nature communications

Article

https://doi.org/10.1038/s41467-024-53746-9

Structural insights into epitope-paratope interactions of a monoclonal antibody targeting CEACAM5-expressing tumors

Received: 4 August 2023

Accepted: 18 October 2024

Published online: 30 October 2024

Check for updates

Anand Kumar[1], Francis Duffieux[2], Marie Gagnaire[2], Chiara Rapisarda[1], Thomas Bertrand[1] & Alexey Rak[1]

Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) are overexpressed in some tumor types. The antibody-drug conjugate tusamitamab ravtansine specifically recognizes the A3-B3 domains of human CEACAM5 (hCEACAM5). To understand this specificity, here we map the epitope-paratope interface between the A3-B3 domains of hCEACAM5 ($hCEACAM5_{A3-B3}$) and the antigen-binding fragment of tusamitamab (tusa Fab). We use hydrogen/deuterium exchange mass spectrometry to identify the tusa Fab paratope, which involves heavy chain (HC) residues 101–109 and light chain residues 48–54 and 88–104. Using surface plasmon resonance, we demonstrate that alanine variants of HC residues 96–108 abolish binding to hCEACAM5, suggesting that these residues are critical for tusa-Fab–antigen complex formation. The cryogenic electron microscopy structure of the $hCEACAM5_{A3-B3}$- tusa Fab complex (3.11 Å overall resolution) reveals a discontinuous epitope involving residues in the A3-B3 domains and an N-linked mannose at residue Asn612. Conformational constraints on the epitope-paratope interface enable tusamitamab to target $hCEACAM5_{A3-B3}$ and distinguish CEACAM5 from other CEACAMs.

Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs), a family of membrane-associated cell-surface glycoproteins that are involved in cell adhesion and signaling and tumor metastasis and upregulated in some tumor cells of lung, colorectal, and gastric tissue, are attractive targets for antibody-drug conjugates[1–3]. However, the extracellular domain (ECD) of CEACAMs, which consists of repeated immunoglobulin (Ig)-like domains, including an N-terminal immunoglobulin variable (IgV)-like domain and up to 6 immunoglobulin constant C2 (IgC2)-like domains (types A or B)[1,2,4], has a high degree of sequence similarity and structural homology among various CEACAMs[3]. In addition, the ECD of CEACAMs has up to 28 potential N-glycosylation sites and is heavily glycosylated[1,5]. Differences in glycosylation affect protein-protein interactions and disease progression, which is consistent with the role of CEACAMs in metastasis[6–9]. As a result, CEACAM-targeted antibodies must be highly specific to minimize cross-reactivity with other CEACAMs, which may lead to false positives in diagnostic assays or to off-target or on-target off-tumor toxicity[3,10,11].

The CEACAM5-directed antibody-drug conjugate tusamitamab ravtansine (SAR408701) was investigated for the treatment of advanced nonsquamous non-small cell lung cancer (NSCLC) and other solid tumors[12–14]. The antibody component, tusamitamab, selectively targets the A3-B3 domains of CEACAM5 and does not bind to CEACAM1, CEACAM6, or CEACAM8[3], which implies the presence of unique binding regions on the surface of these two domains that confer antibody recognition and specificity. Understanding the mechanism of binding and the structural determinants of this specificity requires identification of the regions of the antibody (tusamitamab) and antigen (CEACAM5) that form and/or stabilize the paratope-epitope interface.

A low-resolution X-ray and neutron solution scattering model of the ECD of CEACAM5 showed that it is shaped like a bottle brush, with

[1]Integrated Drug Discovery, Sanofi R&D, Paris, France. [2]Large Molecules Research, Sanofi R&D, Paris, France. e-mail: Alexey.Rak@sanofi.com

the bristles formed by the Ig-like domains, each of which consists of two β sheets[1,15]. This structure was supported by subsequent homology modeling based on the crystal structure of another member of the Ig superfamily, cluster of differentiation 2 (CD2), and partially with the crystal structures of the N-terminal IgV-like domain of CEACAM5[16,17]. However, the structures of the A and B IgC2-like domains have not been elucidated.

Here, we describe the high-resolution structure of the A3-B3 domains of CEACAM5 in complex with the antigen-binding fragment of tusamitamab (hereafter referred to as "tusa Fab") by cryogenic electron microscopy (cryo-EM). Using hydrogen-deuterium exchange coupled with mass spectrometry (HDX-MS) and surface plasmon resonance (SPR) analysis of alanine variants, we further characterize the epitope-paratope interface between the A3-B3 domains of human CEACAM5 (hCEACAM5$_{A3\text{-}B3}$) and tusa Fab.

Results

Cryo-EM structure of the hCEACAM5$_{A3\text{-}B3}$-tusa Fab complex

The structure of the hCEACAM5$_{A3\text{-}B3}$-tusa Fab complex (Fig. 1a) was solved by single-particle cryo-EM at an overall resolution of 3.11 Å and a local resolution at the epitope-paratope interface of <2.8 Å (Supplementary Fig. 1 and Supplementary Table 1). The MolProbity score was 1.89, and 94.9% of amino acids were in Ramachandran favored conformations (Supplementary Table 1).

The structures of both the tusa Fab and hCEACAM5$_{A3\text{-}B3}$ consist entirely of β strands and sheets (Fig. 1b). As expected, the A3 and B3 domains of hCEACAM5 have a β-sandwich structure typical of the IgC2-set domain (InterPro domain IPR008424)[18]. Seven glycans were identified in the hCEACAM5$_{A3\text{-}B3}$ region, including an N-linked mannose at residue Asn612 that was hydrogen bonded to the tusa Fab heavy chain residue Gly54 and Gly56 (Fig. 1c). Surface exposure of glycans and the hCEACAM5-tusa Fab complex is shown in Fig. 1d.

The hCEACAM5-tusa Fab structure revealed an epitope across the surface of hCEACAM5$_{A3\text{-}B3}$ (mostly in B3) that binds both heavy and light chains of tusa Fab. The buried surface area of the hCEACAM5-tusa Fab heavy chain interface covered 455 Å2 involving 12 residues in the hCEACAM5 B3 domain (Pro621 to Ser626, Gln635 to Val639, and Phe641) and 14 tusa Fab heavy chain residues (Val28, Ser30 to Asp33, Ser53, and His99 to Pro106). A network of 8 hydrogen bonds was detected within this interface, with several amino acids contributing multiple hydrogen bonds. In particular, tusa Fab heavy chain Ser31 was hydrogen bonded with 3 hCEACAM5 residues (Gln635, Thr637, and Val639). Similarly, tusa Fab heavy chain Ser104 was hydrogen bonded to both hCEACAM5 Ser622 and Gln624, and the CEACAM5 Ser622 was hydrogen bonded to 2 tusa Fab heavy chain residues (Ser103 and Ser104). Although several residues in the interface had positive free energies of solvation (ΔG_{solv}, Supplementary Table 2), indicating a hydrophobic effect, most were too far apart to interact (>5 Å). Only 1 potential hydrophobic interaction in the interface was identified, that between hCEACAM5 Phe641 and tusa Fab heavy chain Val28, residues which are ~4.2 Å apart.

The buried surface area of the hCEACAM5-tusa Fab light chain interface covered 344 Å2 involving 12 hCEACAM5 amino acids (A3 domain: Asn509, Lys511, Val513, Asp588, and Leu590; B3 domain: Pro621, Ser622, Gln624, His636, Asn659, Leu660, and Ala661) and 9 tusa Fab light chain amino acids (Glu27, Asn28, Phe30, Ser31, Tyr32, Tyr49, Asn50, His91, and Tyr92). Three hydrogen bonds were detected between hCEACAM5 residues Leu660, Lys511, and Gln624 and tusa Fab light chain residues Tyr32, Tyr92, and Asn50, respectively. In addition, the interface contained a small hydrophobic region, involving 4 hCEACAM5 residues (Pro621, Leu590, Leu660, and Ala661) and 3 tusa Fab light chain residues (Phe30, Tyr32, and Tyr92), with positive ΔG_{solv} (Supplementary Table 2) at distances between 3.7 Å and 4.7 Å.

Notably, Gln624 in the hCEACAM5 B3 domain was hydrogen bonded with both tusa Fab heavy and light chain residues. No salt bridges or disulfide bonds were involved in epitope-paratope interactions between hCEACAM5$_{A3\text{-}B3}$ and tusa Fab (Fig. 2).

Furthermore, multiple sequence alignment revealed 61%–80% identity between the ECDs of CEACAM5 and CEACAMs 1, 6, and 8; Supplementary Fig. 2), and 72% and 67% identity for A1-B1 and A2-B2 domains in hCEACAM5, respectively (Supplementary Fig. 3).

Critical epitope and paratope residues identified using SPR

The binding of tusa Fab plus the fragment crystallizable [Fc] region of the CEACAM5-directed antibody (tusa Fab+Fc) to human CEACAM5 extracellular domain (hCEACAM5$_{ECD}$) was investigated in 2 SPR experiments. In the initial screening, the impact of tusa Fab+Fc alanine substitutions on hCEACAM5$_{ECD}$ binding (the percentage of the maximum observed binding signal [%R_{max}]; Supplementary Fig. 4a) and on the stability of the hCEACAM5$_{ECD}$-tusa Fab+Fc complex (the percentage of binding signal remaining after 50 seconds of dissociation [%Remaining]; Supplementary Fig. 4b) was measured for 27 heavy chain and 18 light chain variants. Five variants in the heavy chain (Asp33Ala, Phe101Ala, Gly102Ala, Gly105Ala, and Pro106Ala) and 2 in the light chain (Tyr32Ala and Asn50Ala) exhibited complete loss of binding to hCEACAM5$_{ECD}$; 4 variants in the heavy chain (Tyr32Ala, Tyr50Ala, His99Ala, and Tyr100Ala) and 4 in the light chain (Phe30Ala, His91Ala, Tyr92Ala, and Pro95Ala) showed partial loss of binding to hCEACAM5. In addition, 7 variants in the heavy chain (Tyr32Ala, Asp33Ala, Tyr100Ala, Phe101Ala, Gly102Ala, Gly105Ala, and Pro106Ala) and 2 in the light chain (Tyr32Ala and Asn50Ala) completely destabilized the hCEACAM5-tusa Fab+Fc complex. Two other variants in the heavy chain (Tyr50Ala and His99Ala) and 4 in the light chain (Phe30Ala, His91Ala, Tyr92Ala, and Pro95Ala) resulted in partial loss of complex stability. Together, these findings suggest that the tusa Fab paratope encompasses the third complementarity-determining region (CDR3) of the heavy chain (residues 96–108). A few residues from heavy chain CDR1 (residues 32 and 33), light chain CDR1 (residue 32), and light chain CDR3 (residues 91 and 92) may also contribute to paratope interactions.

In the second set of SPR experiments, 20 tusa Fab+Fc variants, including those with %R_{max} and %Remaining that differed from wild-type (WT) tusa Fab+Fc in the SPR screening experiments, were selected for further evaluation. When tested at higher concentrations, the tusa Fab+Fc alanine variants exhibited %R_{max} and %Remaining patterns (Supplementary Fig. 5) similar to those in the initial screening. Kinetic analyses determined the binding affinity and dissociation kinetics of tusa Fab+Fc WT and alanine variants binding to hCEACAM5$_{ECD}$ (Table 1 and Supplementary Fig. 6). Of note, 2 variants in the heavy chain (Tyr50Ala and His99Ala) and 4 in the light chain (Phe30Ala, His91Ala, Tyr92Ala, and Pro95Ala) had reduced binding affinity for hCEACAM5$_{ECD}$ (K_D = 306.5 nM to 13.5 μM) relative to WT tusa Fab+Fc (K_D = 14.0 nM) (Table 1).

Critical epitope and paratope residues identified by HDX-MS

We identified 25 peptides in hCEACAM5, 30 in the tusa Fab heavy chain, and 20 in the tusa Fab light chain, which covered 89%, 77%, and 68% of the respective sequences (Supplementary Fig. 7). The average peptide length was 14.6, 12.7, and 12.0 amino acids, respectively.

Six of the 25 CEACAM5 peptides showed a difference in HDX (ΔHDX) between deuterated free CEACAM5 and complexed CEACAM5-tusa Fab. Peptides spanning residues 606–613 and 626–641 showed significant ΔHDX after 2 minutes of deuteration at 4 °C, which suggested that the CEACAM5 epitope was localized to these regions. Within the latter region, 4 peptides (residues 626–640, 628–640, 628–641, and 638–641) had a maximum absolute reduction in deuterium uptake of 1.6 ± 0.2 Da (15% ± 2% difference) between free CEACAM5 and complexed CEACAM5-tusa Fab.

The epitope boundaries were refined by comparing the deuterium uptake in overlapping peptides. After 2 minutes of deuteration, deuterium uptake did not show a significant change in the peptides

https://doi.org/10.1038/s41467-024-53746-9

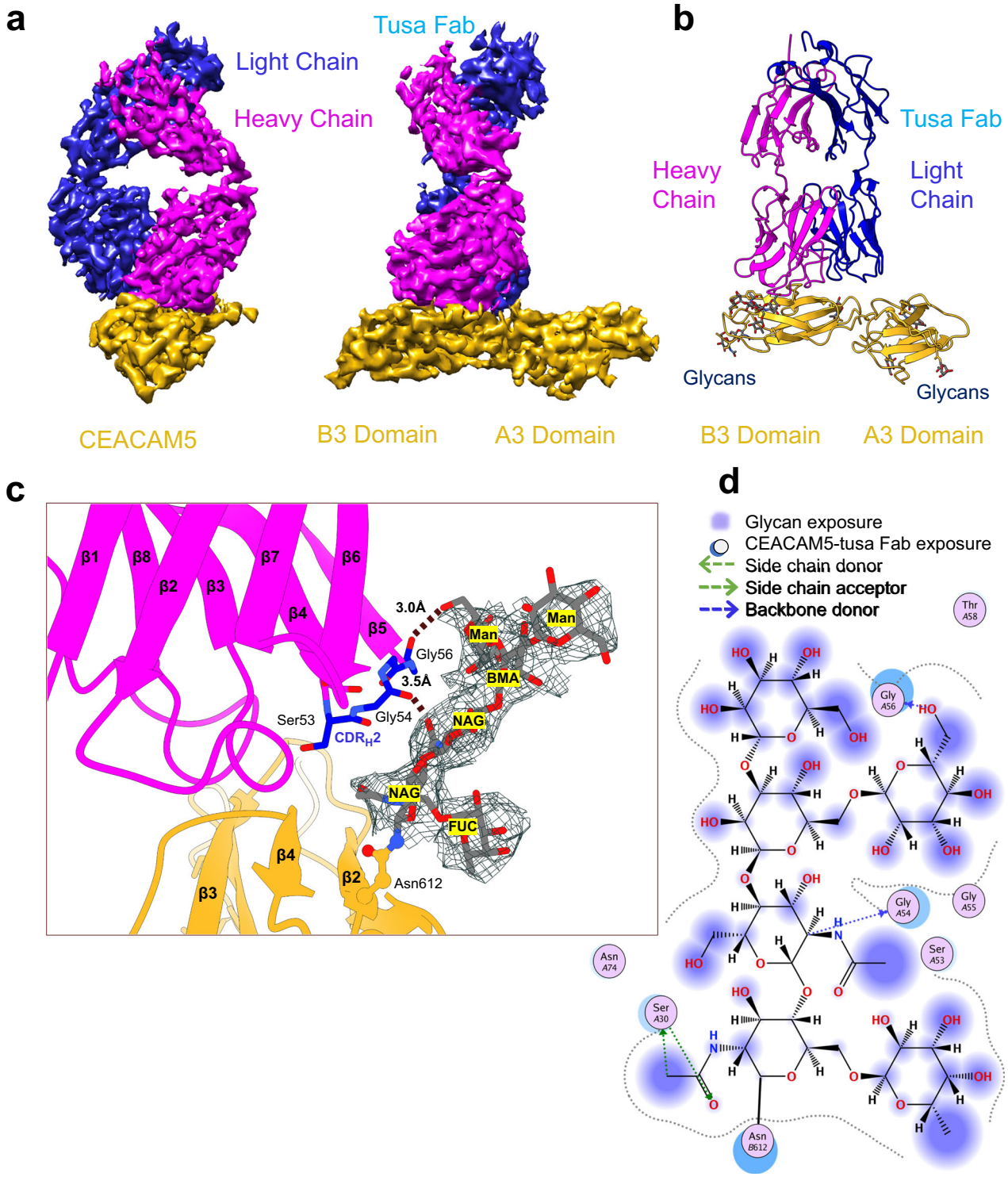

Fig. 1 | Single-particle cryo-EM structure of the hCEACAM5$_{A3\text{-}B3}$ in complex with tusa Fab, the Fab fragment of a humanized CEACAM5-directed monoclonal antibody. a Cryo-EM map of hCEACAM5$_{A3\text{-}B3}$ complexed with tusa Fab. The map on the right is rotated by 90° on the vertical axis relative to that on the left. **b** Ribbon representation of hCEACAM5$_{A3\text{-}B3}$ (gold ribbons) complexed with tusa Fab (magenta and blue ribbons). Each arrow represents a β strand. N-linked glycans are shown as ball-and-stick representations. **c** Close-up view of interactions between the N-linked mannose at residue Asn612 in the hCEACAM5 B3 domain (gold ribbons) and the tusa Fab heavy chain (magenta ribbons). Interactions between Gly54 and Gly56 from the CDR$_H$2 (blue) of the tusa Fab heavy chain and glycans (dark gray) are shown with the dotted line. The wire mesh around the glycan moiety shows the fit of the derived model to the cryo-EM density map. **d** Ligand interaction map of glycans generated by Molecular Operating Environment (MOE), 2022.02 Chemical Computing Group ULC, 910–1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada. Interacting residues from hCEACAM5 are shown in circles with A and B designating residues from the A3 and B3 domains, respectively. BMA β-d-mannose, CDR$_H$2 complementarity-determining region 2 of heavy chain, CEACAM5 carcinoembryonic antigen-related cell adhesion molecule 5, cryo-EM cryogenic electron microscopy, FUC α-L-fructose, hCEACAM5 human carcinoembryonic antigen-related cell adhesion molecule 5, hCEACAM5$_{A3\text{-}B3}$ A3-B3 domains of human carcinoembryonic antigen-related cell adhesion molecule 5, Man α-D-mannose, NAG *N*-acetyl-β-D-glucosamine, tusa Fab antigen-binding fragment of tusamitamab.

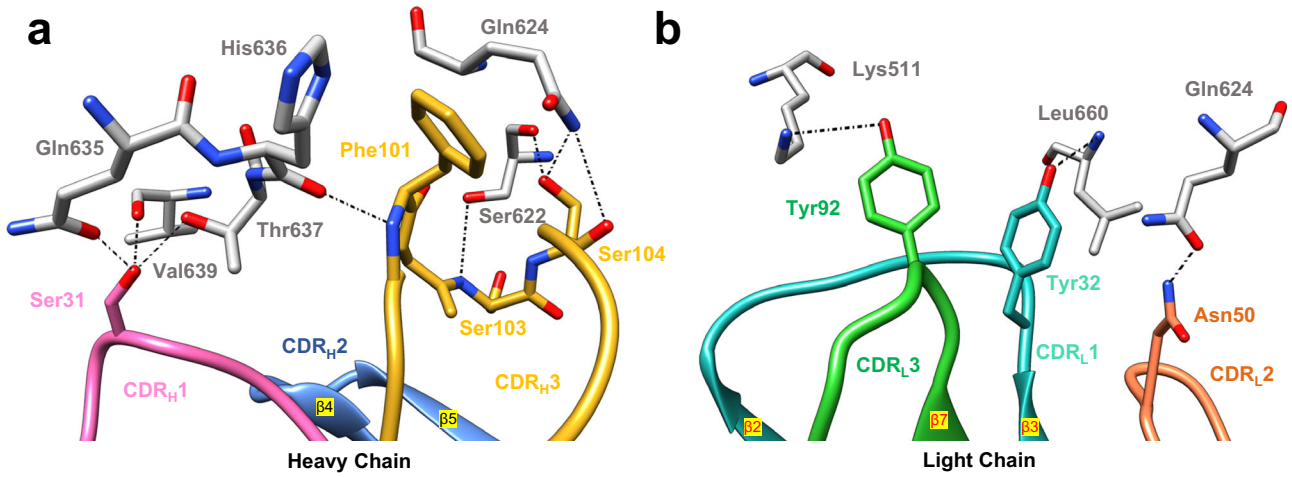

Fig. 2 | Detailed views of key epitope-paratope interacting residues between the A3-B3 domains of hCEACAM5 and tusa Fab. The backbone and side chains of hCEACAM5 are colored in gray. **a** Cartoons represent CDR_H1 (magenta) and CDR_H3 (yellow) that contribute to the interaction by forming hydrogen bond networks with the B3 domain of CEACAM5; hydrogen bonds are shown by the dotted lines. Residues from CDR_H3 are dominantly forming the strong hydrogen bond interaction network between Ser104, Ser103, and Phe101 of CDR_H3 and Ser622, Gln624, and His636 of the hCEACAM5 B3 domain. Residue Ser31 from CDR_H1 has hydrogen bonds with Gln635, Val639, and Thr637 of the hCEACAM5 B3 domain. **b** Cartoons representing CDR_L1 (cyan), CDR_L2 (coral), and CDR_L3 (green) interacting with $hCEACAM5_{A3\text{-}B3}$ domains. Residues Tyr32 from CDR_L1, Asn50 from CDR_L2, and Tyr92 from CDR_L3 have hydrogen bond interactions with Leu660, Gln624, and Leu511 of the $hCEACAM5_{A3\text{-}B3}$ domains, respectively. *CDR_H* heavy-chain complementarity-determining region, *CDR_L* light chain complementarity-determining region, hCEACAM5 human carcinoembryonic antigen-related cell adhesion molecule 5, $hCEACAM5_{A3\text{-}B3}$ A3-B3 domains of human carcinoembryonic antigen-related cell adhesion molecule, tusa Fab antigen-binding fragment of tusamitamab.

comprising residues 626–640 and 628–640 (1.3–1.4 Da). Similarly, deuterium uptake in the peptides comprising residues 626–641 and residues 628–641 did not show a significant change (1.6 Da). These results suggested that the amides of residues 627 and 628 are not involved in the CEACAM5 epitope. Comparison of the deuterium uptake in the peptide from residues 626–640 with that in the peptide from residues 626–641 and in the peptide from residues 628–640 with that in the peptide from residues 628–641 showed a small reduction in deuteration for each pair (~0.2 Da), suggesting that the amide of residue 641 is part of the CEACAM5 epitope. Similarly, a reduction in deuterium uptake between free and complexed CEACAM5 in the peptide spanning residues 638–641 (~0.3 Da) suggested that the amides of residues 639 and 640 also contribute to the epitope. Several of the amides between Ile629 and Gln638 are also involved in the epitope (total ΔHDX ~1 Da). Collectively, these differences in deuterium uptake suggest that the CEACAM5 epitope involves residues 607–613 and 629–641.

Of the 30 peptides identified in the tusa Fab heavy chain, only 1 had a significant ΔHDX between deuterated free tusa Fab and complexed CEACAM5-tusa Fab. Specifically, the peptide from residues 100–109 exhibited a ΔHDX of 0.7 ± 0.2 Da (11% ± 2% difference) after the equivalent of 200 minutes of deuteration at 4 °C. Similarly, of the 20 peptides identified in the tusa Fab light chain, only 2 had a significant ΔHDX between deuterated free tusa Fab and complexed CEACAM5-tusa Fab after 20 minutes of deuteration at 4 °C. The peptide from residues 47–54 showed a ΔHDX of 0.6 ± 0.2 Da (10% ± 2% difference), and the peptide from residues 87–104 had a ΔHDX of 0.9 ± 0.2 Da (5% ± 2% difference). Together, these differences in deuterium uptake suggest that the tusa Fab paratope involves residues 101–109 in the heavy chain and residues 48–54 and 88–104 in the light chain. It is important to note that data at 4 °C and 26 °C showed similar trends in the exchange outcome.

Residues with interactions identified using multiple methods

Overall, 16 amino acid residues in the tusa Fab paratope (10 in the heavy chain and 6 in the light chain) were identified by more than one of the methods used (cryo-EM, SPR, and HDX-MS) (Fig. 3a). When mutated to alanine, 6 of the residues (heavy chain: Tyr32, Asp33, Tyr100, and Phe101; light chain: Tyr32 and Asn50) completely eliminated binding to $hCEACAM5_{A3\text{-}B3}$, and 2 residues in the light chain (Phe30 and Tyr92) had weaker binding and a difference in binding free energy change between WT tusa Fab and the alanine variant ($\Delta G_{WT} - \Delta G_{ala}$) of $\Delta\Delta G > -2.5$ kcal/mol (Table 1).

Tusa Fab paratope residues that were identified by alanine mutagenesis mostly overlapped with those that were identified by cryo-EM (Fig. 3b). In general, residues that resulted in complete loss of binding to hCEACAM5 when mutated to alanine were clustered in the core region of the cryo-EM paratope, whereas those that only resulted in partial loss of binding when mutated to alanine were more distant.

The hCEACAM5 epitope that was identified by HDX covered a smaller surface area than that identified by cryo-EM; however, both methods detected a core region of interacting residues in the B3 domain. (Fig. 3c).

Discussion

Here we describe the high-resolution structure of the A3-B3 domains of CEACAM5, including N-linked glycans, and the detailed epitope mapping of a monoclonal antibody against CEACAM5. The structure presented includes a <70 kDa complex of the flexible and highly glycosylated antigen CEACAM5, which poses challenges for structural studies. Until now, no structure of the CEACAM5 A3-B3 domains has been determined; four other human CEACAM structures have been solved, but they are all limited to the N-terminal IgV-like domain (CEACAM1 [PDB: 6XNO], CEACAM5 [PDB: 2QSQ], CEACAM6 [PDB: 4WHC], and CEACAM8 [PDB: 4YIQ]). This high-resolution structure demonstrates the advancement of the cryo-EM technique in drug discovery, and that it can now be used for wide-spectrum small targets.

Because the CEACAM protein family contains highly similar IgG-like domains, identifying the unique epitope for CEACAM5 is crucial for understanding the specific binding and sheds light on avoiding off-target binding. In addition, the epitope location on the A3-B3 domains is important for the invagination of the cell membrane, which is significant with regard to the antibody-drug conjugate mechanism of action. These data may be used to rationally design the next generation of CEACAM-targeting monoclonal antibodies.

Table 1 | Binding affinity and kinetics of wild-type tusa Fab+Fc and alanine variants of tusa Fab+Fc binding to hCEACAM5$_{ECD}$ analyzed by surface plasmon resonance

Chain	Variant	k_a (1/M • s)	k_d (1/s)	$t_{1/2}$ (min)	K_D (nM)	ΔΔG (kcal/mol)[a]
	WT	2.75×10^5	3.86×10^{-3}	3.0	14.0	–
Heavy	Gly26Ala	3.95×10^5	3.22×10^{-3}	3.6	8.2	0.311
	Phe29Ala	4.5×10^5	3.68×10^{-3}	3.1	8.2	0.311
	Ser30Ala	3.07×10^5	2.05×10^{-3}	5.6	6.7	0.429
	Tyr50Ala	2.64×10^4	5.31×10^{-2}	0.2	2013.3	−2.893
	Ser53Ala	3.21×10^5	1.8×10^{-3}	6.4	5.6	0.534
	His99Ala	6.48×10^4	1.99×10^{-2}	0.6	306.5	−1.797
Light	Phe30Ala	2.7×10^4	3.62×10^{-2}	0.3	1338.9	−2.655
	Ser31Ala	2.91×10^5	2.38×10^{-3}	4.9	8.2	0.311
	His91Ala	7.32×10^4	6.37×10^{-2}	0.2	870.0	−2.404
	Tyr92Ala	3.08×10^4	5.52×10^{-2}	0.2	1789.7	−2.824
	Pro95Ala	1.31×10^3	1.77×10^{-2}	0.7	13,534.4	−4.002

hCEACAM5$_{ECD}$ human carcinoembryonic antigen-related cell adhesion molecule 5 extracellular domain, k_a association constant, k_d dissociation constant, K_D binding constant, $t_{1/2}$ half-life time of dissociation, *tusa Fab+Fc* tusamitamab antigen-binding fragment region plus the fragment crystallizable region, *WT* wild-type.

[a]Difference in binding free energy change (conformational stability) between wild-type tusa Fab and the alanine variant ($\Delta G_{WT} - \Delta G_{Ala}$). Negative values indicate that the alanine variant is energetically unfavorable (i.e., requires more energy to fold than the WT protein).

The epitopes and paratopes identified by HDX-MS, SPR, and cryo-EM in this study were in general agreement. Although there were some differences, this is not unexpected given the differences in how these techniques probe protein structure and macromolecular interactions. The identification of a set of residues in both the hCEACAM5$_{A3\text{-}B3}$ epitope and tusa Fab paratope by multiple methods provides greater confidence that these amino acids play a critical role in antibody recognition and specific binding between hCEACAM5 and tusa Fab, compared with epitope/paratope residues that were only identified by a single method.

Antibody paratopes preferentially involve aromatic residues (Tyr, Trp, and Phe), residues with short hydrophilic side chains (Ser, Thr, Asp, and Asn), and glycine[19,20]. Consistent with these propensities, 16 amino acid residues in the tusa Fab paratope that were identified by multiple methods belong to one of these groups. These residues tend to be enriched in antibody paratopes because they can form multiple, diverse interactions, including hydrogen bonds and van der Waals interactions, that stabilize the paratope-epitope interface. In addition, the clustering of these residues in the heavy chain, particularly CDR3, is consistent with the importance of this region in antibody-binding specificity[21]. Furthermore, the large number of hydrogen bonds, predominantly in the tusa Fab heavy chain, is in agreement with the mean (standard deviation) number of hydrogen bonds that occur in the heavy chain variable region (5.2 ± 3.0) and light chain variable (2.4 ± 2.0) domains[22]. These findings suggest that the 16 amino acids in the tusa Fab paratope that were identified by multiple methods contribute to the specificity and affinity of tusamitamab for hCEACAM5. Particularly, the 6 residues that eliminated binding to hCEACAM5$_{A3\text{-}B3}$ and the 2 residues that had weaker binding when mutated to alanine are likely to be key.

The hCEACAM5 epitope involves residues in both the A3 and B3 domains; however, the latter provides a greater overall contribution to binding. Tusa Fab primarily binds to the B3 domain, with few interactions with A3, explaining CDRs of heavy and light chains. Most epitope interactions within the B3 domain involve 6 residues (Ser622, Gln624, Gln635, His636, Thr637, and Val639) that account for all of the hydrogen bonds with the tusa Fab heavy chain and 2 of the 3 hydrogen bonds with the light chain. One amino acid in particular, Gln624, hydrogen bonds with both the heavy and light chains. In addition, 3 hydrophobic residues (Phe641, Leu660, and Ala661) may contribute hydrophobic interactions. Four of the 6 hydrogen-bonding residues (Gln635, His636, Thr637, and Val639) and 1 of the 3 hydrophobic residues (Phe641) were identified by multiple methods and are likely to be critical to the hCEACAM5 epitope.

In addition, an N-linked mannose at residue Asn612 in the B3 domain forms a hydrogen bond with the tusa Fab heavy chain. Because N-glycosylation tends to decrease protein flexibility and increase protein stability globally[23,24], all N-linked glycans in hCEACAM5$_{A3\text{-}B3}$, including the one linked to Asn612, may play an important role in shaping the overall conformation that is recognized by the tusa Fab and by other proteins involved in the cellular function of hCEACAM5. The role of glycans in CEACAM5 is also important due to the absence of salt bridges in the hCEACAM5-tusa Fab binding interface. Salt bridges play an important role in antibody-antigen interactions by imposing geometric constraints that limit conformational flexibility[25–27]. It is possible that the N-linked mannose at residue Asn612 may have a compensatory, stabilizing function, similar to critical anchor residues in other protein complexes without salt bridges[28]. This would increase the rigidity of the epitope-paratope interface and, consequently, the specificity of tusa Fab for hCEACAM5$_{A3\text{-}B3}$.

We hypothesize that the specificity of the tusa Fab for the A3-B3 domains of hCEACAM5 is due to diverse types of interactions between the tusa Fab heavy and light chains and hCEACAM5$_{A3\text{-}B3}$ combined with conformational constraints on the epitope-paratope interface (shape complementarity), which are known to play an important role in molecular recognition[22]. This binding mechanism may enable tusamitamab to distinguish hCEACAM5 from other human CEACAMs (ie, CEACAMs 1, 6, and 8) and distinguish the A3-B3 domain from the A1-B1 and A2-B2 domains in hCEACAM5.

Human CEACAM5$_{A3\text{-}B3}$ epitopes such as the one reported here may not be unique. Recently, Baek et al.[29] isolated from a human Fab phage-displayed library a monoclonal antibody, 1G9, that also specifically binds the A3-B3 domains of hCEACAM5. Initial epitope-mapping experiments using negative-stain transmission electron microscopy and site-directed mutagenesis suggested that the 1G9 epitope involves glycans at residues 612 and 650 in the B3 domain. Assuming that the epitope is located between these residues, it may partially overlap with the epitope recognized by tusamitamab but is likely to be distinct (Supplementary Fig. 8). However, because no epitope determination was presented in Baek et al.[29], this is only a hypothesis. Whether tusamitamab definitively competes for binding to hCEACAM5 with the 1G9 Fab may be answered with a competitive binding experiment, but because the amino acid sequence for the 1G9 antibody is not yet available, such an experiment is not feasible at this time. In agreement with our findings, the 1G9 epitope supports the importance of the B3 domain in specific antibody recognition of hCEACAM5.

The main strength of this study is the use of multiple epitope-mapping methods. The high quality of the cryo-EM structure enabled the determination of a high-resolution structure of hCEACAM5$_{A3\text{-}B3}$, including linked glycans (Supplementary Fig. 9). In addition, the high sequence coverage of hCEACAM5 peptides that were identified by HDX-MS provides confidence that other possible epitopes were not missed in that analysis.

Taken together, these findings define the epitope-paratope interface between hCEACAM5 and a clinically relevant and highly specific antibody. The provided structural information was used for patient stratification in the clinical trials for the development of the drug. These studies may inform the future development of CEACAM5-directed antibody-based therapies for patients with cancer.

Methods

Protein expression and purification

CEACAM5. Glycosylated CEACAM5 was expressed using plasmids containing the sequences encoding the ECD (amino acids 35–685) or

https://doi.org/10.1038/s41467-024-53746-9

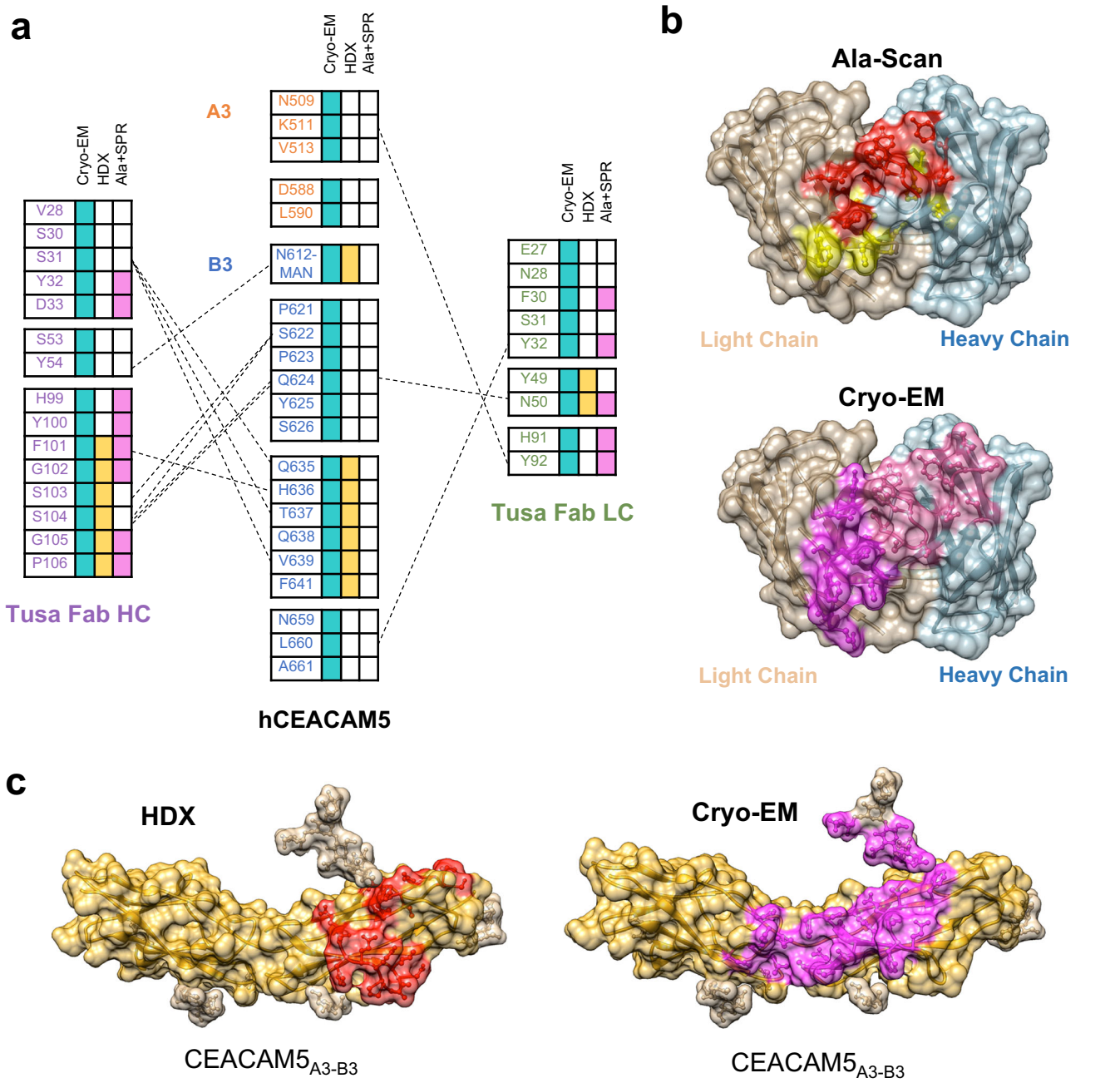

Fig. 3 | Comparison of the epitope-paratope interface between hCEACAM5$_{A3\text{-}B3}$ and tusa Fab identified by 3 different structural analysis techniques. **a** Comparison of interacting residues in hCEACAM5$_{A3\text{-}B3}$ complexed with tusa Fab. Interacting residues in hCEACAM5 (A3 domain, residues in orange text; B3 domain, residues in blue text) and tusa Fab (heavy chain, residues in purple text; light chain, residues in green text) that were identified by cryo-EM. A subset of these residues was also identified by HDX (yellow) and/or surface plasmon resonance analysis of alanine variants (Ala+SPR, magenta). Not all residues involved in epitopes or paratopes that were identified by HDX or alanine mutagenesis are shown. Hydrogen bonds are depicted by dashed lines. **b** Comparison of tusa Fab paratope residues identified by alanine mutagenesis (yellow highlighting indicating partial binding and red highlighting indicating no binding) and cryo-EM (light pink highlighting indicating the interacting residues of the heavy chain and magenta highlighting indicating the interacting residues from the light chain). The solvent-accessible surface of the tusa Fab heavy and light chains is illustrated as light brown and light blue, respectively. **c** Comparison of hCEACAM5 epitope residues identified by HDX (highlighted in red) and cryo-EM (highlighted in magenta). The solvent-accessible surface of hCEACAM5$_{A3\text{-}B3}$ and N-linked glycans are depicted as golden brown and light brown, respectively. **b**, **c** Secondary structures (darker ribbons and ball-and-stick moieties) are shown. Ala-Scan, alanine scanning mutagenesis; Ala+SPR, surface plasmon resonance of alanine variants; cryo-EM, cryogenic electron microscopy; HC, heavy chain; hCEACAM5, human carcinoembryonic antigen-related cell adhesion molecule 5; hCEACAM5$_{A3\text{-}B3}$, A3-B3 domains of human carcinoembryonic antigen-related cell adhesion molecule 5; HDX, hydrogen-deuterium exchange; LC, light chain; MAN, mannose; tusa Fab, antigen-binding fragment of tusamitamab.

A3-B3 domains (amino acids 499–685) of hCEACAM5 (GenBank accession ID: M29540.1). HEK293FS cells (Thermo Fisher) were transiently transfected using 293fectin™ transfection reagent according to manufacturer (Gibco) recommendations and then grown in suspension culture for 8 days at 37 °C, 115 rpm, with 0.8% CO_2 for pH regulation, in FreeStyle 293 Expression Medium (Gibco, 12-338-018). The culture supernatant was collected and purified using immobilized metal-affinity chromatography (IMAC), followed by size-exclusion

chromatography (SEC) on a HiLoad Superdex 200 PG preparative SEC column (Cytiva, product code 28989331) for the A3-B3 domain or anion-exchange chromatography for the ECD on a Q Sepharose HP column (Cytiva, 29018182) with a 0–500 mM NaCl gradient. The purity, molecular weight, and sequence of the CEACAM5 protein were confirmed by analytical SEC, liquid chromatography/mass spectrometry (LC/MS) or N-terminal sequencing, and sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Analytical SEC was performed with a Superdex 200 Increase 5/150 column (Cytiva, 28990945) that was equilibrated with Dulbecco's phosphate-buffered saline (DPBS, Gibco, 12559069) at room temperature on an Agilent 1290 Infinity LC system with OpenLab software. LC/MS was performed using a QExactive Plus MS spectrometer coupled with a Vanquish Core HPLC system (Thermo Scientific, VQ-CORE-BIN-01) equipped with an MAbPac RP analytical column (Thermo Scientific, 088644) at 30 °C and run with LC eluant mix WF2 UpS (ROMIL Ltd, R4905) as the mobile phase. Raw data retreatment was performed using Expressionist software (Genedata). When no LC/MS data could be obtained, N-terminal sequencing (Edman degradation) was performed using a PPSQ-33A sequencer according to manufacturer (Shimadzu) recommendations. Protein batches that were used for mouse immunization were confirmed to be free of endotoxin.

Deglycosylated CEACAM5$_{A3\text{-}B3}$ was produced in a similar manner, except cell cultures were grown in the presence of 0.2 mM kifunensine (Enzo Life Sciences, BML-S114), an inhibitor of mannosidase I, a glycoprotein processing enzyme[30]. Subsequently, the culture supernatant was incubated with up to 625 U/mL endoglycosidase H (New England Biolabs, P0702S) for 3 hours at 37 °C to cleave N-linked high-mannose oligosaccharides. The purity and composition of deglycosylated CEACAM5$_{A3\text{-}B3}$ were confirmed as described for glycosylated CEACAM5 previously.

Recombinant tusa Fab and tusa Fab + Fc. Prior development of tusamitamab is described in the Supplementary Note. To generate the tusa Fab used in the current study, DNA sequences encoding the tusamitamab variable regions were cloned into proprietary (Sanofi) expression vectors in fusion with human heavy chain constant domain 1 (C_H1) with a six-histidine tag and kappa light chain constant domain (Cκ). These vectors were transiently transfected and expressed in HEK293FS cells, and the tusa Fab protein was purified from the culture supernatant by IMAC as described above for recombinant CEACAM5.

To generate the tusa Fab+Fc variants used in the SPR experiments, proprietary (Sanofi) plasmids encoding alanine point mutations in tusa Fab (Supplementary Table 3) along with the immunoglobulin G1 Fc region were transiently transfected and expressed in HEK293FS cells (Supplementary Methods) and then purified from the culture supernatant using protein A affinity chromatography.

Cryo-EM and structure analysis

For cryo-EM, a 1.2 molar excess of tusa Fab was mixed with glycosylated hCEACAM5$_{A3\text{-}B3}$ to form the hCEACAM5$_{A3\text{-}B3}$-tusa Fab complex, concentrated to 2–3 mg/mL using a centrifugal protein concentrator with a 10 kDa molecular weight cutoff (Merck Millipore), then purified by SEC on a Superdex 200 column (Cytiva).

Purified hCEACAM5$_{A3\text{-}B3}$-tusa Fab in DPBS (0.87 mg/mL) was spotted onto glow-discharged UltraAuFoil R0.6/1.0 Au grids (Quantifoil Micro Tools GmbH). Grids were vitrified using a Vitrobot Mark IV system (FEI) at 4 °C, then stored in liquid nitrogen.

Cryo-EM data were recorded on a Glacios electron microscope equipped with a Falcon4 direct electron detector (Thermo Fisher Scientific) (Supplementary Table 1). The electron microscope was operated at 200 kV, and data were collected in counting mode using EPU software (version 2.9) using a defocus range between −0.6 and −2.0 μm with a pixel size of 0.58 Å. Each movie contained 54 frames (162 electron-event representation exposure fractions) with a dose per frame of 1.14 electrons/Å2 and a total exposure time of 4.72 seconds. Data were processed with CryoSPARC (version 3.2). Full-frame motion correction and estimation of contrast-transfer function (CTF) parameters were carried out using the built-in patch-motion correction and patch-CTF estimation routines, respectively. Using a 6 Å cutoff for the CTF fit parameter for micrograph selection, 4455 movies were manually selected, from which 2,614,488 particles were auto picked. After 2–3 rounds of 2D classification, 554,660 selected particles were further used for ab initio model generation with three classes. Following heterogeneous refinement, 213,470 particles were used for 3D nonuniform refinement and a final local refinement, which resulted in a map at 3.11 Å resolution with a locally higher resolution (Supplementary Fig. 1 and Supplementary Fig. 10).

Initial models were obtained from AlphaFold[31], fit into the cryo-EM density using rigid body fit in UCSF CHIMERA (version 1.17.3), and then manually adjusted using Coot (version 0.9.4.1)[32]. All models were real-space refined in PHENIX (version 1.19.2)[33], and the quality was assessed with MolProbity (Supplementary Table 1). Structural analysis of the complex interface was performed with Proteins, Interfaces, Structures, and Assemblies (PISA)[34–37] and UCSF Chimera[38]. Standard PISA criteria were used for detecting interactions. Specifically, a cutoff distance of 3.89 Å between donor and acceptor atoms was used to detect hydrogen bonds and of 4.00 Å for salt bridges (Supplementary Table 1).

Protein sequence analysis

Amino acid sequences for human CEACAM1, 5, 6, and 8 were obtained from the UniProt Knowledgebase (accession IDs P13688, P06731, P40199, and P31997, respectively)[39]. Multiple sequence alignments were performed using Clustal Omega version 1.2.4[34]. The proportion of identical residues in an alignment was calculated using pseqsid (available from github.com/amaurypm/pseqsid).

Surface plasmon resonance

Binding of CEACAM5-directed monoclonal antibodies (tusa Fab+Fc) to hCEACAM5$_{ECD}$ was analyzed using SPR, which detects biomolecular interactions by measuring changes in the refractive index on the surface of a sensor chip[35], on a Biacore™ 3000 system (Cytiva). Binding interactions were assessed by 2 types of SPR analyses: screening and kinetic. For each, hCEACAM5$_{ECD}$ binding to tusa Fab+Fc WT or its alanine variants was measured, with lower %R_{max} indicating reduced binding and lower %Remaining signal after 50 seconds of dissociation indicating decreased stability of the hCEACAM5$_{ECD}$-tusa Fab+Fc complex.

For both analyses, mouse anti-human IgG antibody (25 μg/mL) was immobilized on an activated CM5 Biacore sensor chip at a flow rate of 5 μL/min for 7 minutes, as per manufacturer instructions in the Human Antibody Capture Kit (Cytiva). In the screening analyses, 1 μg/mL of tusa Fab+Fc was captured over the anti-human Fc surface at 10 μL/min for 1 minute. Subsequently, 50 nM hCEACAM5$_{ECD}$ was injected at a flow rate of 30 μL/min. Association was monitored for 1 minute, followed by dissociation for 1 minute. In the kinetic analyses, 0.3 μg/mL of tusa Fab +Fc was captured over anti-huFc surface at 10 μL/min for 1 minute. Subsequently, 150 nM hCEACAM5$_{ECD}$ was injected at a flow rate of 30 μL/min. Association was monitored for 5 minutes, followed by dissociation for a variable period. Kinetic parameters were calculated using a 1:1 Langmuir model. For both analyses, the biosensor chip surface was regenerated using 3 M $MgCl_2$ for 30 seconds at 10 μL/minute, and sensorgrams were analyzed with BIAevaluation software (Cytiva).

Hydrogen-deuterium exchange mass spectrometry

Deglycosylated hCEACAM5$_{A3\text{-}B3}$ was mixed with 1.5 molar excess of the tusa Fab to form the hCEACAM5$_{A3\text{-}B3}$-tusa Fab complex. Excess unbound hCEACAM5$_{A3\text{-}B3}$ was removed by SEC, and the fraction corresponding to the hCEACAM5$_{A3\text{-}B3}$-tusa Fab complex was used for HDX studies.

Reactions were automated using a PAL autosampler (CTC Analytics). HDX was initiated by diluting stock solutions of hCEACAM5, tusa Fab, or hCEACAM5$_{A3-B3}$-tusa Fab complex to 60 µM, 26 µM, or 4 µM, respectively, in PBS in D_2O. The protein solutions were diluted by a factor of 5 for HDX labeling and by a factor of 2 for quenching, resulting in final concentrations of 6 µM, 2.6 µM, and 0.4 µM for hCEACAM5, tusa Fab, and hCEACAM5$_{A3-B3}$-tusa Fab complex, respectively. HDX labeling was performed for periods of 2, 20, and 200 minutes at 4 °C or 20 minutes at 26 °C (equivalent to 200 minutes at 26 °C) for the last period (Supplementary Table 4). The 26 °C temperature was used to increase the exchange rate. Subsequently, a solution of 2 M guanidine hydrochloride, 0.8 M tris(2-carboxyethyl) phosphine, and 1 M glycine pH 2.2 was added and incubated for 2 minutes at 4 °C to quench the exchange reaction, minimize back-exchange (reversion of incorporated deuterium to hydrogen), and reduce disulfides. All reactions were performed in triplicate.

For mass spectrometry (MS) analysis, proteins were first digested with pepsin and nepenthesin in solution. The resulting peptides were desalted using a peptide micro-trap microcolumn (Michrom Bioresources, Bruker Daltonics, Bremen, Germany) with 0.03% trifluoroacetic acid (TFA) at a flow rate of 100 µL/min, then separated on an analytical column (Poroshell 120 EC-C18, 1 × 50 mm, 2.7 µm from Agilent Technologies) with 0.03% TFA in water (solvent A) and a 15%–100% gradient of solvent B (90% acetonitrile, 0.03% TFA in water) over 20 minutes. The HPLC eluent was introduced into an Agilent 6210 electrospray time-of-flight LC/MS mass spectrometer (Agilent Technologies) to perform peptide mass analysis in the 300–1300 m/z range as described in Giladi et al.[40]. Peptides were identified by tandem mass spectrometry (MS/MS) with a Bruker 9.4 T Apex-Q FTMS (Bruker Daltonics) as reported in Giladi et al.[41]. MS data were processed using Data Analysis (Bruker) and MassHunter (Agilent Technologies) software. Similarly, MS/MS data were processed with Data Analysis and Mascot (Matrix Science) software. MassHunter (Agilent Technologies) and HDExaminer (Sierra Analytics) software were used for HDX data processing with the user-selected option of excluding the first amide residue in HDExaminer parameters.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM structure of the hCEACAM5$_{A3-B3}$-tusa Fab complex has been deposited in the RCSB Protein Data Bank under accession code 8BW0 and the Electron Microscopy Data Bank under accession code EMD-16279, which also contains the tusa Fab amino acid sequences. Other data generated in this study are available within the Supplementary Information file. Source data are available as a Source Data file. Due to circumstances outside of the authors' control, the raw HDX-MS data generated by a contracted organization could not be located. The report detailing the methods and description of the results and conclusions reached are included as Supplementary Data 1. Source data are provided with this paper.

References

1. Hammarström, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin. Cancer Biol.* **9**, 67–81 (1999).
2. Beauchemin, N. & Arabzadeh, A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev* **32**, 643–671 (2013).
3. Decary, S. et al. Preclinical activity of SAR408701: a novel anti-CEACAM5-maytansinoid antibody-drug conjugate for the treatment of CEACAM5-positive epithelial tumors. *Clin. Cancer Res.* **26**, 6589–6599 (2020).
4. Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
5. Beauchemin, N., Benchimol, S., Cournoyer, D., Fuks, A. & Stanners, C. P. Isolation and characterization of full-length functional cDNA clones for human carcinoembryonic antigen. *Mol. Cell Biol.* **7**, 3221–3230 (1987).
6. Roda, G. et al. Characterizing CEACAM5 interaction with CD8alpha and CD1d in intestinal homeostasis. *Mucosal Immunol.* **7**, 615–624 (2014).
7. Zhao, Q. et al. Glycan analysis of colorectal cancer samples reveals stage-dependent changes in CEA glycosylation patterns. *Clin. Proteomics* **15**, 9 (2018).
8. Chiang, W.-F. et al. Carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) promotes EGF receptor signaling of oral squamous cell carcinoma metastasis via the complex N-glycosylation. *Oncogene* **37**, 116–127 (2018).
9. Pont, L. et al. Site-specific N-linked glycosylation analysis of human carcinoembryonic antigen by sheathless capillary electrophoresis-tandem mass spectrometry. *J. Proteome Res.* **20**, 1666–1675 (2021).
10. Tate, J. & Ward, G. Interferences in immunoassay. *Clin. Biochem. Rev* **25**, 105–120 (2004).
11. Jin, S. et al. Emerging new therapeutic antibody derivatives for cancer treatment. *Signal Transduct. Target Ther.* **7**, 39 (2022).
12. Gazzah, A. et al. Safety, pharmacokinetics, and antitumor activity of the anti-CEACAM5-DM4 antibody-drug conjugate tusamitamab ravtansine (SAR408701) in patients with advanced solid tumors: first-in-human dose-escalation study. *Ann. Oncol.* **33**, 416–425 (2022).
13. Gazzah, A. et al. Efficacy and safety of the antibody-drug conjugate (ADC) SAR408701 in patients (pts) with non-squamous non-small cell lung cancer (NSQ NSCLC) expressing carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5). *J. Clin. Oncol.* **38**, 9505 (2020).
14. Ricordel, C. et al. Safety and efficacy of tusamitamab ravtansine (SAR408701) in long-term treated patients with nonsquamous non-small cell lung cancer (NSQ NSCLC) expressing carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5). American Society for Clinical Oncology Annual Meeting; Chicago, IL; June 3-7, 2022.
15. Boehm, M. K. et al. Extended glycoprotein structure of the seven domains in human carcinoembryonic antigen by X-ray and neutron solution scattering and an automated curve fitting procedure: implications for cellular adhesion. *J. Mol. Biol.* **259**, 718–736 (1996).
16. Bates, P. A., Luo, J. & Sternberg, M. J. A predicted three-dimensional structure for the carcinoembryonic antigen (CEA). *FEBS Lett.* **301**, 207–214 (1992).
17. Bonsor, D. A., Günther, S., Beadenkopf, R., Beckett, D. & Sundberg, E. J. Diverse oligomeric states of CEACAM IgV domains. *Proc. Natl. Acad. Sci. USA.* **112**, 13561–13566 (2015).
18. EMBL-EBI. InterPro: IPR008424–immunoglobulin C2-set. Available at https://www.ebi.ac.uk/interpro/entry/InterPro/IPR008424/.
19. Nguyen, M. N., Pradhan, M. R., Verma, C. & Zhong, P. The interfacial character of antibody paratopes: analysis of antibody-antigen structures. *Bioinformatics* **33**, 2971–2976 (2017).
20. Peng, H.-P., Lee, K. H., Jian, J.-W. & Yang, A.-S. Origins of specificity and affinity in antibody-protein interactions. *Proc. Natl. Acad. Sci. USA.* **111**, E2656–E2665 (2014).
21. D'Angelo, S. et al. Many routes to an antibody heavy-chain CDR3: necessary, yet insufficient, for specific binding. *Front. Immunol.* **9**, 395 (2018).
22. Kuroda, D. & Gray, J. J. Shape complementarity and hydrogen bond preferences in protein-protein interfaces: implications for antibody modeling and protein-protein docking. *Bioinformatics* **32**, 2451–2456 (2016).

23. Lee, H. S., Qi, Y. & Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci. Rep.* **5**, 8926 (2015).
24. Zhou, Q. & Qiu, H. The mechanistic impact of N-glycosylation on stability, pharmacokinetics, and immunogenicity of therapeutic proteins. *J. Pharm. Sci.* **108**, 1366–1377 (2019).
25. Sinha, N., Mohan, S., Lipschultz, C. A. & Smith-Gill, S. J. Differences in electrostatic properties at antibody-antigen binding sites: implications for specificity and cross-reactivity. *Biophys. J.* **83**, 2946–2968 (2002).
26. Tsumoto, K. et al. Role of salt bridge formation in antigen-antibody interaction. Entropic contribution to the complex between hen egg white lysozyme and its monoclonal antibody HyHEL10. *J. Biol. Chem.* **271**, 32612–32616 (1996).
27. Xu, D., Tsai, C. J. & Nussinov, R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* **10**, 999–1012 (1997).
28. Yugandhar, K. & Gromiha, M. M. Protein-protein binding affinity prediction from amino acid sequence. *Bioinformatics* **30**, 3583–3589 (2014).
29. Baek, D.-S. et al. A highly-specific fully-human antibody and CAR-T cells targeting CD66e/CEACAM5 are cytotoxic for CD66e-expressing cancer cells in vitro and in vivo. *Cancer Lett.* **525**, 97–107 (2022).
30. Elbein, A. D., Tropea, J. E., Mitchell, M. & Kaushal, G. P. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J. Biol. Chem.* **265**, 15599–15605 (1990).
31. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
33. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
34. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
35. Nguyen, H. H., Park, J., Kang, S. & Kim, M. Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel)* **15**, 10481–10510 (2015).
36. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* **43**, W314–W319 (2015).
37. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
38. Pettersen, E. F. et al. UCSF chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
39. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
40. Giladi, M. et al. Dynamic distinctions in the Na(+)/Ca(2+) exchanger adopting the inward- and outward-facing conformational states. *J. Biol. Chem.* **292**, 12311–12323 (2017).
41. Giladi, M. et al. Asymmetric preorganization of inverted pair residues in the sodium-calcium exchanger. *Sci. Rep.* **6**, 20753 (2016).

Acknowledgements

The authors acknowledge Dr. Armelle Buzy, Dr. Tarik Dabdoubi, Dr. Cendrine Lemoine, and Dr. Pierrick Rival (Sanofi) for critically reviewing the manuscript. Writing assistance was provided by Alexander Simon of inScience Communications (Philadelphia, PA). This work was performed in accordance with current Good Publication Practice guidelines and was funded by Sanofi.

Author contributions

A.K. optimized and performed cryo-EM sample preparation; did data collection, model building, refinement, structure description and analysis; and wrote the manuscript. F.D. designed the constructs, produced and purified recombinant $CEACAM5_{A3\text{-}B3}$ and Fab, and wrote the manuscript. M.G. performed the SPR and alanine scanning experiments. C.R. and T.B. contributed to the discussion. A.R. did HDX analysis, managed, and supervised the project and manuscript writing.

Competing interests

The authors declare the following competing interests: All authors are employees of Sanofi and may hold shares and/or stock options in the company.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53746-9.

Correspondence and requests for materials should be addressed to Alexey Rak.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.