ch28/11 mAb have provided unique and surprising insights into its recognition of SSEA-4, which represents a promising glycolipid target in a range of cancers (11,13-15). The SSEA-4 glycan headgroup is almost completely engaged by snugly fitting into a deep groove of ch28/11 Fab in a horseshoe-shaped conformation. SSEA-4 is securely bound by numerous hydrogen bonds, van der Waals interactions and a large network of well-defined water molecules. The high level of binding site occupancy enables the secure antigen specific binding between ch28/11 and SSEA-4, which is replicated across the five high-resolution crystal structures. In contrast, Kannagi et al. showed that SSEA-4 antibody MC813 recognizes the terminal 3 sugars (NeuAcα2-3Galβ1-3GalNAc) whereas anti-SSEA-3 antibody MC631 recognises the middle 3 sugars (GalNAcβ1-3Galα1-4Gal). This could explain the exquisite specificity displayed by ch28/11 towards SSEA-4 compared to the cross-reactivity of MC813 with SSEA-4, SSEA-3 and Forssman glycans (see Fig. 1) (20).

In the context of an intact ch28/11 mAb binding to the surface SSEA-4 positive cells, binding must involve intimate contact with the cell membrane where the glycan epitope is pushed flat against the lipid bilayer. Interestingly, the terminal sialic acid of SSEA-4 dominates interactions with the ch28/11 mAb in the section of the groove formed exclusively by H chain CDRs. An induced fit mechanism of SSEA-4 recognition is evident by comparisons of the bound and free ch28/11 Fab structures. The ligand-induced conformational changes in the binding groove and sialic acid pocket may partly explain the specificity of ch28/11 mAb for SSEA-4. More compelling is the increased mobility of the related SSEA-3 glycan, which lacks the terminal sialic acid and is not recognized by ch28/11. All these observations point towards binding and specificity for SSEA-4 by the ch28/11 mAb requiring at least 5 monosaccharides of the glycan headgroup and critically depending on the presence of the sialic acid residue. Also considered as a cancer-related glycolipid, Globo-H is an α1-2 fucose substituted SSEA-3 and is not recognized by ch28/11 (Fig. 1). The proximity of the Gal O2 to the H3 loop and the prominent Tyr 102H side-chain (see Fig. 5) suggests that Globo-H does not cross-react with ch28/11 due steric clashes imposed by the terminal fucose.

To date, ch28/11 is only the second glycolipid binding antibody for which a structure of the complex with the glycan has been resolved. The GD2 binding antibody 14G2a recognizes the branched glycan headgroup by end-on insertion and interacts with 4 of the 5 monosaccharides. This suggests a perpendicular binding mode of the 14G2a mAb for the GD2 ganglioside (30) as compared to the parallel binding proposed for the ch28/11 mAb.

Crystal structures for several other glycolipid mAbs have been determined as free Fabs, with docking and molecular modeling used to predict the carbohydrate binding mode. Three studies suggest that the terminal sialic acid of GD3 and NeuGc-GM3 bind in pockets formed by the H chain only (31-33). In two of the reported Fab structures (R24 and chP3 mAbs), the CDR loops of the H chain form binding pockets large enough to accommodate the terminal sialic acid by end-on insertion (31,32). For the NeuGc-GM3 binding antibody 14F7, the structure revealed an extended CDR H3 loop that divides the binding site into 2 subsites, with the model showing the ligand binding to the subsite formed solely by H chain (33). In another structure, molecular docking simulations of the anti-GD2 antibody 3F8 showed the terminal sialic acid of GD2 binding between the H and L chains, with the second sialic acid binding primarily to the H chain (34). Similar to these predicted interactions, ch28/11 also binds sialic acid entirely through a subsite formed by the H chain CDRs, which supports the concept that the H chain repertoire may be predisposed towards sialic acid recognition.

By comparison, the structure of another GD2 binding antibody ME36.1 has a relatively straight groove-shaped binding site (not a pocket), where it is proposed that all six CDRs are involved in binding to the tetrasaccharide (Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc) in an extended conformation (35). This type of groove-shaped recognition has been observed

for antibody binding to an extended pentasaccharide epitope expressed by many microbes (36). In contrast, ch28/11 binds the glycan in a horseshoe-shaped conformation in a groove that accommodates the hexasaccharide glycan headgroup of SSEA-4.

Inspection of the sialic acid binding pocket of ch28/11 shows the carboxylate anchored by two hydrogen bonds to main-chain and the N-acetyl nitrogen and O8 positions also hydrogen bonded to binding site residues (Fig. 5). The remaining oxygen atoms of the sialic acid interact with ordered water molecules. Similar binding modes have been observed in viral and bacterial lectins where the carboxylate of the sialic acid is anchored by at least 2 hydrogen bonds. However, other lectins interact with the sialic acid carboxylate by ionic interactions with arginine or lysine residues (based on PDB structures annotated by ProCarbDB) (37). Furthermore, most sialic acid binding lectins interact with the terminal sialic acid or only a couple of monosaccharides of longer glycan ligands, compared to ch28/11 interacting with all 6 monosaccharides of the SSEA-4 glycan headgroup. Sialic acids of tumor glycans can be incorporated from the diet as the N-glycolyl form that is not normally found in humans (i.e., Neu5Gc) or can be O-acetylated usually at the O9 position (38,39). Although we have not experimentally tested these modifications, both Neu5Gc or O9-acetylation appears compatible with the ch28/11:SSEA-4 interaction as both sites are solvent exposed (see Fig. 5) and relatively unobstructed by binding site residues.

Here we define the molecular basis for recognition of SSEA-4 by ch28/11, a newly developed mouse-human chimeric antibody. Our findings reveal the structural basis for the ch28/11 mAb selectively binding to SSEA-4. Understanding the molecular details of glycan recognition will allow for future structure-based design of highly effective glycan specific mAbs for cancer targeting.

EXPERIMENTAL PROCEDURES

Direct binding ELISAs of glycan antigens coupled to BSA. Glycan-BSA conjugates were obtained from Elicityl, France. Glycans were characterised by NMR prior to conjugation and TLC and HPAEC-PAD after conjugation with 10-40 glycan molecules per BSA. ELISA plates were coated overnight at 4°C with 100 ng/well glycan-BSA antigens in PBS, blocked with PBS 10% w/v BSA and incubated with primary antibodies (5 µg/ml). Primary antibodies were FG2811 (anti-SSEA-4) (University of Nottingham, UK), MC631 (anti-SSEA-3) (eBioscience, USA), MC813 (anti-SSEA-4) (Abcam, UK), M1/87 (anti-Forssman) (Santa Cruz Biotech, USA), and anti-sialyl-Lewis X (Biorbyt, UK). Glycan antigen binding was detected using anti-mouse, anti-human or anti-rat IgG Fc specific biotinylated antibody (Sigma, UK) [1/5000 dilution in PBS 1% w/v BSA], followed by streptavidin horseradish peroxidase (HRPO) conjugate (Sigma, UK) [1/5000 dilution in PBS 1% w/v BSA] and development with 3,3',5,5'-Tetramethylbenzidine (TMB). Absorbance was measured at 450nm using a Tecan infinite F50 plate reader.

Glycan array analysis. FG28/11 and MC631 were screened for binding to 585 glycans on the Consortium for Functional Glycomics (CFG) mammalian glycan array (core H group, version 5.1). Slides were incubated with 10µg/ml of antibody for 1 hour, before detection with Alexa Fluor488-conjugated secondary mAb.

Immunohistochemistry assessment. Tumor tissue binding was analyzed by immunohistochemistry as described previously (40). Briefly, after antigen-retrieval, blocking of endogenous peroxidase activity and non-specific binding sites, the sections were incubated with primary mAbs at room temperature for 1 hour. Primary mAb binding was detected by biotinylated secondary mAb (Vector Labs, USA) followed by preformed streptavidin-biotin/HRPO (Dako Ltd, UK) and 3, 3'-Diaminobenzidine as the substrate. Finally, sections were counterstained with hematoxylin. Staining was analyzed via New Viewer software 2010.

Patient cohorts. The study populations include cohorts of consecutive series of 350 ovarian cancer (41) samples (1982-1997; median follow up 192 months: censored November 2005: patients with stage II to IV disease received standard adjuvant chemotherapy which in later years was platinum based), 798 primary breast cancer (42) samples (1987-1998; median follow up, 177 months). Patients were selected for systemic adjuvant treatment on the basis of

Nottingham Prognostic Index (NPI) score and hormone receptor status [patients with NPI score ≤ 3.4 received no adjuvant therapy; those with an NPI score higher than 3.4 received tamoxifen if they were estrogen receptor (ER) positive (± goserelin if premenopausal) or classical cyclophosphamide, methotrexate and fluorouracil if they were ER negative and if capable of tolerating chemotherapy].

Flow cytometry. Cells (1 x 10^5) were incubated with primary mAbs at 4ºC for 1 hour followed by FITC-conjugated anti-mouse IgG Fc specific secondary mAb (Sigma, UK). Cells were resuspended in PBS and run on a Beckman Coulter FC-500 with WinMDI 2.9 for analysis.

Recombinant ch28/11 IgG and Fab production. Recombinant ch28/11 IgG was expressed in Expi293 cells (Thermo Fisher Scientific) and purified by filtration and fast protein liquid chromatography (FPLC) using HiTrap protein G columns (GE Healthcare). A Pierce Fab Preparation Kit (Thermo Fisher Scientific) was used to produce Fab from IgG. Briefly, 0.5 mL of an 8 mg/mL IgG sample was digested with Papain-agarose for 6 hours at 37°C. Fc and residual intact IgG was separated from Fab using protein A affinity chromatography. Coomassie stained SDS-PAGE (precast 4-15% Bis-Tris Mini Gels and MES running buffer, BioRad) was used to examine the purity of the Fab compared to the Fc and intact IgG samples under non-reducing and reducing (β-mecaptoethanol) conditions. Fab was quantitated by absorbance at 280 nm (Nanodrop) assuming a mass extinction (E, derived concentrations in mg/mL) value of 1.0 (1.37 for intact IgG). DLS was used to determine protein size and polydispersity using a Zetasizer nano ZS instrument (Malvern Instruments). Disposable cuvettes containing 100 µL of protein sample were measured (five replicates) at 25°C, to generate intensity reports showing size distribution by intensity to identity population peaks (by size (d.nm) and % intensity), and Z-average hydrodynamic diameters (D_H, in nm) and polydispersity were estimated by the cumulants method.

Crystallization of ch28/11 Fab in complex with SSEA-4 hexasaccharide. For crystallization, ch28/11 Fab was dialysed overnight into ultrapure water (Milli-Q) using Slide-A-Lyser MINI Dialysis devices with a 10-kDa-cutoff. Fab samples were concentrated to 10-20 mg/mL using Pall Corp Nanosep OMEGA devices with a 3 kDa cut-off membrane. Co-crystallization of the SSEA-4 glycan headgroup (SSEA-4 hexaose, Carbosynth) with ch28/11 Fab was achieved with a molar ratio of 5:1 (carbohydrate:protein). Initial crystals were generated by the sitting-drop vapor diffusion method at 18°C in a 96-well sitting drop plate (Corning) using the Crystal Screen High Throughput kit (Hampton Research). Reservoirs of 80 µL of each crystallization condition were used with 1 µL of Fab:SSEA-4 sample mixed with 1 µL of the same condition for the sitting drop. Improved crystals were obtained in 24-well hanging drop plates (18°C) with 1 mL reservoirs and 1µl of Fab:SSEA-4 mixed with 1 µL of crystallization condition for the hanging drop. Crystals suitable for X-ray diffraction were produced in the following crystallization conditions: 25% v/v polyethylene glycol (PEG) 4000, 0.1 M sodium acetate pH 4.6 and 0.2 M ammonium sulfate ($P4_12_12$ crystal form); 20% PEG 8000, 0.1 M Sodium Cacodylate pH 6.5 and 0.1 M Magnesium Acetate ($P2_1$ crystal form); 20% PEG 4000, 10% iso-Propanol and 0.1 M Sodium HEPES pH 7.5 ($P6_2$ crystal form). The free ch28/11 Fab crystals crystallized in 20% PEG 4000, 10% iso-Propanol and 0.1 M Sodium HEPES pH 7.5. For data collection, crystals were mounted in nylon loops (Hampton Research) in the drop solution and plunged into liquid nitrogen. Some crystals from the highest-resolution condition were frozen in the presence of 10% ethylene glycol in crystallization media as a cryoprotectant, but a cryoprotectant was not needed for most crystals.

Data collection and structure determination. Diffraction data were collected at the Australian Synchrotron using the MX2 beamline, by the ultrafine φ-slicing data-collection method using an EIGER X 16M detector and the qeGUI graphical user interface. Diffraction data were auto processed on the MX2 beamline using automated indexing with *xdsme* (using *XDS* and *Pointless*) and *AIMLESS* (43,44). Data processing and *hkl* file conversions were implemented in the XDS and the CCP4 program package (45,46). All modeling and crystallographic refinements were performed using Phenix, COOT and REFMAC (47-49). Validation of SSEA-4 was performed in the CCP4 program package with PRIVATEER

(45,50). LigPlot+ version 1.4 was used to generate the ligand interaction diagrams (51). Figures were generated using Discovery Studio (Dassault Systèmes BIOVIA, USA).

Crystals belonged to space groups $P4_12_12$ (1.5 Å resolution), $P2_1$ (1.9 Å resolution) and $P6_2$ (2.7 Å resolution), with the highest resolution structure determined by molecular replacement using Protein Data Bank (PDB) code 1A7N for VL:VH and PDB code 3IU4 for CL:CH1. Crystals of the free Fab belonged to space group $P2_1$ (2.5 Å resolution). Several rounds of fitting of the atomic model to electron density and crystallographic refinement were conducted. The final atomic model of the highest-resolution structure was used for molecular replacement of the other versions. Following multiple rounds of crystallographic and amino acid refinement, all ch28/11 Fab: SSEA-4 structures were of high resolution with strong fit of the ligands in the binding site. Data collection and refinement statistics are reported in Table 1.

MD simulations of complexes of ch28/11 Fab with SSEA-3 and SSEA-4. MD simulations were run using GROMACS 2018.2 (52,53). Proteins were parameterized using the AMBER99SB-ILDN forcefield (54). Carbohydrate topologies were generated using the Glycam06j forcefield via GLYCAM-Web (http://glycam.org), converted to GROMACS format using ACPYPE (55,56) and combined with the protein topology to form the complete protein-carbohydrate system. Initial coordinates for the protein-carbohydrate system for both SSEA-3 and SSEA-4 were based on the 1.5 Å resolution ch28/11 Fab:SSEA-4 structure (with the terminal sialic acid residue removed for SSEA-3). The protein-carbohydrate system was placed in a rhombic dodecahedral box with ≥10 Å distance between the molecule and the edge of the box. The system was solvated using the TIP3P water model, then ionised and neutralised with Na^+ and Cl^- to a concentration of 0.15 M. Energy minimization was performed using the steepest descent algorithm (5000 steps), follow by equilibration at constant volume and temperature (NVT ensemble) for 100 ps, with each replicate initialized with random velocities sampled from a Maxwell distribution at 300 K. This was followed by equilibration with constant pressure and temperature (NPT ensemble) for 300 ps. Pressure coupling was achieved via the Parrinello-Rahman barometer with a time constant of 2 ps and a reference pressure of 1 bar. Temperature coupling was achieved using the modified Berendsen thermostat, with separate temperature coupling groups for ligand/protein and water/ions respectively. A time constant of 0.1 ps was used for temperature coupling, with a reference temperature of 300 K. Equations of motion were integrated using the leap-frog integrator with a time-step of 2 fs. Hydrogen bonds were constrained during all steps using the LINCS constraint algorithm (57). Long range electrostatics were calculated using the Particle Mesh Ewald (PME) method with cubic interpolation. Neighbor searching was performed using the Verlet cut-off scheme with a distance cut-off of 1.0 nm for both Coulomb and van der Waals interactions. Simulations were run for 500 ns with 3 replicates per complex (for a total of 1.5 μs per complex) using the Spartan high-performance computer system (University of Melbourne).

Analysis of MD simulations was performed using a combination of GROMACS 2018.2 and the MDTraj Python package (v1.9.1). Root-mean-square fluctuation (RMSF) of carbohydrate atoms was calculated relative to the protein backbone carbon atoms (i.e. protein backbone carbon atoms were used to remove rotational and translational movement across frames, and RMSF calculated for these transformed trajectories). The glycosidic *phi* and *psi* angles were calculated using the MDTraj package. All visualization was performed using the Matplotlib 1.5 and Seaborn 0.9.0 Python libraries. Figures were generated in VMD 1.9.3 and Discovery Studio (Dassault Systèmes BIOVIA, USA).

ACKNOWLEDGEMENTS

This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the ACRF detector. This research was also undertaken using the LIEF HPC-GPGPU Facility hosted at the University of Melbourne. This Facility was established with the assistance of LIEF Grant LE170100200. CS is supported by a research training program stipend

scholarship from the Australian Government, Department of Education and Training. PAR is supported by a Vice Chancellor's senior research fellowship from RMIT University.

CONFLICT OF INTEREST STATEMENT

Caroline Soliman, Andrew J. Guy, Richard S. McIntosh, Ian Spendlove and Paul A. Ramsland declare that they have no conflicts of interest with the contents of this article. Lindy G. Durrant is director and CSO of Scancell Ltd. and has ownership interest (including patents) in Scancell Ltd. Mireille Vankemmelbeke and Jia Xin Chua are employees of Scancell Ltd.

AUTHOR CONTRIBUTIONS

C.S., L.G.D. and P.A.R. designed research; C.S., J.X.C. A.G., and P.A.R. performed research; C.S., J.X.C., A.G., and P.A.R. analyzed the data; J.X.C., M.V., R.S.M., I.S., L.G.D. contributed new reagents/analytic tools; C.S., and P.A.R. wrote the paper; and A.G., J.X.C., M.V., R.S.M., I.S. and L.G.D. reviewed and commented on the paper.

DATA DEPOSITION

The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org [PDB ID codes 6UG7, 6UG8, 6UG9 (ch28/11 Fab:SSEA-4), and 6UGA (ch28/11 Fab)].

REFERENCES

1. Meezan, E., Wu, H. C., Black, P. H., and Robbins, P. W. (1969) Comparative studies on the carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. II. Separation of glycoproteins and glycopeptides by sephadex chromatography. *Biochemistry* **8**, 2518-2524
2. Munkley, J., and Elliott, D. J. (2016) Hallmarks of glycosylation in cancer. *Oncotarget* **7**, 35478-35489
3. Soliman, C., Yuriev, E., and Ramsland, P. (2017) Antibody recognition of aberrant glycosylation on the surface of cancer cells. *Curr. Opin. Struct. Biol.* **44**, 1-8
4. Durrant, L. G., Noble, P., and Spendlove, I. (2012) Immunology in the clinic review series; focus on cancer: glycolipids as targets for tumour immunotherapy. *Clin. Exp. Immunol.* **167**, 206-215
5. Daniotti, J. L., Vilcaes, A. A., Torres Demichelis, V. E., Ruggiero, F. M., and Rodriguez-Walker, M. (2013) Glycosylation of glycolipids in cancer: basis for development of novel therapeutic approaches. *Front. Oncol.* **3**, 306
6. Zhang, T., de Waard, A. A., Wuhrer, M., and Spaapen, R. M. (2019) The Role of Glycosphingolipids in Immune Cell Functions. *Front. Immunol.* **10**, 90
7. Wang, Z., Wen, L., Ma, X., Chen, Z., Yu, Y., Zhu, J., Wang, Y., Liu, Z., Liu, H., Wu, D., Zhou, D., and Li, Y. (2012) High expression of lactotriaosylceramide, a differentiation-associated glycosphingolipid, in the bone marrow of acute myeloid leukemia patients. *Glycobiology* **22**, 930-938
8. Hakomori, S. I. (2008) Structure and function of glycosphingolipids and sphingolipids: recollections and future trends. *Biochim. Biophys. Acta* **1780**, 325-346
9. Nudelman, E., Hakomori, S., Kannagi, R., Levery, S., Yeh, M. Y., Hellstrom, K. E., and Hellstrom, I. (1982) Characterization of a human melanoma-associated ganglioside antigen defined by a monoclonal antibody, 4.2. *J. Biol. Chem.* **257**, 12752-12756
10. Scott, A. M., Lee, F. T., Hopkins, W., Cebon, J. S., Wheatley, J. M., Liu, Z., Smyth, F. E., Murone, C., Sturrock, S., MacGregor, D., Hanai, N., Inoue, K., Yamasaki, M., Brechbiel, M. W., Davis, I. D., Murphy, R., Hannah, A., Lim-Joon, M., Chan, T., Chong, G., Ritter, G., Hoffman, E. W., Burgess, A. W., and Old, L. J. (2001) Specific targeting, biodistribution, and lack of immunogenicity of chimeric anti-GD3 monoclonal antibody KM871 in patients with metastatic melanoma: results of a phase I trial. *J. Clin. Oncol.* **19**, 3976-3987
11. Chang, W. W., Lee, C. H., Lee, P., Lin, J., Hsu, C. W., Hung, J. T., Lin, J. J., Yu, J. C., Shao, L. E., Yu, J., Wong, C. H., and Yu, A. L. (2008) Expression of Globo H and SSEA3 in breast cancer stem cells and the involvement of fucosyl transferases 1 and 2 in Globo H synthesis. *Proc. Natl. Acad. Sci. USA* **105**, 11667-11672
12. Suzuki, S., Ishida, T., Yoshikawa, K., and Ueda, R. (2016) Current status of immunotherapy. *Jap. J. Clin. Oncol.* **46(3)**, 191-203
13. Gang, E. J., Bosnakovski, D., Figueiredo, C. A., Visser, J. W., and Perlingeiro, R. C. (2007) SSEA-4 identifies mesenchymal stem cells from bone marrow. *Blood* **109**, 1743-1751
14. Noto, Z., Yoshida, T., Okabe, M., Koike, C., Fathy, M., Tsuno, H., Tomihara, K., Arai, N., Noguchi, M., and Nikaido, T. (2013) CD44 and SSEA-4 positive cells in an oral cancer cell line HSC-4 possess cancer stem-like cell characteristics. *Oral Oncol.* **49**, 787-795
15. Yu, C.-C., Yu, C.-H., and Chang, Y.-C. (2016) Aberrant SSEA-4 upregulation mediates myofibroblast activity to promote pre-cancerous oral submucous fibrosis. *Sci. Rep.* **6**, 37004
16. Aloia, A., Petrova, E., Tomiuk, S., Bissels, U., Dnas, O., Saini, M., Zickgraf, F. M., Wagner, S., Spaich, S., Sutterlin, M., Schneeweiss, A., Reitberger, M., Ruberg, S., Gerstmayer, B., Agorku, D., Knobel, S., Terranegra, A., Falleni, M., Soldati, L., Sprick, M. R., Trumpp, A., Judde, J.-G., Bosio, A., Cairo, S., and Hardt, O. (2015) The sialyl-glycolipid stage-specific embryonic antigen 4 marks a subpopulation of chemotherapy-resistant breast cancer cells with mesenchymal features. *Breast Cancer Res.* **17**, 146
17. Parte, S. C., Batra, S. K., and Kakar, S. S. (2018) Characterization of stem cell and cancer stem cell populations in ovary and ovarian tumors. *J Ovarian Res.* **11**, 69
18. Sivasubramaniyan, K., Harichandan, A., Schilbach, K., Mack, A. F., Bedke, J., Stenzl, A., Kanz, L., Niederfellner, G., and Bühring, H.-J. (2015) Expression of stage-specific embryonic antigen-

4 (SSEA-4) defines spontaneous loss of epithelial phenotype in human solid tumor cells. *Glycobiology* **25**, 902-917
19. Lou, Y.-W., Wang, P.-Y., Yeh, S.-C., Chuang, P.-K., Li, S.-T., Wu, C.-Y., Khoo, K.-H., Hsiao, M., Hsu, T.-L., and Wong, C.-H. (2014) Stage-specific embryonic antigen-4 as a potential therapeutic target in glioblastoma multiforme and other cancers. *Proc. Natl. Acad. Sci. USA* **111**, 2482
20. Kannagi, R., Cochran, N. A., shigami, F., Hakomori, S., Andrews, P. W., Knowles, B. B., and Solter, D. (1983) Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. *EMBO J.* **2**, 2355-2361
21. Fan, Z. C., Goldsteen, B. Z., Guddat, L. W., Thakur, A., Landolfi, N. F., Co, M. S., Vasquez, M., Queen, C., Ramsland, P. A., and Edmundson, A. B. (1999) Comparison of the three-dimensional structures of a humanized and a chimeric Fab of an anti-gamma-interferon antibody. *J. Mol. Recognit.* **12**, 19-32
22. Hudis, C. A. (2007) Trastuzumab — Mechanism of Action and Use in Clinical Practice. *N Engl J Med* **357**, 39-51
23. Galizia, G., Lieto, E., De Vita, F., Orditura, M., Castellano, P., Troiani, T., Imperatore, V., and Ciardiello, F. (2007) Cetuximab, a chimeric human mouse anti-epidermal growth factor receptor monoclonal antibody, in the treatment of human colorectal cancer. *Oncogene* **26**, 3654
24. Dhillon, S. (2015) Dinutuximab: first global approval. *Drugs* **75**, 923-927
25. Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J. C., Akerley, W., van den Eertwegh, A. J. M., Lutzky, J., Lorigan, P., Vaubel, J. M., Linette, G. P., Hogg, D., Ottensmeier, C. H., Lebbé, C., Peschel, C., Quirt, I., Clark, J. I., Wolchok, J. D., Weber, J. S., Tian, J., Yellin, M. J., Nichol, G. M., Hoos, A., and Urba, W. J. (2010) Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N Engl J Med* **363**, 711-723
26. Markham, A. (2016) Atezolizumab: First Global Approval. *Drugs* **76**, 1227-1232
27. The Antibody Society. (2019) Approved Antibodies: Therapeutic monoclonal antibodies approved or in review in the EU or the US.
28. Wang, S., and Jia, M. (2016) Antibody Therapies in Cancer. in *Progress in Cancer Immunotherapy* (Zhang, S. ed.), Springer Netherlands, Dordrecht. pp 1-67
29. Pento, J. T. (2017) Monoclonal Antibodies for the Treatment of Cancer. *Anticancer Res.* **37**, 5935-5939
30. Horwacik, I., Golik, P., Grudnik, P., Kolinski, M., Zdzalik, M., Rokita, H., and Dubin, G. (2015) Structural Basis of GD2 Ganglioside and Mimetic Peptide Recognition by 14G2a Antibody. *Mol. Cell. Proteomics* **14**, 2577-2590
31. Kaminski, M. J., MacKenzie, C. R., Mooibroek, M. J., Dahms, T. E., Hirama, T., Houghton, A. N., Chapman, P. B., and Evans, S. V. (1999) The role of homophilic binding in anti-tumor antibody R24 recognition of molecular surfaces. Demonstration of an intermolecular beta-sheet interaction between vh domains. *J. Biol. Chem.* **274**, 5597-5604
32. Talavera, A., Eriksson, A., Okvist, M., Lopez-Requena, A., Fernandez-Marrero, Y., Perez, R., Moreno, E., and Krengel, U. (2009) Crystal structure of an anti-ganglioside antibody, and modelling of the functional mimicry of its NeuGc-GM3 antigen by an anti-idiotypic antibody. *Mol. Immunol.* **46**, 3466-3475
33. Krengel, U., Olsson, L. L., Martinez, C., Talavera, A., Rojas, G., Mier, E., Angstrom, J., and Moreno, E. (2004) Structure and molecular interactions of a unique antitumor antibody specific for N-glycolyl GM3. *J. Biol. Chem.* **279**, 5597-5603
34. Ahmed, M., Goldgur, Y., Hu, J., Guo, H. F., and Cheung, N. K. (2013) In silico driven redesign of a clinically relevant antibody for the treatment of GD2 positive tumors. *PLoS One* **8**, e63359
35. Pichla, S. L., Murali, R., and Burnett, R. M. (1997) The crystal structure of a Fab fragment to the melanoma-associated GD2 ganglioside. *J. Struct. Biol.* **119**, 6-16
36. Soliman, C., Walduck, A. K., Yuriev, E., Richards, J. S., Cywes-Bentley, C., Pier, G. B., and Ramsland, P. A. (2018) Structural basis for antibody targeting of the broadly expressed microbial polysaccharide poly-N-acetyl glucosamine. *J. Biol. Chem.* **293**, 5079-5089
37. Copoiu, L., Torres, P. H. M., Ascher, D. B., Blundell, T. L., and Malhotra, S. (2019) ProCarbDB: a database of carbohydrate-binding proteins. *Nucleic Acids Res.*, gkz860

38. Dhar, C., Sasmal, A., and Varki, A. (2019) From "Serum Sickness" to "Xenosialitis": Past, Present, and Future Significance of the Non-human Sialic Acid Neu5Gc. *Front. Immunol.* **10**, 807
39. Parameswaran, R., Lim, M., Arutyunyan, A., Abdel-Azim, H., Hurtz, C., Lau, K., Muschen, M., Yu, R. K., von Itzstein, M., Heisterkamp, N., and Groffen, J. (2013) O-acetylated N-acetylneuraminic acid as a novel target for therapy in human pre-B acute lymphoblastic leukemia. *J. Exp. Med.* **210**, 805-819
40. Chua, J. X., Vankemmelbeke, M., McIntosh, R. S., Clarke, P. A., Moss, R., Parsons, T., Spendlove, I., Zaitoun, A. M., Madhusudan, S., and Durrant, L. G. (2015) Monoclonal Antibodies Targeting LecLex-Related Glycans with Potent Antitumor Activity. *Clin. Cancer. Res.* **21(13)**, 2963-2974
41. Duncan, T. J., Rolland, P., Deen, S., Scott, I. V., Liu, D. T., Spendlove, I., and Durrant, L. G. (2007) Loss of IFN gamma receptor is an independent prognostic factor in ovarian cancer. *Clin. Cancer. Res.* **13**, 4139-4145
42. Mahmoud, S. M., Paish, E. C., Powe, D. G., Macmillan, R. D., Grainge, M. J., Lee, A. H., Ellis, I. O., and Green, A. R. (2011) Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J. Clin. Oncol.* **29**, 1949-1955
43. Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Noll, A., Tampe, R., Brandstetter, S., Forster, A., Mueller, M., Schulze-Briese, C., Bunk, O., and Wang, M. (2016) EIGER detector: application in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **72(Pt 9)**, 1036–1048
44. Aragao, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N., Ericsson, D. J., Gee, C., Macedo, S., Mudie, N., Panjikar, S., Price, J. R., Riboldi-Tunnicliffe, A., Rostan, R., Williamson, R., and Caradoc-Davies, T. T. (2018) MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J Synchrotron Radiat* **1**, 885-891
45. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242
46. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132
47. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221
48. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501
49. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255
50. Agirre, J., Iglesias-Fernandez, J., Rovira, C., Davies, G. J., Wilson, K. S., and Cowtan, K. D. (2015) Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833-834
51. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778-2786
52. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* **91**, 43-56
53. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701-1718
54. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958
55. Sousa da Silva, A. W., and Vranken, W. F. (2012) ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **5**, 367

56. Bernardi, A., Faller, R., Reith, D., and Kirschner, K. N. (2019) ACPYPE update for nonuniform 1–4 scale factors: Conversion of the GLYCAM06 force field from AMBER to GROMACS. *SoftwareX* **10**, 100241
57. Hess, B. (2008) P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **4**, 116-122

ABBREVIATIONS

BGC, glucose; CDR, complementarity-determining region; Fab, fragment antigen-binding; Fv, variable fragment; GAL, β-galactose; GLA, α-galactose; GSLs, glycosphingolipids; HPAEC-PAD, High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection; HRPO, horseradish peroxidase; mAb, monoclonal antibody; MD, molecular dynamics; NGA, N-acetyl-galactosamine; NMR, nuclear magnetic resonance; PEG, polyethylene glycol; RMSF, root-mean-square fluctuation; SIA, sialic acid; SLe^x, sialyl-Lewis x; SSEA, stage-specific embryonic antigen; SSEA-3, stage-specific embryonic antigen-3; SSEA-4, stage-specific embryonic antigen-4; ST3GAL2, β-galactoside α2,3-sialyltransferase 2; TLC, thin layer chromatography; TMB, 3,3',5,5'-Tetramethylbenzidine.

FIGURES AND FIGURE LEGENDS

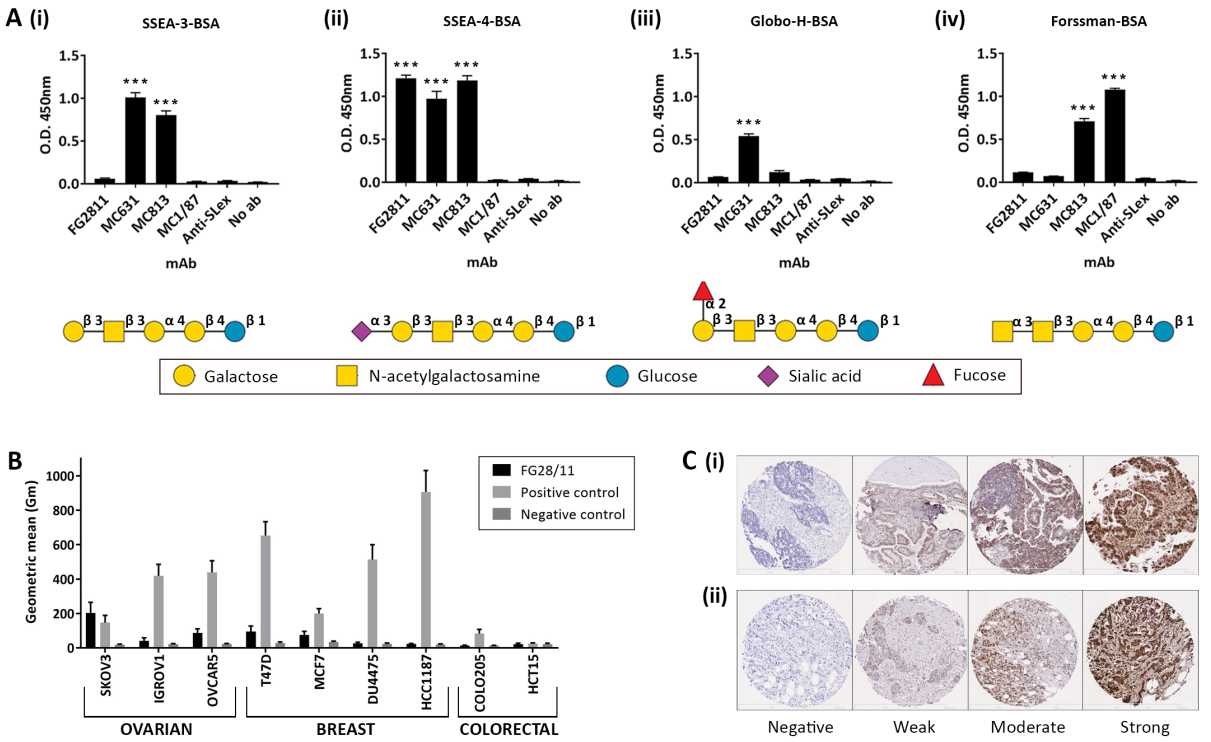

Figure 1. Assessment of FG28/11 specificity towards SSEA-4. **A)** Binding of FG28/11 mAb to **(i)** SSEA-3-BSA, **(ii)** SSEA-4-BSA, **(iii)** Globo-H-BSA and **(iv)** Forssman-BSA glycans was assessed by ELISA. MC631, MC813-70 and M1/87 were included as positive control mAbs for SSEA-3, SSEA-4 and Forssman glycans, respectively. Anti-sialyl-Lewis x antibody and biotinylated anti-mouse IgG secondary antibody (no ab) were used as negative controls. Antibody activity was measured by absorbance at 450nm. Error bars represent mean ± standard deviation of quadruplicate wells (*** $p <$ 0.0001 versus control, ANOVA followed by Bonferroni multiple comparisons test). The cartoon representation of each glycan (in SNFG nomenclature) is shown below the corresponding panel, along with a key for the monosaccharides. **B)** A panel of cancer cell lines were stained with FG28/11 mAb (5µg/ml) and analysed by flow cytometry. The anti-HLA-A,B,C mAb W6/32 (1µg/ml) and no primary mAb were included as positive and negative controls, respectively. Results were expressed as Geometric mean (Gm) with % CV. **C)** Binding of FG28/11 (1µg/ml) antibody to **(i)** ovarian and **(ii)** breast tumors showing a range of negative to strongly positive immunostaining results.

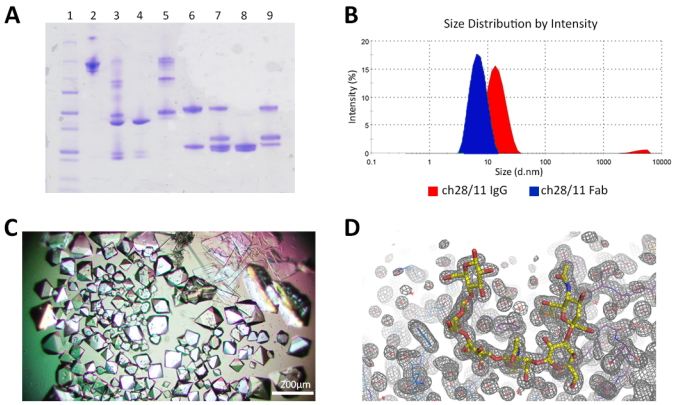

Figure 2. Production and crystallization of ch28/11 Fab. **A)** Coomassie stained SDS-PAGE gel. Lanes correspond to: 1, protein marker; 2 and 6, ch28/11 IgG; 3, and 7, antibody digest; 4 and 8, purified Fab; 5 and 9, undigested IgG and Fc fraction. Samples in lanes 6-9 were reduced using β-mercaptoethanol. **B)** Size distribution analysis of DLS data for ch28/11 IgG and ch28/11 Fab. **C)** Co-crystals of ch28/11 Fab in complex with SSEA-4 ($P4_12_12$ crystal form; scale bar, 200 microns). **D)** Composite omit $2F_o - F_c$ map (displayed at 1.5σ level) for the 1.5 Å resolution ch28/11 Fab:SSEA-4 complex. Residues from Fab are depicted as lines with L chain in blue and H chain in purple, with the ligand displayed as sticks colored by atom type (C, yellow; O, red; N, blue).

Table 1. Data collection and crystallographic refinement statistics. Values in parentheses refer to the highest resolution shell for each data set. Collection statistics were compiled from XDS, and refinement statistics from Phenix and CCP4, using MolProbity for Ramachandran statistics.

Parameter	ch28/11 Fab:SSEA-4	ch28/11 Fab:SSEA-4	ch28/11 Fab:SSEA-4	ch28/11 Fab
Data collection				
Space group	$P4_12_12$	$P2_1$	$P6_2$	$P2_1$
Unit cell dimensions (Å)	a = 67.8, b = 67.8, c = 234.0	a = 37.5, b = 69.1, c = 97.7	a = 170.1, b = 170.1, c = 95.4	a = 75.2, b = 69.6, c = 93.9
Unit cell angles (°)	α = 90.0, β = 90.0, γ = 90.0	α = 90.0, β = 94.5, γ = 90.0	α = 90.0, β = 90.0, γ = 120.0	α = 90.0, β = 98.3, γ = 90.0
Resolution range (Å)	50 – 1.5 (1.6 – 1.5)	50 – 1.9 (2.0 – 1.9)	50 – 2.7 (2.8 – 2.7)	50 – 2.5 (2.6 – 2.5)
Number of unique reflections	84916 (13419)	39580 (6187)	41405 (6561)	33478 (5126)
Data completeness (%)	99.9 (99.2)	99.3 (96.8)	99.7 (98.6)	98.1 (93.9)
Average multiplicity	13.3 (13.4)	3.4 (3.4)	5.3 (5.2)	3.4 (3.5)
R-factor	0.10 (0.46)	0.06 (0.48)	0.13 (0.69)	0.06 (0.41)
R_{meas}	0.10 (0.48)	0.07 (0.57)	0.14 (0.77)	0.07 (0.48)
Mean I/σ (I)	15.8 (4.2)	10.1 (1.9)	10.6 (2.3)	14.4 (2.6)
Crystallographic refinement				
R_{work}	0.172	0.176	0.169	0.164
R_{free}	0.197	0.230	0.246	0.239
Average B-factor from Wilson plot (Å^2)	19.5	34.4	52.7	52.1
rmsd from ideal values				
Bond lengths (Å)	0.018	0.018	0.012	0.014
Bond angles (^{O})	1.9	1.9	1.7	1.7
Ramachandran plot values (%)				
Favored regions	98.6	97.9	96.2	96.8
Allowed regions	1.2	2.1	3.4	3.0
Outliers	0.2	0	0.4	0.2
Average B-factor (Å^2)				
Protein atoms	23.1	41.6	54.9	58.8
Carbohydrate atoms	26.4	38.3	61.2	-
Water	37.2	47.3	40.9	51.0

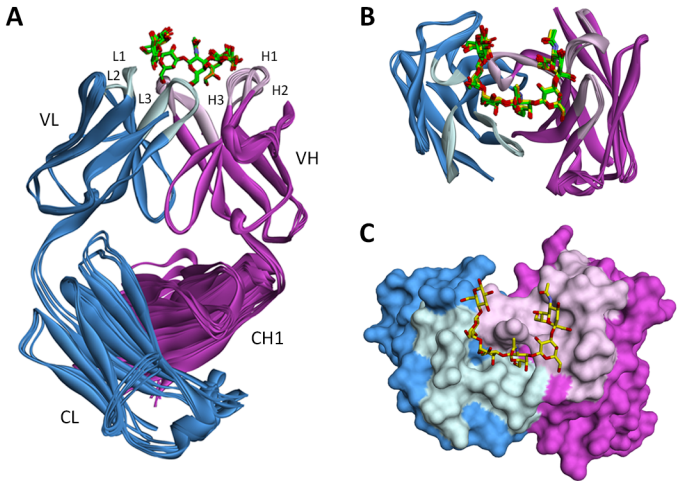

Figure 3. Overview of the crystal structures of ch28/11 Fab bound to SSEA-4. **A)** Overlay of the five complexes, with the Fab shown as ribbons with the L chain in blue and the H chain in magenta. **B)** Overlay of the five ch28/11 Fab:SSEA-4 complexes looking down into the binding site (end-on view). For clarity only the VL and VH domains of each Fab are shown. CDRs are shown in pale blue (L chain) and pale pink (H chain). SSEA-4 glycan ligands are shown with carbon atoms in yellow ($P4_12_12$ crystal form), orange ($P2_1$ crystal form) and green ($P6_2$ crystal form). **C)** End-on solvent-accessible view of the 1.5 Å resolution ch28/11 Fab:SSEA-4 complex (VL, blue; VH, magenta; L chain CDRs, pale blue; H chain CDRs, pale pink). SSEA-4 is depicted as sticks colored by atom type (C, yellow; O, red; N, blue).

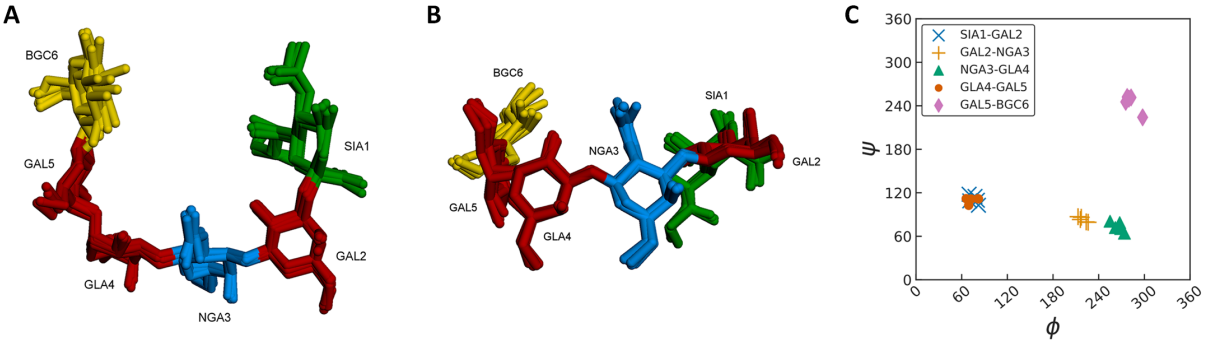

Figure 4. Overlay of SSEA-4 from the five ch28/11 Fab crystal structures. **A)** Top view. **B)** Side view (rotated 90 degrees). The ligands are depicted as sticks colored by sugar type, with glucose (BGC) in yellow, galactose (GAL and GLA) in red, N-acetyl-galactosamine (NGA) in blue and sialic acid (SIA) in green. **C)** Sugar residue conformations of SSEA-4 bound to ch28/11 in the five crystal structures, plotted by *phi* (ϕ) vs. *psi* (ψ) angles.

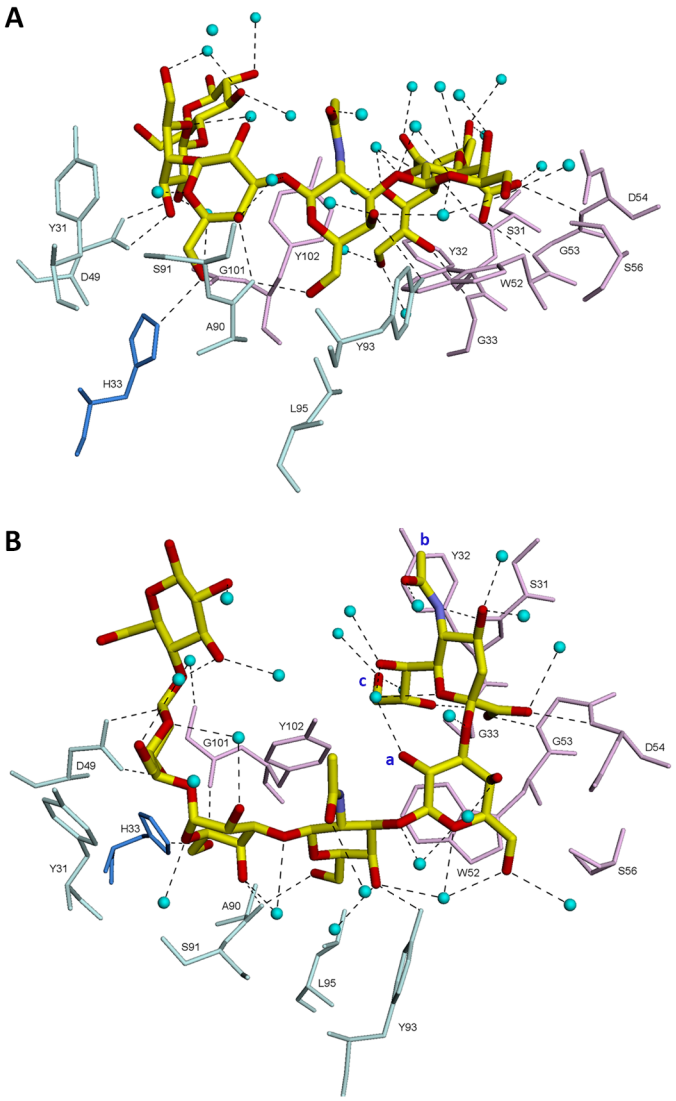

Figure 5. Details of the interactions between ch28/11 Fab and SSEA-4 from the highest resolution crystal structure. **A)** Side view and **B)** Top view (rotated 90 degrees). Residues from Fab are depicted as thin sticks: L chain CDR residues, pale blue; H chain CDR residues, pale pink; L chain framework residue, blue. Glycan ligands are depicted as thick sticks colored by atom type (C, yellow; O, red; N, blue). From left to right, glycans are BGC6, GAL5, GLA4, NGA3, GAL2, SIA1 (see Fig 4). Amino acids are labeled (1 letter code) and sequential numbering is shown (see Fig. S4 for IMGT numbering scheme). Hydrogen bonds (dashed black lines) are displayed between the SSEA-4 hexasaccharide and Fab or water molecules (light blue spheres). Labels a, b and c (shown in blue) correspond to the O2 position of GAL2, the C11 position of SIA1 and the O9 position of SIA1 respectively.

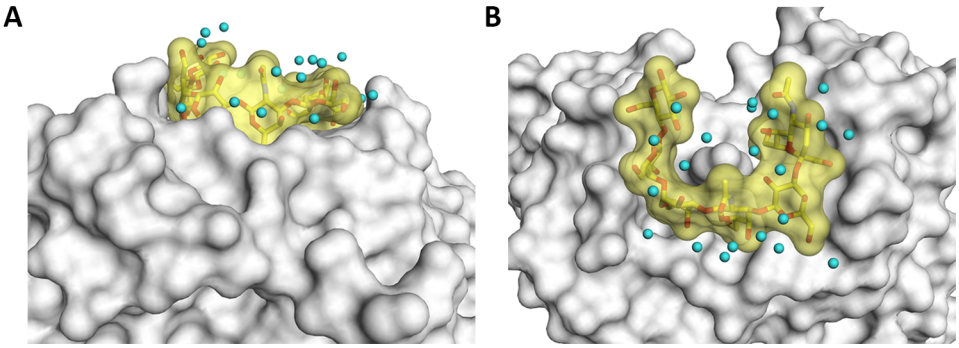

Figure 6. Surface representation of the ch28/11 Fab:SSEA-4 complex. **A)** Side view. **B)** Top view (rotated 90 degrees). Solvent-accessible surfaces for the ch28/11 Fab (white) and SSEA-4 (yellow, transparent) are shown, with water molecules displayed as light blue spheres. The SSEA-4 ligand completely fills the binding groove.

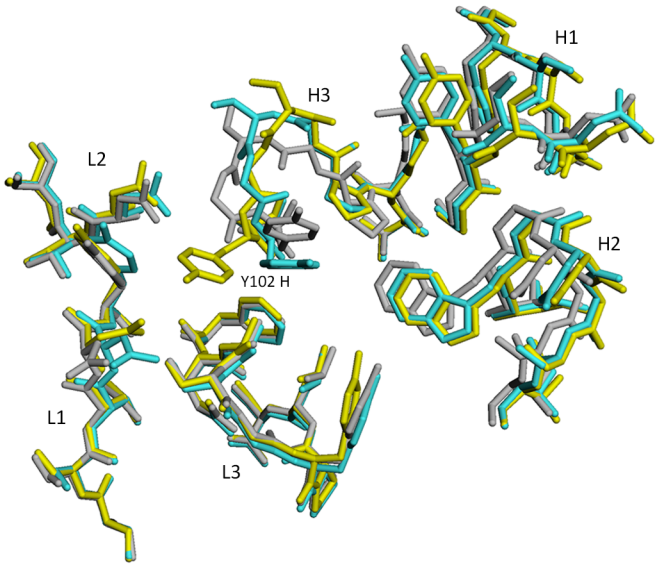

Figure 7. The unliganded ch28/11 Fab structure compared to the 1.5 Å resolution ch28/11 Fab bound to SSEA-4. CDR residues from ch28/11 Fab:SSEA-4 (light grey sticks) with ch28/11 Fab 1 (yellow sticks) and ch28/11 Fab 2 (light blue sticks). Differences between free and bound Fabs are more prominent in the heavy chain CDRs, particularly H3.

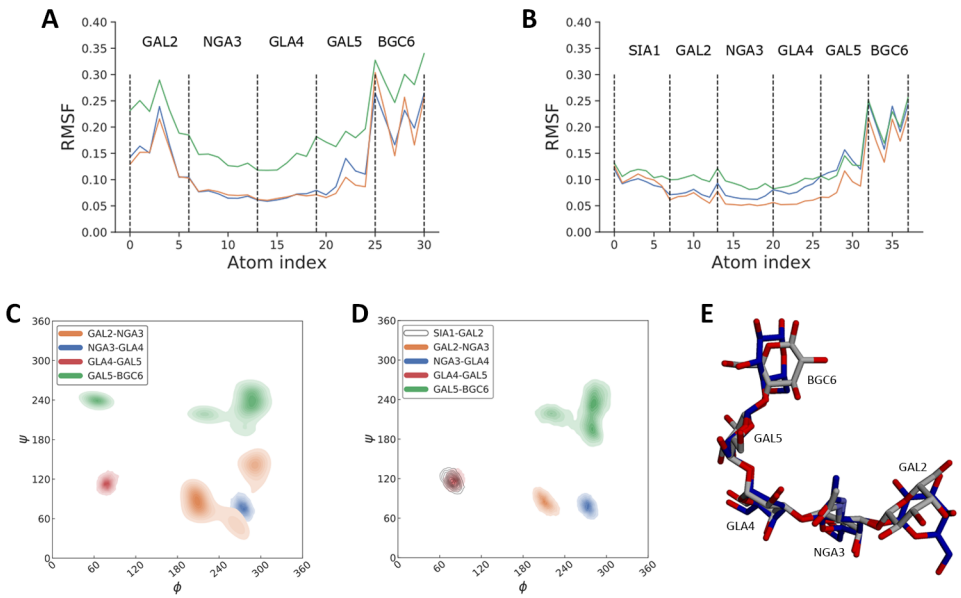

Figure 8. MD simulations of ch28/11 Fab complexes with SSEA-3 and SSEA-4. RMSF plots of glycan atoms with residues indicated by dashed lines for **A)** SSEA-3 and **B)** SSEA-4. MD simulations (500 ns) were performed in triplicate, averaged over time, and are represented as green, orange and blue traces. *Phi* (ϕ) vs. *psi* (ψ) density plots of each glycosidic linkage of **C)** SSEA-3 and **D)** SSEA-4. Angles were averaged over time for the 3 simulations. **E)** The 2 main conformations of SSEA-3, represented as sticks with carbon atoms in grey and dark blue.