HHS Public Access

Author manuscript

Structure. Author manuscript; available in PMC 2023 May 05.

Published in final edited form as:

Structure. 2022 May 05; 30(5): 658–670.e5. doi:10.1016/j.str.2022.02.003.

Human CEACAM1 N-domain Dimerization is Independent from Glycan Modifications

Meagan Belcher Dufrisne[a], **Nicole Swope**[a], **Marissa Kieber**[a], **Jeong-Yeh Yang**[b], **Ji Han**[a], **Jason Li**[a], **Kelley W. Moremen**[b], **James H. Prestegard**[b], **Linda Columbus**[a,*]

[a]Department of Chemistry, University of Virginia, Charlottesville, VA 22904, USA

[b]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA

Summary

Carcinoembryonic cellular adhesion molecules (CEACAMs) serve diverse roles in cell signaling, proliferation and survival and are made up of one or several immunoglobulin (Ig)-like ectodomains glycosylated *in vivo*. The physiological oligomeric state and how it contributes to protein function are central to understanding CEACAMs. Two putative dimer conformations involving different CEACAM1 N-terminal Ig-like domain (CCM1) protein faces (ABED and GFCC′C″) were identified from crystal structures. GFCC′C″ was identified as the dominant CCM1 solution dimer, but ambiguity regarding the effect of glycosylation on dimer formation calls its physiological relevance into question. We present the first crystal structure of minimally glycosylated CCM1 in the GFCC′C″ dimer conformation and characterization in solution by continuous-wave and double electron-electron resonance electron paramagnetic resonance spectroscopy. Our results suggest the GFCC′C″ dimer is dominant in solution with different levels of glycosylation, and structural conservation and co-evolved residues support the GFCC′C″ dimer is conserved across CEACAMs.

Graphical Abstract

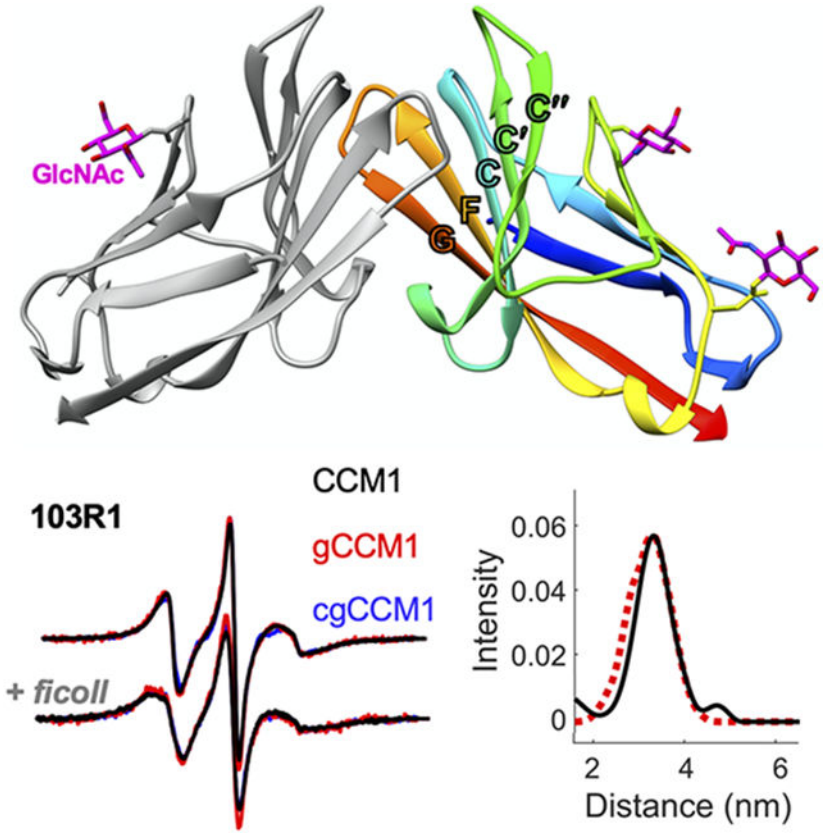
*Lead Contact/Corresponding author. columbus@virginia.edu.

Author Contributions

M.B.D. performed crystallization experiments with protein purified and provided by J.Y.Y. M.B.D froze crystals and shipped them to APS where she collected data remotely. M.B.D. processed the data, built, and refined the model. M.K., N.S. and M.B.D. purified and spin labeled all protein used for EPR experiments with help from J.L. and J.H. and input from J.Y.Y. EPR data were collected and analyzed by M.K., N.S., and M.B.D. N.S. performed structural conservation analysis and analysis of co-evolved residues. L.C. supervised all aspects of the project, providing essential intellectual contributions. K.W.M. and J.H.P. provided feedback and biological insight on experimental design and data interpretation. M.B.D., N.S., and L.C. wrote the manuscript with input from all authors.

Declaration of Interests

The authors declare no financial conflict of interest.



eTOC Blurb

Details of the glycosylated CEACAM1 oligomeric state have remained ambiguous. Belcher Dufrisne, et al present the crystal structure of minimally glycosylated N-terminal CEACAM1 domain in a dimer involving the GFCC′C″ face and use electron paramagnetic resonance spectroscopic techniques to show the dimer is dominant in solution regardless of glycosylation state.

Keywords

carcinoembryonic antigen-related cell adhesion molecule (CEACAM); X-ray crystallography; electron paramagnetic (EPR) spectroscopy; dimerization; glycosylation

Introduction

Carcinoembryonic antigen cellular adhesion molecules (CEACAMs) are cell surface adhesion proteins with immunoglobin (Ig)-like ectodomains. CEACAMs mediate many different functions in the cell such as proliferation, signaling, survival, and tumor suppression (Kuespert et al., 2006). The diverse functional roles of CEACAMs may be attributed to structural variation within the protein family, which are encoded by 12 genes (CEACAM1, 3-8 (Figure 1A), 16, 18-21) and can exist in multiple isoforms (Beauchemin and Arabzadeh, 2013). Isoforms can be secreted by the cell or tethered to the membrane

through a transmembrane helix that may include cytoplasmic signaling domains, or a glycosylphosphatidylinositol (GPI) anchor (Beauchemin and Arabzadeh, 2013). CEACAMs with transmembrane helices and cytoplasmic domains often contribute to signaling pathways dependent on immunoreceptor tyrosine-based inhibition/activation (ITIM/ITAM) motifs (Kammerer and Zimmermann, 2010).

Some CEACAMs form homo- and hetero-oligomers, which modulate intracellular signaling. For example, equilibrium between monomeric and dimeric CEACAM1 on the apical cell surface is crucial for tight control of cell-cell adhesion, facilitated by activation of its cytoplasmic domain (Patel et al., 2013). Cellular loss of CEACAM1 monomer/dimer equilibrium control can promote tumor formation (Lawson et al., 2012) and may be disrupted by pathogen receptors for adhesion and/or immune suppression (Bonsor et al., 2018; Korotkova et al., 2008; Moonens et al., 2018). The extracellular N-terminal domain (CCM) is required for the adhesion properties of CEACAM1, and mediates interactions with itself (Tchoupa et al., 2014), other CEACAM isoforms (Öbrink, 1997), and pathogens (Bonsor et al., 2018; Korotkova et al., 2006; Martin et al., 2016; Virji et al., 1996). CEACAM1 distribution varies extensively across tissue types, with expression on epithelial cells, endothelial cells, and immune cell types (Kuespert et al., 2006). Molecular interactions between N-terminal CEACAM1 domains (CCM1) are, therefore, important to understand CEACAM1 signaling, recognition by pathogens, and tumor formation.

The monomeric structures of CCMs are remarkably similar (Bonsor et al., 2015a; Bonsor et al., 2015b; Bonsor et al., 2018; Fedarovich et al., 2006; Huang et al., 2014). Most CCMs crystallize as dimers, even if the concentration of dimer in solution is small (Bonsor et al., 2015b). CCM1 in particular has a strong propensity to dimerize exhibiting K_D values for the homodimer between ~ 0.5 (Bonsor et al., 2015b) and 2.0 μM (Bonsor et al., 2018) (Supplemental Table S1). Though the crystallographic interfaces observed in CCM structures are diverse (Supplemental Table S4), two CCM1 dimer interfaces identified from crystal structures have been named as putative physiological dimers to date: one occurring at the ABED β-strand interface (PDB ID: 2GK2; (Fedarovich et al., 2006), and the other occurring at the GFCC′C″ interface (PDB ID: 4QXW; Huang et al., 2014; Figure 1B). While studies have reported CCM1 interactions at the ABED interface, the asymmetric unit can be expanded to identify GFCC′C″ interface dimers (Zhuo et al., 2016; Supplemental Table S3). In all cases in which the ABED interface is observed, the GFCC′C″ interface is as well; however not all structures containing the GFCC′C″ dimer also have the ABED interface represented in the crystal packing (Supplemental Table S4). CCM1 homotypic and heterotypic interactions are typically attributed to the GFCC′C″ interface (Bonsor et al., 2015b; Bonsor et al., 2018; Korotkova et al., 2008; Markel et al., 2004; Taheri et al., 2000; Virji et al., 1999; Watt et al., 2001). Electron tomography data with soluble rat CEACAM1 ectodomains containing all four Ig domains supports dimers of the GFCC′C″ interface and shows that an additional distinct orientation of the dimer can occur but is a minor population (Klaile et al., 2009). Non-glycosylated CCM1 in solution was reported to have the GFCC′C″ dimer interface (Zhuo et al., 2016) and alanine mutations on the GFCC′C″ dimerization face were shown to increase the monomer population (Gandhi et al., 2021). Collectively, these data suggest the dimer observed in the 4QXW structure is the dominant domain orientation in solution, though is not conclusive (Zhuo et al., 2016).

Understanding oligomerization behavior of CCMs is further complicated by post-translational modifications. All reported human CCM crystal structures to date lack post-translational modifications, so the atomic-level understanding of CCM dimerization in this context is limited. Three Asn residues in human CCM1 can be glycosylated (N70, N77, and N81; Zhuo et al., 2016; Figure 1B), and these residues are conserved among CCM1, 3, 5, 6, 7, and 8 (Supplemental Figure S1) with N70 being conserved across the CEACAM protein family (Tan et al., 2002). Based on the structures of CCM, it has been suggested that complex glycosylation would obstruct formation of the ABED-interface dimer observed in PDB ID 2GK2 (Fedarovich et al., 2006; Klaile et al., 2009; Zhuo et al., 2016). Oligomerization of human CCM1 constructs with single GlcNAc modifications (gCCM1, Figure 1C) was investigated with NMR and other biophysical methods with variable results suggesting a heterogeneous preparation can impact oligomerization (Zhuo et al., 2016, 2020). Furthermore, crystal structures of glycosylated mouse CEACAM1a (PDB ID: 1L6Z; Peng et al., 2017; Tan et al., 2002) and CEACAM1b (PDB ID: 5VST; Peng et al., 2017) demonstrate the GFCC′C″ dimer as well as a unique interface at the ABED face that is non-superimposable with the ABED dimer of PDB ID 2GK2 and has a buried surface area less than 200 $Å^2$and, thus, more likely a crystal contact than a true dimer interface (Peng et al., 2017; Tan et al., 2002). In light of all this, questions about the impact of glycosylation on the oligomerization of CCM1 remain.

In this work, the impact of glycosylation on CCM1 dimerization and protomer orientation was investigated with X-ray crystallography, continuous-wave (CW) and double electron-electron resonance (DEER) electron paramagnetic resonance (EPR) spectroscopy. The crystal structure of gCCM1 was determined, revealing the same GFCC′C″ interface dimer in the presence of GlcNAc (N-acetylglucosamine) modifications. To obtain spin-labeled glycosylated proteins, cysteine mutants were expressed in HEK293S GnT1- and FreeStyle™ 293-F cells as previously described (Moure et al., 2018), purified, and reacted with a sulfhydryl reactive nitroxide spin label. The solution CCM1 structure in the absence of glycosylation was determined using experimental DEER restraints and simulated annealing with crystal structures of CCM1. Two DEER distances were used to confirm the glycosylated CCM1 forms are dimeric in solution and formed at the GFCC′C″. Combined, the results support CCM1 GFCC′C″ interface dimers in solution with and without glycosylation.

Results and Discussion

Crystal structure of gCCM1 reveals the GFCC′C″ dimer

Minimally glycosylated gCCM1 was crystallized (Supplemental Figure S2A), producing diffraction sufficient (Supplemental Figure S2B and C) to solve the structure to 1.7 Å resolution (Figure 2A; Table 1). The overall architecture of the dimer aligns well to the previously solved GFCC′C″ CCM1 dimer (PDB ID 4QXW; Supplemental Figure S2D) with a rmsd of 0.67 Å (Supplemental Table S3). Density is observed for GlcNAc modifications of three Asn residues within the dimer (N70 of both monomers, and N77 of only the chain B monomer; Figure 2B). Electron density for GlcNAc modifications were not observed at N77 of the chain A monomer or at N81 of either monomer (Figure

2B, Supplemental Figure S2E). This may be because of decreased occupancy due to natural variation in glycosylation patterns, crystal packing and relative flexibility of the regions containing these residues, or both. Poor density of GlcNAc modifications due to flexibility of the region is supported by higher average B-factor values for residue N81 of both monomer chains (50.33 and 57.75 for chain A and B respectively) compared to N70 (33.51 and 28.89 for chain A and B respectively), N77 (36.27 and 39.68 for chain A and B respectively) and the overall protein (32.38). Glycosylation of N70 has previously been hypothesized as important as a shield for a hydrophobic patch of the mouse CEACAM1a homolog (Tan et al., 2002) and is conserved among all CEACAMs (Tan et al., 2002; Supplemental Figure S1). When structurally aligned with a non-glycosylated CCM1 structure (PDB ID 4QXW), the GlcNAc of gCCM1 aligned with an OG molecule bound to the non-glycosylated dimer (Supplemental Figure S2D). In this structure, N70 of both monomers are the highest quality electron density of the GlcNAc modifications. These data support the biological importance of glycosylation at N70, however the specific function remains unclear.

Many previous crystal structures of non-glycosylated CCM proteins have revealed the GFCC′C″ dimer, which is recapitulated in the structure of gCCM1 presented here (Supplemental Table S4). Other interfaces observed in the crystallographic symmetry were identified as potential biologically relevant dimers. For example, the ABED dimer (Fedarovich et al., 2006) is found in five of the eleven unique space groups of CECAM structures (Supplemental Table S4). However, these observed interfaces have less surface area buried than the GFCC'C" interface, are diverse, and vary with space group. Predicting diverse biological dimers from interfaces of crystal structures is limited and complicated by the fact that CCM domains are relatively small, form transient interactions for their diverse biological functions, and the surface areas of interfaces observed from CCM crystal structures are lower than the range observed for permanent dimers (approximately 1000-3000 $Å^2$; Luo et al., 2015; Supplemental Table S4) and within, or lower than, the range observed for both weak, transient protein-protein interactions and general crystal packing contacts (approximately 500-900 $Å^2$ and 400-700 $Å^2$ respectively; Luo et al., 2015; Supplemental Table S4). Many observed crystal interfaces may in fact be biologically relevant, but it is unclear. Investigating the dominant oligomeric state of CCM1 glycosylated forms in solution is essential to understand CCM1 biological assemblies. For this reason, we continued our inquiry into the dimer conformation of the solution state CCM1 with varying glycosylation.

CCM1 oligomeric states persist with glycan modifications

To investigate the oligomeric state of CCM1 in solution, CW EPR spectra were recorded for 10 spin-labeled sites on CCM1 (Figure 1C, Supplemental Figure S4A and C; the spin label side chain is referred to as R1 as seen in Figure 3B) to observe if EPR line shapes are consistent with CCM1 oligomers or show evidence of interface contacts between protomers (Cornish et al., 1994; Hubbell et al., 2000). EPR line shapes are dependent on the rotational motion of the nitroxide, which is influenced by (i) internal rotations of the nitroxide side chain, (ii) backbone motions, and (iii) domain or overall protein correlation time for spin-labeled proteins (Hubbell et al., 2000). For X-band CW EPR, Brownian rotational diffusion

of proteins ≥ 50 kD is too slow to contribute to spectral averaging. For smaller proteins, experiments can be performed in solutions of increased viscosity such as ficoll or sucrose. A comparison of EPR lineshapes with and without the viscous agent allows the assessment of the overall correlation time of the protein and, thus, the oligomeric state (Equations 1-3). In addition, with the contributions of overall rotation removed, the internal rotations of the side chain can be assessed with respect to an oligomeric model. Residues at an oligomeric interface will have restricted side chain motion due to tertiary interactions where in the monomeric form there lacks steric modulation of the internal rotation and a more mobile line shape is observed (Supplemental Figure S3).

A comparison of CW EPR line shapes of ten spin-labeled sites of CCM1 with and without ficoll (Supplemental Figure S4A) shows apparent spectral line broadening upon the addition of ficoll, indicating that the overall protein correlation time is contributing to nitroxide dynamics (Lopez et al., 2009; McHaourab et al., 1996). For most of the sites, EPR line shapes alone cannot distinguish monomeric from dimeric CCM1. However, the observed line shape for 94R1 (Figure 3A) indicates that nitroxide is at a tertiary contact site with restricted mobility, supporting that CCM1 is forming an oligomer at the GFCC′C″ interface (Figure 3B). If CCM1 was a monomer or ABED dimer, 94R1 would be a solvent-exposed site based on the crystal structures, thus producing a mobile nitroxide and EPR line shape similar to other residues (Supplemental Figure S3).

From the experimental $2A_{zz}'$ values measured from CCM1 94R1 EPR spectra, τ_n (the overall correlation time of the spin label with all rotational contributions) without ficoll is 3.3 ns and with ficoll (correlation time of the rotational dynamics of the side chain and backbone) is 4.5 ns (Equations 2 and 3). The CCM1 94R1 nitroxide correlation time with and without ficoll can be substituted into $(1/\tau_s + 1/\tau_b$, and $1/\tau_n$, respectively, in Equation 1, giving an experimentally-determined CCM1 tumbling correlation time (τ_p) of 12.4 ns. The estimated correlation time range (Cavanagh et al., 2007) for the CCM1 monomer is 5.5 – 7.1 ns and 9.8 – 14.0 ns for the dimer based on a monomeric molecular weight of 12 kDa. Therefore, the protein correlation time determined with CW EPR experiments is consistent with a CCM1 dimer. Higher-order CCM1 oligomers (tetramers) are not apparent from CW EPR experiments because the overall tetramer correlation time is theoretically > 20 ns, and the line shape would not be expected to change with ficoll addition.

To compare with CCM1 EPR spectra, a subset of residues was chosen to spin label in CCM1 with single GlcNAc (gCCM1) and complex glycan (cgCCM1) modifications (Figure 1C; Supplemental Figure S4B and C). The EPR line shapes for CCM1, gCCM1, and cgCCM1 are very similar and upon addition of ficoll, decrease in overall mobility (Figure 3C; Supplemental Figure S4B) with the same magnitude, indicating that CCM1 glycoforms are identical oligomeric states – specifically, they are dimers. The similarity of the line shapes for the labeled gCCM1 and cgCCM1 constructs with and without ficoll compared to CCM1 indicates that different glycosylation patterns do not alter the oligomeric state of the protein. Lineshapes of cgCCM1 5R1 (Supplementary Figure S4) are less affected by ficoll than gCCM1 and CCM1, which is consistent with the higher molecular weight and slower rotational diffusion of the fully glycosylated protein.

CCM1 forms GFCC′C″-interface dimers in solution

To investigate the structure of the CCM1 dimers, which is evident in both glycosylated and non-glycosylated constructs, DEER-derived distances were determined for each of the labeled residues. DEER spectroscopy, like CW EPR, is well-suited for distinguishing between CCM1 monomeric and dimeric species because monomers are expected to have no DEER signal, and different CCM1 dimers will have unique DEER distance distributions (Supplemental Figure S5A, B and C). The background-corrected dipolar evolution functions (DEFs; Jeschke and Polyhach, 2007) and the corresponding distance distributions derived from the DEFs (Jeschke et al., 2006) for each of the spin-labeled CCM1 sites are shown in Figure 4 and Supplemental Figure S6. All ten spin-labeled CCM1 proteins produced DEER signals, agreeing with CW EPR results and previously established CCM1 dimer studies (Bonsor et al., 2015b; Zhuo et al., 2016). High populations of CCM1 dimers are recapitulated in size-exclusion chromatography (Supplemental Figure S7A), which is evident from comparing retention times of CCM1 to CCM3 and CCM8 (both known to be monomeric by analytical ultracentrifugation; Bonsor et al., 2015b; Bonsor et al., 2018); Supplemental Table S1). To demonstrate that DEER can appropriately distinguish between monomeric and dimeric CCMs using the methods in this study, CCM3 was spin-labeled and produces no DEER signal as expected for a monomer (Supplemental Figure S7B and C).

Reliability of distance distributions in the four-pulse DEER experiment depends on the maximum dipolar evolution time, t_{max}, which is observed in the decay time of dipolar oscillations in the DEF (Jeschke, 2012). An upper limit of 5.0 nm can reliably be achieved with t_{max} = 2 μs, scaling as $t_{max}^{1/3}$. All DEER data were obtained with $t_{max} \geq 3$ μs, corresponding to an upper limit of ~ 5.7 nm for an accurate mean distance. With the exception of A12R1 and E16R1 (Supplemental Figure S6), all mean distances are ≤ 5.2 nm and fall within the reliable limit for data collected out to ~ 3 μs. Although non-zero DEF signals were observed for A12R1 and E16R1, these distances were greater than 6.0 nm and require $t_{max} > 3.5$ μs for generating a reliable distance distribution. Most distance distributions are unimodal with widths ranging from ~ 0.5 – 2.0 nm. Many of the narrower distributions correspond to more immobilized spin labels as assessed by CW EPR line shapes (e.g. 94R1, 56R1, 20R1, and 9R1), and the broader distributions (e.g. 5R1) correspond to residues with more mobile line shapes (Supplemental Figure S4) indicating the distance distribution widths are due to conformations sampled by side chain rotamers.

Two approaches were used to investigate if CCM1 ABED (PDB ID: 2GK2) or GFCC′C″-interface (PDB ID: 4QXW) dimers are supported by the DEER-derived distance distributions. First, using Multiscale Modeling of Macromolecules (MMM; (Kamisetty et al., 2013), predicted distance distributions were calculated by modeling MTSL at each spin-labeled site for both dimer structures and compared to experimental distance distributions. All ten DEER distance distributions were consistent with modeled distributions of the GFCC′C″ CCM1 dimer (Figure 5A and Supplemental Figure S8) whereas only 3 experimental DEER distributions (T56R1, T83R1, and Q103R1) matched those calculated with the ABED dimer (Figure 5B and Supplemental Figure S9). At position 7R1 in the ABED dimer, *in silico* modeling with MTSL did not produce any possible rotameric states from the MMM spin label library due to restricted conformational space (Supplemental

Figure S9). These data support that one predominant dimer structure is observed in solution, specifically a GFCC′C″-interface dimer. If a mixture of both GFCC′C″ and ABED CCM1 dimers existed in solution (at sufficiently high concentrations), two distance populations might be expected for E5R1, M7R1, F9R1, L20R1, and D94R1 as these positions deviate in simulated DEER distributions for each dimer structure (Supplementary Figures S5D, S8 and S9). Additionally, A12R1 and E16R1 DEFs would possess evidence of short distance contributions, which is not observed (Supplemental Figure S6).

The second approach performed simulated annealing of each dimer model with DEER-derived distance restraints using Xplor-NIH (Schwieters et al., 2003). Each resulting DEER-refined structure was compared to the parent crystal structure to assess which CCM1 dimer is supported by experimental DEER data. Simulated annealing of the CCM1 GFCC′C″ dimer with the DEER restraints resulted in an average structure that closely resembles the parent structure (Figure 6A). The average backbone RMSD between the DEER-refined and parent GFCC′C″ structures was 0.4 ± 0.1 Å with no distance restraint violations (Table 2). In contrast, refinement of the CCM1 ABED dimer using the same DEER restraints and simulated annealing protocol produced an average structure that deviates significantly from the protomer orientations in the parent structure (Figure 6B). The backbone RMSD of the refined ABED dimer to its parent was 13.0 ± 0.3 Å with seven DEER restraint violations. A maximum distance restraint violation of 17.6 Å was observed for CCM1 ABED dimer refinement and corresponded to the T83R1 distance restraint (Table 2).

CCM1 homodimer conformers are consistent between glycoforms

To determine the glycosylated gCCM1 dimer conformation in solution, DEER distributions were obtained for E5R1 and Q103R1 gCCM1 and cgCCM1 and compared with non-glycosylated CCM1 measurements (Figure 7). Glycosylation, for either gCCM1 or cgCCM1 spin-labeled mutant, did not change the mean DEER distribution with respect to non-glycosylated CCM1, with E5R1 gCCM1 and cgCCM1 producing mean distances of 5.0 nm and 5.1 nm, respectively (Figure 7A) and Q103R1 gCCM1 and cgCCM1 producing mean distances of 3.4 nm and 3.3 nm, respectively (Figure 7B). Furthermore, DEER distributions and CW-EPR spectra for Q103R1 and E5R1 are consistent across all glycosylation states of (CCM1, gCCM1 and cgCCM1; Figure 7B and Figure 3C). Since glycosylated CCM1 has the same DEER distributions as CCM1, this suggests that the GFCC′C″-interface dimer is the dominant species in solution with or without glycan modifications. Previous reports of glycosylation modulating CCM1 dimers noted that the phenomenon was difficult to rationalize, as N70, N77, and N81 are located far from the GFCC′C″ interface (Zhuo et al., 2016). Furthermore, it is important to note that a correction of Zhou et al. (Zhuo et al., 2016) stated that results characterizing a glycosylated monomer in solution by NMR were difficult to reproduce and that more recent samples showed evidence of dimerization even with different glycosylation states. Results presented here are consistent with this correction.

Aside from CCM1, only CCM6-CCM8 heterodimerization studies have been performed with glycosylated constructs *in vitro*. The CCM6-CCM8 heterodimer (PDB ID: 4YIQ) and CCM1 homodimer (4QXW) rmsd is 1.2 Å (Supplemental Table S3), and the K_D of each complex is ~2 and 0.450 μM, respectively ((Bonsor et al., 2015b); Supplemental Table S1).

Glycosylated CCM6-CCM8 was reported to decrease the heterodimer affinity roughly two-fold (4.5 μM; Bonsor et al., 2015b; Supplemental Table S1). A similar reduction in CCM1 homodimer affinity (CCM1 K_D = 2.8 μM to K_D ~ 6-7 μM with glycosylation; Supplemental Table S1) would result in roughly 7% glycosylated CCM1 monomer at concentrations used for NMR and EPR experiments (−100 μM). Seven percent CCM1 under conditions used for this study is not expected to noticeably change the observed DEER DEF echo intensities and EPR line shapes.

In vivo, Mammalian glycosylation is highly regulated and important for immune recognition, glycan-dependent protein binding, and cell signaling events (Marth and Grewal, 2008). CCM1 GFCC′C″ dimers importantly have glycans exposed and may promote specific types of CEACAM1 association *in vivo*. CEACAM1 has been reported to form dimers *in vivo* both in *cis* (apical cell surface) and *trans* (between cells), with GFCC′C″ interface mutations abrogating CEACAM1 *trans*-homophilic binding (Patel et al., 2013). Thus, solution-based studies of glycosylated CCM1 *in vitro* support observations of *trans* binding between HeLa cells expressing CEACAM1. However, electron tomography structures of soluble rat CEACAM1 containing all four extracellular Ig domains reported a minor population of dimers distinct from the major GFCC′C″ interface species (Klaile et al., 2009) that was not captured by DEER spectroscopy. The minor CEACAM1 dimer was identified based on in-parallel, close binding between three or all four Ig domains. Interestingly, the authors report enhanced formation of this alternative CEACAM1 dimer upon membrane attachment in which CEACAM1 clusters were greater on free liposome surfaces (*cis*-type). As the CEACAM1 N-terminal domain alone was assessed in this study, additional CEACAM1 domains and/or membrane tethering may be necessary to observe the minor dimer population (possibly ABED dimers) reported (Klaile et al., 2009).

GFCC′C″ interface residue contacts are predicted from co-evolution patterns

Due to the prevalence of CCM1 GFCC′C″ interface dimers observed both *in vitro* and *in vivo*, the evolutionary basis for its formation across the CCM protein family was investigated by co-evolution based prediction of residue-residue contacts with the GREMLIN server (Anishchenko et al., 2017; Kamisetty et al., 2013). Co-evolution measurements of CCM proteins have not been described to date and may be important for identifying GFCC′C″ interface contacts that are relevant to dimer formation and evolutionarily selected for, supplementing experimental CCM1 mutagenesis studies. Using the monomeric CCM1 protein sequence, contact predictions were generated from conservation and co-evolution patterns based on a multiple-sequence alignment (MSA) containing 5,964 homologous, non-redundant (no pair of sequences > 90% identical) protein sequences. Contact predictions are most accurate when the number of sequences in the MSA is at least five times the query sequence length (Kamisetty et al., 2013). Given the CCM1 query sequence length of 106 amino acids, the sequence/length parameter was well above the minimum value required for accurate contact predictions.

The evolutionarily-coupled residues (ECs) within the CCM1 sequence predicted to be most probable by the GREMLIN server are provided in Supplemental Table S5. Each EC pair was mapped onto both the CCM1 GFCC′C″ and ABED dimer structures. To identify EC pairs at

the dimer interface, a 5 Å distance separation cutoff between residues in opposite protomers was applied using terminal side chain atoms. Based on the initial filter, Figure 8A shows the following proposed GFCC′C″ interface residues that are evolutionarily coupled: Y34/Q89 (prob. = 0.999), Y34/E99 (prob. = 0.551), and G41/N97 (prob. = 0.393). EC pairs identified across the CCM1 ABED interface (Figure 8B) are L18/N70 (prob. = 0.912) and L18/Y68 (prob. = 0.848). To distinguish between intraprotomer ECs and those specific to GFCC′C″ and ABED interfaces, each EC pair was additionally mapped onto the same protomer in CCM1 crystal structures (Figure 8). Using an intraprotomer distance separation cutoff of 5 Å between terminal side chain atoms, only Y34/E99 and G41/N97 EC pairs specific to the CCM1 GFCC′C″ interface could be identified as residues that are coupled and at a dimer interface (rather than a tertiary contact within the CCM1 monomer fold).

Basing meaningful interpretations of CCM1 co-evolved dimer interface residues on EC scores alone is difficult. However, extensive mutagenesis studies of CCM proteins both *in vivo* and *in vitro* have been performed to date, which crucially allows assessment of CCM1 ECs using both bioinformatics data and published experimental observations. For example, neither Y34 nor Q89 play roles in CEACAM1 homophilic adhesion (Watt et al., 2001) and dimer formation (Bonsor et al., 2018) even though the pair's EC probability score is considered significant (0.999). However, Y34 in CEACAM5, but not the analogous residue to Q89 (H89; Korotkova et al., 2008), is suggested to be important for homophilic adhesion (Taheri et al., 2000). Mutagenesis studies of Y34 and Q89 combined with mapping the EC pair onto the CCM1 structure thus supports that this pair has likely not co-evolved to form the dimer interface. However, both Y34 and Q89 are crucial residues for binding Opa proteins, Dr adhesins, and HopQ proteins expressed by *Neisseria* bacteria (Virji et al., 1999), pathogenic *Escherichia coli* (Korotkova et al., 2008), and *Helicobacter pylori* (Bonsor et al., 2018), respectively.

The proposed CCM1 dimer interface pair Y34/E99 is highly conserved among the CCM protein family (Supplemental Figure S1), except for CCM4 (H34 and Q99). E99 point mutations have only been performed with recombinantly-expressed CCM5, in which the E99A mutation converts CCM5 from a dimer to a monomer in solution but does not impact Dr-adhesin binding (Korotkova et al., 2008). It is likely that CCM1 E99 plays a similar role in dimer formation, as both CCM1 and CCM5 are highly related in terms of residues important for CCM recognition, homotypic and heterotypic interactions, as well as homodimer affinities (Korotkova et al., 2008). Finally, the EC pair G41/N97 is the most convincing CCM dimer interface pair based on structure mapping, albeit with a low EC probability score (0.393). Recently, mutation of CCM1 N97 to alanine was shown to convert the protein from dimer to monomer (Bonsor et al., 2018). G41 and N97 additionally form hydrophilic contacts at the CCM5 dimer interface (Korotkova et al., 2008). While N97A appears to disrupt the CCM1 homodimer in solution, the mutation has little effect on binding HopQ (Bonsor et al., 2018). CCM G41 on the other hand plays an important role in binding select Opa proteins (Virji et al., 1999) and HopQ (Bonsor et al., 2018), though G41 mutations have not been explored in CCM homodimerization. Combined, data from evolutionary coupling analysis and previous CCM literature support that Y34/E99 and G41/N97 are residue pairs that have likely co-evolved to facilitate CCM1 homodimer interactions. Additionally, E99 and N97 may be critical for maintaining the CCM1, and

possibly CCM5, dimer interface, strengthening experimental observations for CCM1 and CCM5 specificity (Korotkova et al., 2008).

Mapping single-nucleotide variations (SNVs) in cancer

Although not confirmed to cause disease, 21 residues associated with SNVs in different cancers (Dingerdissen et al., 2018; Pan et al., 2014; Wu et al., 2014) are mapped to the N-domain of CEACAM1 (Supplementary Figure S10). Of these mutations, eight map to the dimeric interface (red spheres in Supplementary FigureS10). Based on the resulting functional groups, many of the mutations would likely disrupt dimer formation. One SNV found in liver cancer mutates one of the evolutionary coupled resides, Q89, to a lysine residue. In the structure, these residues point directly at each other and are occluded from solvent. Thus, it is very likely the Q89K mutation would prevent dimer formation. Similarly, G51E and E37K, both found in melanoma, would introduce repulsive charges at the dimer interface. One SNV variant found in uterine cancer results in the mutation R43C, which may result in the loss of a repulsion between the dimer, which may enhance the association of the dimer. The specific effect of each SNV on the monomer-dimer transition and on the biology of CEACAM1 is a rich area for future studies.

STAR METHODS

Resource Availability

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Linda Columbus (columbus@virginia.edu).

Materials availability—This study did not generate unique reagents. Plasmids for expression of N-terminal CEACAM1 domains (WT and all cysteine mutants) in *E. coli* (pGEX-2T) and in HEK293 cells (pGen2) are available upon request to the lead contact.

Data and code availability—Coordinates and structure factors for the structure of minimally glycosylated N-terminal CEACAM1 domain have been deposited in the Protein Data Bank under the accession code PDB ID 7MU8 and are publicly available as of the date of publication. Further data for this manuscript are available in the supporting materials document. All data reported in this paper will be shared by the lead contact upon request. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental model and subject details

For protein expression in *E. coli*, MC1061 strain was used. This bacteria was transformed with pGEX-2T plasmid containing the N-terminal CEACAM1 domain construct (WT or containing cysteine mutants). Cells were grown at 37°C in LB medium containing 100 μg/ml ampicillin and 100 μg/ml streptomycin. For protein expression in mammalian cells, HEK293S or Freestyle™ 293-F cells were used. These cells were transfected using PEI with pGen2 plasmid containing the GFP-fused N-terminal CEACAM1 domain construct (WT or containing cysteine mutants). HEK cells were maintained in suspension at 37°C

and 8% CO_2 in FreeStyle™ 293 expression medium. Transfections were carried out in a mixture of FreeStyle™ 293 expression medium and EX-CELL media at a ratio of 9:1 at the same temperature and CO_2 conditions. Expression and purification details are found in the Method Details section.

Method Details

CCM1 mutagenesis, expression and purification.—A pGEX-2T vector containing the N-terminal D1 domain of the human *CEACAM1* gene (amino acids 35-141 of the full-length protein) was generously provided by Rob Nicholas (University of North Carolina at Chapel Hill) for the production of non-glycosylated protein. Glycosylated protein was produced using a pGEn2 expression vector (Moremen et al., 2018) containing a N-terminal sequence that signals protein secretion into the medium, followed by His_8 tag, AviTag, green fluorescent protein (GFP), TEV cleavage site, and sequence encoding the N-terminal domain of CEACAM1 (CCM1, residues 34-141, UniProt P13688; (Zhuo et al., 2016). All cysteine mutations were introduced using PIPE mutagenesis (Klock and Lesley, 2009) using the *Pfu* Turbo DNA Polymerase (Agilent). The sequences of the primers used are provided in Supplemental Table S6. Mutations were confirmed using gene sequencing (Genewiz Inc, South Plainfield, NJ).

Expression and purification of non-glycosylated CCM1 was performed as previously described (Martin et al., 2016) having been adapted from (Fedarovich et al., 2006). Plasmids containing the mutated *CEACAM1* gene were transformed into MC1061 *E. coli* cells, which were grown in LB media to an OD_{600} of 0.6. Protein expression was induced with 1 mM IPTG overnight. Cells were harvested, resuspended in lysis buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 2 mM ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT), and 10% glycerol), and lysed with a microfluidizer (Microfluidics model 110L, Newton, MA). Cell debris was removed via centrifugation, and the addition of ammonium sulfate to 55% precipitated proteins within the supernatant. These proteins were pelleted and resuspended in 20 mM Tris pH 7.3, 150 mM NaCl, 2 mM DTT, and 10% glycerol at 4°C. CCM1 was purified using a glutathione resin column at 4°C, eluting in 20 mM Tris, pH 7.3, 150 mM NaCl, 2 mM DTT, 10% glycerol, and 10 mM reduced glutathione. The GST tag was cleaved from CCM1 using tobacco etch virus (TEV) protease (at approximately 3.5 μM), which was added to the eluent and dialyzed overnight at 4°C against 20 mM Tris pH 7.3, 150 mM NaCl, 10% glycerol, and 2 mM DTT. CCM1 was isolated from TEV and cleaved GST using a HR Sephacryl S-200 Gel Filtration column (26/60 mm, GE Healthcare) equilibrated with 20 mM Tris, pH 8.0, 500 mM NaCl, 10% glycerol, and 2 mM DTT. SDS-PAGE was utilized to analyze CCM1 purity within the eluted fractions.

An expression and purification protocol for glycosylated CCM1 (gCCM1) proteins has been adapted from (Zhuo et al., 2016). HEK293S GnTI$^-$ cells (Reeves et al., 2002), expressing primarily $Man_5GlcNAc_2$-Asn glycans, were maintained using FreeStyle™ 293 expression medium (Thermo Fisher Scientific) in a humidified CO_2 platform shaker incubator at 37°C. A 250 mL suspension culture of HEK293S (GnT1$^-$) cells in a 9:1 ratio of FreeStyle^T 293 media and EX-CELL media (Sigma) was transfected with the CCM1-pGEn2 plasmid using polyethyleneimine (PEI; Polysciences, Inc., Warrington, PA) as described previously

(Zhuo et al., 2016). After incubating for 24 hours, 250 mL of a 9:1 ratio of FreeStyle^{T} 293: EX-CELL medias and 2.2 mM valproic acid (Sigma) were added to the suspension culture. Glycosylated CCM1 proteins were produced over the course of five days at 37°C, after which cell debris was removed via centrifugation (20 min, 150 x g, 4°C). Glycosylated CCM1 proteins were purified from the supernatant via Co^{2+} immobilized metal affinity chromatography (IMAC), eluting in ten column volumes of elution buffer (25 mM HEPES, 300 mM NaCl, 680 mM imidazole, pH 7.0) at 4°C. The eluent was dialyzed into 4L of 25 mM HEPES, 300 mM NaCl, 10% glycerol, pH 7.0 containing approximately 3.5 μM TEV and endoglycosidase F1 (both enzymes were expressed in BL21(DE3) *E. coli* and purified via Co^{2+} IMAC; (Meng et al., 2013)), removing the GFP tag from CCM1 and truncating the glycans to single GlcNAc residues on CCM1 (gCCM1). Excess GFP was removed using Co^{2+} IMAC, where the flow-through containing gCCM1 proteins was collected.

The final step of all purifications was size exclusion chromatography using an HR Sephacryl S-200 column (26/60 mm, GE Healthcare) equilibrated with 20 mM Tris, pH 8.0, 500 mM NaCl, 10% glycerol, and 2 mM DTT.

Expression and purification of cgCCM1.—An expression and purification protocol for CCM1 proteins with complex glycan modifications (cgCCM1) has been adapted (Zhuo et al., 2016) using a similar approach described above for preparation of gCCM1, with the following modifications: FreeStyle™ 293-F cells (ThermoFisher Scientific; which express complex glycans), rather than HEK293S GnTI^{-} cells, were used. Transfection of FreeStyle™ 293-F cells with the CCM1-pGEn2 plasmid, protein expression, and IMAC purification were carried out as for gCCM1 production described above. The cgCCM1 IMAC eluent was simultaneously cleaved with 3.5 μM TEV protease to remove the GFP tag and dialyzed against 4L of 25 mM HEPES, 300 mM NaCl, 10% glycerol, pH 7.0, and 2mM DTT. Further polishing steps, spin labeling, and final sample concentrations were conducted as described for gCCM1 excluding size exclusion chromatography.

Crystallization of gCCM1.—For initial screening for crystals, protein fractions resulting from size exclusion chromatography were concentrated to 6 and 12 mg/ml using a centrifugal concentrator with a 10 kDa cutoff (Amicon) at 10,000 x g at 4°C. The concentrated protein sample was centrifuged at 16,100 x g at 4°C for 15 minutes to clear any insoluble aggregates. Using the Mosquito (TTP Labtech) liquid handling robot, 100 nl of protein was mixed with 100 nl of precipitant solution (from commercial screens) on Intelli-plate 96-3 sitting drop plates (Hampton Research). Commercial screens used include JCSG Core I, JCSG Core II, JCSG Core III, and JCSG Core IV from Qiagen and PEG/Ion and Crystal Screen from Hampton Research. Initial crystal hits were screened for diffraction at an in-house diffractometer (Bruker Kappa APEXII Duo) at the University of Virginia Nanoscale Materials Characterization Facility Core. Optimization screens were made by varying pH, buffer identity and concentration of precipitant components of the initial crystal hit conditions. When ready to freeze and ship to the synchrotron crystals were harvested using 100-400 μm Mounted CryoLoops (Hampton Research). The crystallization condition + 20% glycerol was used as cryoprotectant. This yielded well-diffracting crystals of *gCCM1* (Supplemental Figure S2A, B and C). The crystal that yielded the structure presented were

grown in the following precipitant condition: 100 mM MES, pH 6.5, 2M Ammonium sulfate.

Data collection and structure determination.—All data were collected at SER-CAT beamline 22-ID at the Advanced Photon Source at Argonne National Laboratory (Lemont, IL, USA). Data were indexed, integrated and scaled with XDS (Kabsch, 2010) and AIMLESS (Evans and Murshudov, 2013). Data were phased with molecular replacement using PHASER (McCoy et al., 2007) with the GFCC′C″ dimer as a search model (PDB ID: **4QXW**). The model was completed in Coot (Emsley et al., 2010). Multiple rounds of model building in Coot and refinement with PHENIX (Adams et al., 2010) were carried out to a final R_{work}/R_{free} of 0.2022/0.2172. Complete data collection and refinement statistics can be found in Table 1.

Spin labeling.—DTT was removed from pure protein with a PD-10 column (GE Healthcare), eluting with buffer containing 20 mM sodium phosphate, 150 mM NaCl, and 10% glycerol directly into five molar excess S-(2, 2, 5, 5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL; Toronto Research Chemicals Inc., Toronto, Canada). The reaction was incubated overnight in the dark at 4°C. Excess spin label was removed with a second PD10 column. Resulting protein was concentrated to 100-200 μM.

Continuous-wave EPR.—CW-EPR experiments were measured using an X-band Bruker EMX continuous wave spectrometer with an ER4123D dielectric resonator (Bruker Biospin, Billerica, MA) at room temperature using a 100 G sweep width. For samples containing ficoll, 50% w/v Ficoll® PM 70 (Sigma) in 20 mM sodium phosphate, 150 mM NaCl, 10% glycerol buffer was prepared. Protein samples were mixed 1:1 with the 50% w/v ficoll solution, such that the final concentration was 25% w/v ficoll. Spectra were recorded using 5 μL of each sample (100 – 300 μM) in pyrex capillaries (0.6 mm id x 0.84 od, Vitrocom, Mountain Lakes, NJ). Spectra were baseline corrected and normalized using Lab-VIEW software (provided by C. Altenbach, University of California at Los Angeles).

Double-electron electron resonance spectroscopy.—Double-labeled samples were measured using pulsed EPR with a Q-band Bruker E580 Spectrometer fitted with an ER5106-QT flexline resonator (Bruker Biospin) at 80 K. All samples were prepared to a final protein concentration between approximately 100 and 200 μM with 10% deuterated glycerol. Samples were loaded into quartz capillaries with a 1.6 mm od x 1.1 mm id (Vitrocom) and were flash frozen in liquid nitrogen. Dipolar evolution data were processed using DEERAnalysis2016 using Tikhonov regularization to generate distance distributions (Jeschke et al., 2006). Background subtraction of distance distribution yields error at each distance which is plotted as ranges that represent fits within 15% root-mean-square-deviation of the best fit.

Pulse sequence for DEER EPR data acquisition.—A four-pulse DEER sequence was used with one 16 ns $\pi/2$, two 32 ns π observed pulses (at an observed frequency υ_1), and a π pump pulse (at a frequency υ_2) optimized at approximately 32 ns (Pannier et al., 2000). A pump frequency (υ_2) was set at the maximum of the nitroxide spectrum and

the observed frequency (υ_1) was set 75 MHz lower. Increasing inter-pulse delays at 16 ns increments were utilized with a 16-step phase cycle during data collection. Accumulation times were typically between 18 and 24 hours, with a dipolar evolution time between 2 and 3 μs.

DEER distribution simulations.—CCM1 crystal structures representing GFCC′C″ (PDB ID: 4QXW(Huang et al., 2014) and ABED (PDB ID: 2GK2; (Fedarovich et al., 2006) dimer interfaces were selected for structure refinement, with water and ligand molecules removed. To generate structures of CCM1 dimers labeled with MTSL, Multi scale Modeling of Macromolecules (MMM) (Kamisetty et al., 2013) was used with implementation of the spin label rotamer library approach (Polyhach et al., 2011). DEER distributions were simulated using MMM for all experimental CCM1 DEER sites. Initial simulated DEER distributions were modeled using the predicted R1 rotamer distributions for each dimer pair. Rotamer pairs from the R1 ensembles were then used to determine the best fit of simulated to experimental data.

Model preparation for CCM1 simulated annealing.—For XPLOR-NIH refinement, the lowest-energy MTSL rotamers for all R1 labels were modeled onto CCM1 GFCC′C″ and ABED dimer structures, yielding models with 9 MTSL labels each. Position 9R1 was removed from the GFCC′C″ dimer structure to prevent clashing with 7R1. Removal of 9R1 was not required for the ABED structure as 7R1 could not be modeled with MMM (Supplemental Figure S9). Ensemble statistics improved when using distance restraints between modeled R1 nitroxide moieties compared to C_β atoms. Since DEER restraints are derived from R1 nitroxide moieties, extending up to ~8 Å from the backbone, utilizing CCM1 modeled with R1 labels was therefore preferable to unmodified crystal structures.

Simulated annealing of CCM1 dimeric structures.—Two separate rounds of structure refinement in Xplor-NIH (Schwieters et al., 2003) were performed for GFCC′C″ and ABED CCM1 parent structures: (1) with spin-labeled CCM1 from MMM and (2) with unmodified CCM1. Simulated annealing protocol was adapted from (Sarver et al., 2018). 200 refined structures were generated from each CCM1 parent structure. CCM1 structural ensembles were produced from the 20 lowest energy structures. Statistics for CCM1 ensemble restraint violations and root mean square deviation (RMSD) calculations from MOLMOL (Koradi et al., 1996) are provided in Table 2.

CCM1 dimer contact predictions.—Evolutionarily selected residues that may be important in stabilizing the solution CCM1 dimer structure were identified using the GREMLIN server (Balakrishnan et al., 2011; Kamisetty et al., 2013) using the monomeric CCM1 amino acid sequence as input. Residue contacts predicted from the GREMLIN server were mapped onto the GFCC′C″ crystal structure to determine which contacts may be formed specifically across the CCM1 dimer interface, but not within the same protomer. GFCC′C″ interface-specific pairs were mapped onto the ABED crystal structure for comparison. Electrostatic-dependent interfacial contacts that may play a role in defining the CCM1 dimer orientation were additionally identified from structure mapping. These

contacts are referenced against GFCC′C″ electrostatic potential maps at pH 7 that were generated using the APBS PyMol plugin.

Quantification and Statistical Analysis

Crystallography.—Statistical data of refinement in Table 1 for the crystal structure were obtained from the outputs of PHENIX (Adams et al., 2010).

DEER data processing.—Dipolar evolution data were processed using DEERAnalysis2016 using Tikhonov regularization to generate distance distributions (Jeschke et al., 2006). Error bars in the distance distributions represent uncertainty in the subtraction of the intermolecular background signal that produces fits within 15% root-mean-square-deviation of the best fit.

Calculation of correlation times for CCM1.—At microwave frequencies used in this study (X-band), the nuclear hyperfine splitting determines the rotational range of ~ 0.1 to 40 ns motions (Columbus and Hubbell, 2002; Nesmelov and Thomas, 2010). The nitroxide correlation time (τ_n) is related to the individual dynamic modes by the following equation:

$$\frac{1}{\tau_n} = \frac{1}{\tau_s} + \frac{1}{\tau_b} + \frac{1}{\tau_p} \quad \text{Equation 1}$$

where τ_s is the correlation time of the side chain rotations, τ_b is the correlation time resulting from backbone dynamics, and τ_p is the overall protein correlation time (Freed, 1976). The relationship assumes that dynamics are not correlated. Overall, protein correlation times greater than ~20 ns do not contribute to the EPR line shape (Lopez et al., 2009). Both the CCM1 monomer (11.8 kDa) and dimer (23.6 kDa) are predicted to have correlation times (Cavanagh et al., 2007) less than 20 ns and, therefore, contribute to the EPR line shapes. For proteins with correlation times less than ~20 ns, a viscous agent such as ficoll or sucrose can be used to increase the correlation time of the protein so as not to contribute to the EPR line shape (Lopez et al., 2009; McHaourab et al., 1996). The addition of a viscous agent on the line shape typically manifests as spectral component broadening (Lopez et al., 2009), of which the magnitude will be increased for CCM1 monomers compared to dimers.

The hyperfine splitting present in the 94R1 CW EPR spectrum alone ($2A_{zz}'$, Figure 2B) can additionally be used to quantify the nitroxide correlation time (τ_n; (Freed, 1976)) with the relationship in Equation 1. CCM1 EPR line shapes without ficoll contain all contributions, while those with ficoll report solely on τ_s and τ_b. Therefore, CCM1 τ_p can be determined from calculating τ_n based on the hyperfine splitting ($2A_{zz}'$) quantified from 94R1 EPR spectra recorded with and without ficoll. Nitroxide correlation times (τ_n) can be estimated from experimental $2A_{zz}'$ values uing the relationship described by Freed (Freed, 1976)

$$\tau_n = a \cdot (1 - S)^b \quad \text{Equation 2}$$

Where constants $a = 5.4 \times 10^{-10}$ and $b = -1.36$ are given for a protein tumbling according to the Brownian diffusion model and a Lorentzian line width of 3 G. A ratio (S) of the measured splitting ($A_{zz}e$) to the maximum splitting for R1 ($A_{zz} = 37$ G) is given by:

$$S = \frac{\left(\frac{2A_{zz}'}{2}\right)}{A_{zz}} = \frac{A_{zz}e}{A_{zz}} \quad \text{Equation 3}$$

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the National Institutes of Health (NIH) National Institute of General Medical Sciences (NIGMS) under grants R01 GM087828 (L.C.), R35 GM131829 (L.C.), P41 GM103390 (K.M. and J.P.), R01 GM033225 (K.M. and J.P.), T32 GM008715 (N.S.), F32 GM136076 (M.B.D), and by the National Science Foundation Graduate Research Fellowship under grant DDGE-1315231 (N.S.). This work also used the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamlines at the Advanced Photon Source (APS), Argonne National Laboratory. SER-CAT is supported by its member institutions and equipment grants from the NIH (S10 RR25528, S10 RR028976, and S10 OD027000). APS is a U.S. Department of Energy (DOE) Office of Science User Facility operated by Argonne National Laboratory with U.S. DOE support under Contract No. W-31-109-Eng-38. We especially thank Norma Duke and Unmesh Chinte from SER-CAT for their helpful contributions regarding beamline operation and data collection. The Bruker Kappa APEXII Duo X-ray diffractometer within University of Virginia's Nanoscale Materials Characterization Facility (NMCF) was used for this work, and we thank Diane Dickie for her contribution in equipment training and operation.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221. [PubMed: 20124702]

Anishchenko I, Ovchinnikov S, Kamisetty H, and Baker D (2017). Origins of coevolution between residues distant in protein 3D structures. Proceedings of the National Academy of Sciences 114, 9122.

Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, and Notredame C (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res 34, W604–608. [PubMed: 16845081]

Balakrishnan S, Kamisetty H, Carbonell JG, Lee S-I, and Langmead CJ (2011). Learning generative models for protein fold families. Proteins: Structure, Function, and Bioinformatics 79, 1061–1078.

Beauchemin N, and Arabzadeh A (2013). Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. Cancer and Metastasis Reviews 32, 643–671. [PubMed: 23903773]

Bonsor DA, Beckett D, and Sundberg EJ (2015a). Structure of the N-terminal dimerization domain of CEACAM7. Acta Crystallogr F Struct Biol Commun 71, 1169–1175. [PubMed: 26323304]

Bonsor DA, Gunther S, Beadenkopf R, Beckett D, and Sundberg EJ (2015b). Diverse oligomeric states of CEACAM IgV domains. Proc Natl Acad Sci U S A 112, 13561–13566. [PubMed: 26483485]

Bonsor DA, Zhao Q, Schmidinger B, Weiss E, Wang J, Deredge D, Beadenkopf R, Dow B, Fischer W, Beckett D, et al. (2018). The *Helicobacter pylori* adhesin protein HopQ exploits the dimer interface of human CEACAMs to facilitate translocation of the oncoprotein CagA. The EMBO Journal 37.

Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, and Skelton NJ (2007). CHAPTER 1 - CLASSICAL NMR SPECTROSCOPY. In Protein NMR Spectroscopy (Second Edition), Cavanagh

J, Fairbrother WJ, Palmer AG, Rance M, and Skelton NJ, eds. (Burlington: Academic Press), pp. 1–28.

Columbus L, and Hubbell WL (2002). A new spin on protein dynamics. Trends in Biochemical Sciences 27, 288–295. [PubMed: 12069788]

Cornish VW, Benson DR, Altenbach CA, Hideg K, Hubbell WL, and Schultz PG (1994). Site-specific incorporation of biophysical probes into proteins. Proceedings of the National Academy of Sciences 91, 2910.

Dingerdissen HM, Torcivia-Rodriguez J, Hu Y, Chang TC, Mazumder R, and Kahsay R (2018). BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. Nucleic Acids Res 46, D1128–D1136. [PubMed: 30053270]

Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501. [PubMed: 20383002]

Evans PR, and Murshudov GN (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69, 1204–1214. [PubMed: 23793146]

Fedarovich A, Tomberg J, Nicholas RA, and Davies C (2006). Structure of the N-terminal domain of human CEACAM1: binding target of the opacity proteins during invasion of Neisseria meningitidis and N. gonorrhoeae. Acta Crystallogr D Biol Crystallogr 62, 971–979. [PubMed: 16929097]

Freed JH (1976). Theory of Slow Tumbling ESR Spectra of Nitroxides. In Spin Labeling Theory and Applications, Berliner LJ, ed. (New York: Academic Press), pp. 53–132.

Gandhi AK, Sun ZJ, Kim WM, Huang YH, Kondo Y, Bonsor DA, Sundberg EJ, Wagner G, Kuchroo VK, Petsko GA, et al. (2021). Structural basis of the dynamic human CEACAM1 monomer-dimer equilibrium. Commun Biol 4, 360. [PubMed: 33742094]

Huang Y-H, Zhu C, Kondo Y, Anderson AC, Gandhi A, Russell A, Dougan SK, Petersen B-S, Melum E, Pertel T, et al. (2014). CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. Nature 517, 386. [PubMed: 25363763]

Hubbell WL, Cafiso DS, and Altenbach C (2000). Identifying conformational changes with site-directed spin labeling. Nature Structural Biology 7, 735–739. [PubMed: 10966640]

Jeschke G (2012). DEER Distance Measurements on Proteins. Annual Review of Physical Chemistry 63, 419–446.

Jeschke G, Chechik V, Ionita P, Godt A, Zimmermann H, Banham J, Timmel CR, Hilger D, and Jung H (2006). DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. Appl Magn Reson 30, 473–498.

Jeschke G, and Polyhach Y (2007). Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. Phys Chem Chem Phys 9, 1895–1910. [PubMed: 17431518]

Kabsch W (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr 66, 133–144. [PubMed: 20124693]

Kamisetty H, Ovchinnikov S, and Baker D (2013). Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. Proceedings of the National Academy of Sciences 110, 15674.

Kammerer R, and Zimmermann W (2010). Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. BMC Biology 8, 12. [PubMed: 20132533]

Klaile E, Vorontsova O, Sigmundsson K, Muller MM, Singer BB, Ofverstedt LG, Svensson S, Skoglund U, and Obrink B (2009). The CEACAM1 N-terminal Ig domain mediates cis- and trans-binding and is essential for allosteric rearrangements of CEACAM1 microclusters. J Cell Biol 187, 553–567. [PubMed: 19948502]

Klock HE, and Lesley SA (2009). The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis. In High Throughput Protein Expression and Purification: Methods and Protocols, Doyle SA, ed. (Totowa, NJ: Humana Press), pp. 91–103.

Koradi R, Billeter M, and Wüthrich K (1996). MOLMOL: A program for display and analysis of macromolecular structures. Journal of Molecular Graphics 14, 51–55. [PubMed: 8744573]

Korotkova N, Cota E, Lebedin Y, Monpouet S, Guignot J, Servin AL, Matthews S, and Moseley SL (2006). A Subfamily of Dr Adhesins of Escherichia coli Bind Independently to Decay-accelerating Factor and the N-domain of Carcinoembryonic Antigen. Journal of Biological Chemistry 281, 29120–29130.

Korotkova N, Yang Y, Le Trong I, Cota E, Demeler B, Marchant J, Thomas WE, Stenkamp RE, Moseley SL, and Matthews S (2008). Binding of Dr adhesins of Escherichia coli to carcinoembryonic antigen triggers receptor dissociation. Mol Microbiol 67, 420–434. [PubMed: 18086185]

Kuespert K, Pils S, and Hauck CR (2006). CEACAMs: their role in physiology and pathophysiology. Current Opinion in Cell Biology 18, 565–571. [PubMed: 16919437]

Lawson EL, Mills DR, Brilliant KE, and Hixson DC (2012). The Transmembrane Domain of CEACAM1-4S Is a Determinant of Anchorage Independent Growth and Tumorigenicity. PLOS ONE 7, e29606. [PubMed: 22235309]

Lopez CJ, Fleissner MR, Guo Z, Kusnetzow AK, and Hubbell WL (2009). Osmolyte perturbation reveals conformational equilibria in spin-labeled proteins. Protein Sci 18, 1637–1652. [PubMed: 19585559]

Luo J, Liu Z, Guo Y, and Li M (2015). A structural dissection of large protein-protein crystal packing contacts. Sci Rep 5, 14214. [PubMed: 26370141]

Markel G, Gruda R, Achdout H, Katz G, Nechama M, Blumberg RS, Kammerer R, Zimmermann W, and Mandelboim O (2004). The Critical Role of Residues ⁴³R and ⁴⁴Q of Carcinoembryonic Antigen Cell Adhesion Molecules-1 in the Protection from Killing by Human NK Cells. The Journal of Immunology 173, 3732. [PubMed: 15356119]

Marth JD, and Grewal PK (2008). Mammalian glycosylation in immunity. Nat Rev Immunol 8, 874–887. [PubMed: 18846099]

Martin JN, Ball LM, Solomon TL, Dewald AH, Criss AK, and Columbus L (2016). Neisserial Opa Protein–CEACAM Interactions: Competition for Receptors as a Means of Bacterial Invasion and Pathogenesis. Biochemistry 55, 4286–4294. [PubMed: 27442026]

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ (2007). Phaser crystallographic software. J Appl Crystallogr 40, 658–674. [PubMed: 19461840]

McHaourab HS, Lietzow MA, Hideg K, and Hubbell WL (1996). Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. Biochemistry 35, 7692–7704. [PubMed: 8672470]

Meng L, Forouhar F, Thieker D, Gao Z, Ramiah A, Moniz H, Xiang Y, Seetharaman J, Milaninia S, Su M, et al. (2013). Enzymatic basis for N-glycan sialylation: structure of rat alpha2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. J Biol Chem 288, 34680–34698. [PubMed: 24155237]

Moonens K, Hamway Y, Neddermann M, Reschke M, Tegtmeyer N, Kruse T, Kammerer R, Mejías-Luque R, Singer BB, Backert S, et al. (2018). Helicobacter pylori adhesin HopQ disrupts trans dimerization in human CEACAMs. The EMBO Journal 37, e98665. [PubMed: 29858229]

Moremen KW, Ramiah A, Stuart M, Steel J, Meng L, Forouhar F, Moniz HA, Gahlay G, Gao Z, Chapla D, et al. (2018). Expression system for structural and functional studies of human glycosylation enzymes. Nat Chem Biol 14, 156–162. [PubMed: 29251719]

Moure MJ, Eletsky A, Gao Q, Morris LC, Yang JY, Chapla D, Zhao Y, Zong C, Amster IJ, Moremen KW, et al. (2018). Paramagnetic Tag for Glycosylation Sites in Glycoproteins: Structural Constraints on Heparan Sulfate Binding to Robo1. ACS Chem Biol 13, 2560–2567. [PubMed: 30063822]

Nesmelov YE, and Thomas DD (2010). Protein structural dynamics revealed by site-directed spin labeling and multifrequency EPR. Biophysical reviews 2, 91–99. [PubMed: 21687819]

Öbrink B (1997). CEA adhesion molecules: multifunctional proteins with signal-regulatory properties. Current Opinion in Cell Biology 9, 616–626. [PubMed: 9330864]

Pan Y, Karagiannis K, Zhang H, Dingerdissen H, Shamsaddini A, Wan Q, Simonyan V, and Mazumder R (2014). Human germline and pan-cancer variomes and their distinct functional profiles. Nucleic Acids Res 42, 11570–11588. [PubMed: 25232094]

Pannier M, Veit S, Godt A, Jeschke G, and Spiess HW (2000). Dead-time free measurement of dipole-dipole interactions between electron spins. J Magn Reson 142, 331–340. [PubMed: 10648151]

Patel PC, Lee HSW, Ming AYK, Rath A, Deber CM, Yip CM, Rocheleau JV, and Gray-Owen SD (2013). Inside-out signaling promotes dynamic changes in the CEACAM1 oligomeric state to control its cell adhesion properties. Journal of Biological Chemistry.

Peng G, Yang Y, Pasquarella JR, Xu L, Qian Z, Holmes KV, and Li F (2017). Structural and Molecular Evidence Suggesting Coronavirus-driven Evolution of Mouse Receptor. J Biol Chem 292, 2174–2181. [PubMed: 28035001]

Polyhach Y, Bordignon E, and Jeschke G (2011). Rotamer libraries of spin labelled cysteines for protein studies. Physical Chemistry Chemical Physics 13, 2356–2366. [PubMed: 21116569]

Reeves PJ, Callewaert N, Contreras R, and Khorana HG (2002). Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. Proc Natl Acad Sci U S A 99, 13419–13424. [PubMed: 12370423]

Robert X, and Gouet P (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42, W320–324. [PubMed: 24753421]

Sarver JL, Zhang M, Liu L, Nyenhuis D, and Cafiso DS (2018). A Dynamic Protein–Protein Coupling between the TonB-Dependent Transporter FhuA and TonB. Biochemistry 57, 1045–1053. [PubMed: 29338257]

Schwieters CD, Kuszewski JJ, Tjandra N, and Marius Clore G (2003). The Xplor-NIH NMR molecular structure determination package. Journal of Magnetic Resonance 160, 65–73. [PubMed: 12565051]

Taheri M, Saragovi U, Fuks A, Makkerh J, Mort J, and Stanners CP (2000). Self Recognition in the Ig Superfamily: IDENTIFICATION OF PRECISE SUBDOMAINS IN CARCINOEMBRYONIC ANTIGEN REQUIRED FOR INTERCELLULAR ADHESION. Journal of Biological Chemistry 275, 26935–26943.

Tan K, Zelus BD, Meijers R, Liu JH, Bergelson JM, Duke N, Zhang R, Joachimiak A, Holmes KV, and Wang JH (2002). Crystal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family. EMBO J 21, 2076–2086. [PubMed: 11980704]

Tchoupa AK, Schuhmacher T, and Hauck CR (2014). Signaling by epithelial members of the CEACAM family - mucosal docking sites for pathogenic bacteria. Cell Commun Signal 12, 27–27. [PubMed: 24735478]

Virji M, Evans D, Hadfield A, Grunert F, Teixeira AM, and Watt SM (1999). Critical determinants of host receptor targeting by Neisseria meningitidis and Neisseria gonorrhoeae : identification of Opa adhesiotopes on the N-domain of CD66 molecules. Mol Microbiol 34, 538–551. [PubMed: 10564495]

Virji M, Watt SM, Barker S, Makepeace K, and Doyonnas R (1996). The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of Neisseria meningitidis and Neisseria gonorrhoeae. Mol Microbiol 22, 929–939. [PubMed: 8971714]

Watt SM, Teixeira AM, Zhou GQ, Doyonnas R, Zhang Y, Grunert F, Blumberg RS, Kuroki M, Skubitz KM, and Bates PA (2001). Homophilic adhesion of human CEACAM1 involves N-terminal domain interactions: structural analysis of the binding site. Blood 98, 1469–1479. [PubMed: 11520797]

Wu TJ, Shamsaddini A, Pan Y, Smith K, Crichton DJ, Simonyan V, and Mazumder R (2014). A framework for organizing cancer-related variations from existing databases, publications and NGS data using a High-performance Integrated Virtual Environment (HIVE). Database (Oxford) 2014, bau022. [PubMed: 24667251]

Zhuo Y, Yang JY, Moremen KW, and Prestegard JH (2016). Glycosylation Alters Dimerization Properties of a Cell-surface Signaling Protein, Carcinoembryonic Antigen-related Cell Adhesion Molecule 1 (CEACAM1). J Biol Chem 291, 20085–20095. [PubMed: 27471271]

Zhuo Y, Yang JY, Moremen KW, and Prestegard JH (2020). Correction: Glycosylation alters dimerization properties of a cell-surface signaling protein, carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1). J Biol Chem 295, 3748. [PubMed: 32169858]

Highlights

- Crystal structure of minimally glycosylated CCM1 dimerizes at the GFCC′C″ face.
- CCM1 forms a dimer in solution regardless of glycosylation state.
- The solution dimer of CCM1 is the same dimer observed in the crystal structure.

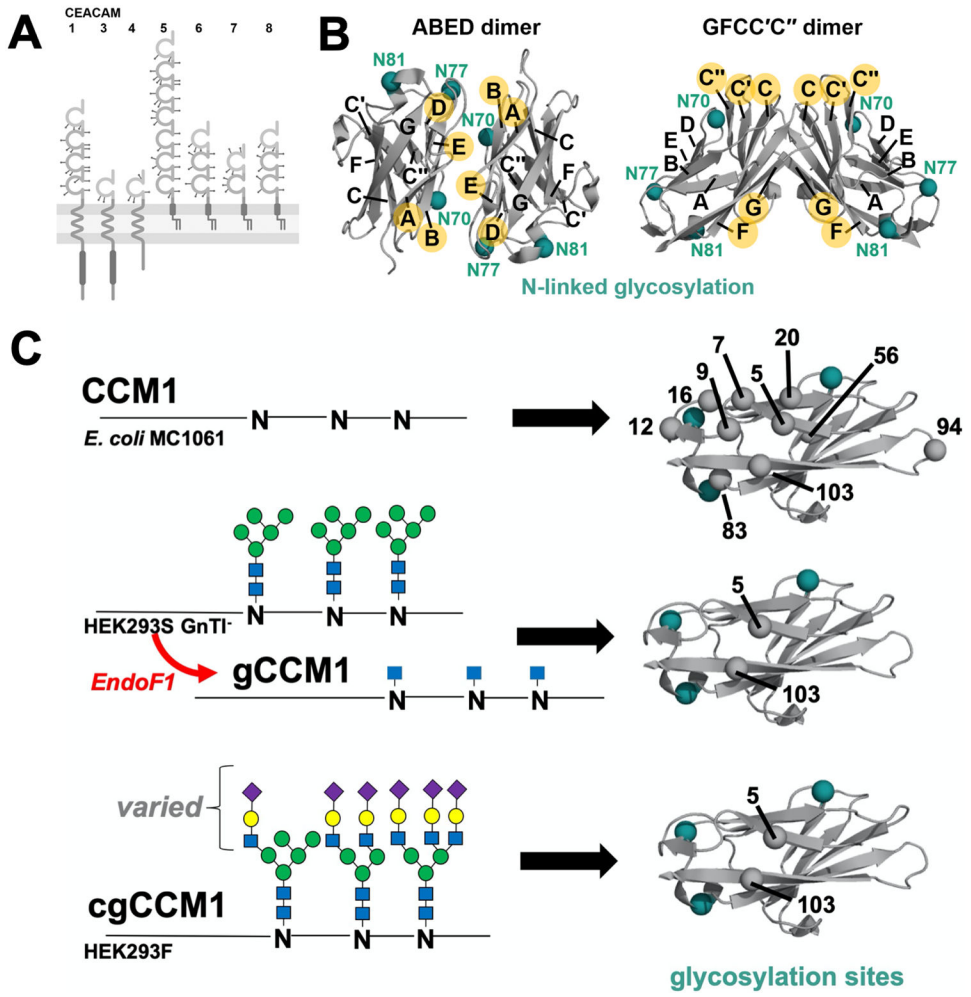


Figure 1. CEACAM protein family, dimerization and glycosylation.

(A) Representative human CEACAM family members (CEACAM1, 3-8) that possess one variable-like Ig domain (light gray; N-domain/CCM) and a different number of constant C2-like Ig domains (gray). CEACAM extracellular domains are generally glycosylated (stick and ball). CEACAM1 and CEACAM3-4 are anchored to the cellular membrane via transmembrane domains and may contain additional cytoplasmic domains (CEACAM1, 3). CEACAM5-8 are associated with the membrane through a glycosylphosphatidylinositol anchor. Image created with BioRender. (B) Two unique CCM1 (N-domain) dimer structures are reported in the literature. β-strands are labeled according to the scheme described by (Fedarovich et al., 2006) Published CCM1 dimer interfaces (yellow letters) are formed by strands ABED (PDB ID: 2GK2; (Fedarovich et al., 2006) and GFCC′C″ (PDB ID: 4QXW(Huang et al., 2014). N-linked glycosylation sites N70, N77, and N81 (Zhuo et al., 2016) are rendered as teal spheres. (C) Schematic of CCM1 constructs with different glycosylation states (left): non-glycosylated CCM1 (CCM1), CCM1 with single N-linked GlcNAc modifications (gCCM1), CCM1 with heterogeneous, complex N-linked glycans (cgCCM1). Glycosylation is represented in standard glycan symbol nomenclature: GlcNAc (blue squares), Man (green circles), Gal (yellow circle), Neu5Ac (purple diamond). Spin

labeled sites of the corresponding constructs (right) are indicated as gray spheres with residue number labels and glycosylated asparagine residues in teal spheres.

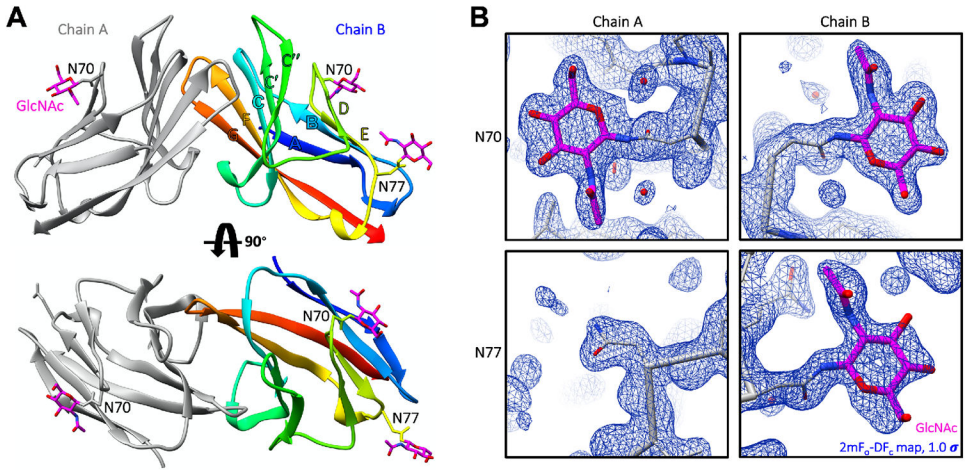


Figure 2. Crystal structure of gCCM1.

A) Crystal structure of gCCM1 reveals the GFCC′C″ dimer and is shown with one monomer in grey (chain A) and one monomer in rainbow (chain B) from the N-terminus (blue) to the C-terminus (red) with the strands labeled in their corresponding color. GlcNAc moieties are shown in stick representation (magenta). B) Electron density ($2mF_o$ - DF_c map; blue mesh) of the GlcNAc moieties (magenta) is shown for the corresponding glycosylation sites. The protein chain is shown in grey. GlcNAc was not modeled at N77 of chain A for lack of clear density.

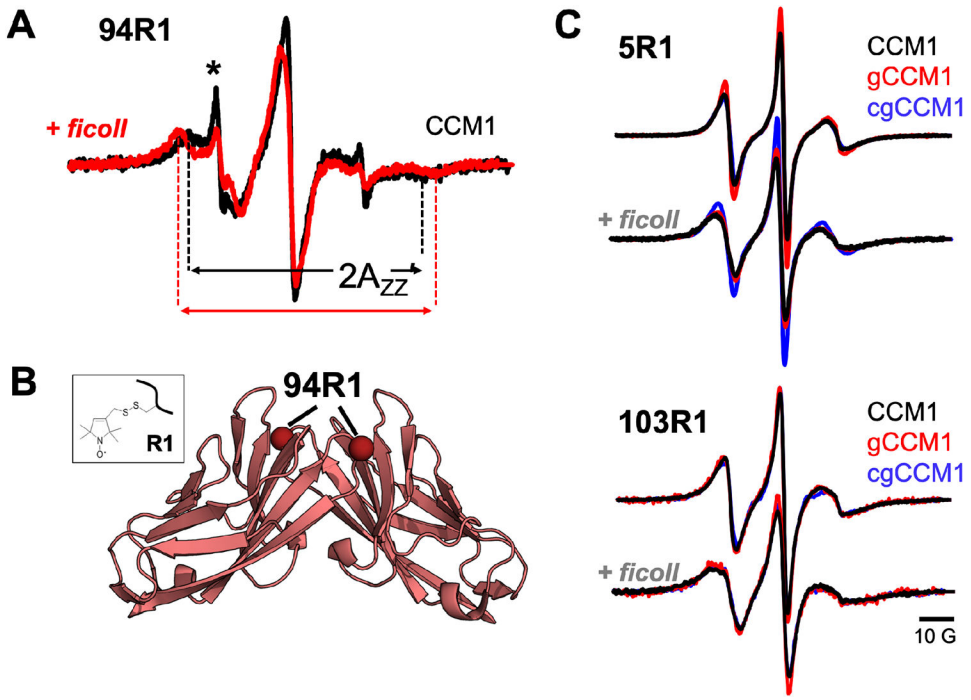


Figure 3. GFCC′C″ dimer conformation is not perturbed by glycosylation state.

CW EPR lineshape (A) of CCM1 94R1 (B) indicates that nitroxide label (R1; insert) is highly restricted and at a buried site. Spectra recorded with (red) and without (black) ficoll. The effective hyperfine splitting, $2A_{zz}'$, of restricted nitroxides used to estimate the nitroxide correlation time. Asterisk denotes component reflecting the presence of free spin label in the sample. Buried CCM1 94R1 line shape is consistent with GFCC′C″-interface dimers (B). (C) EPR spectra of 5R1 and 103R1 CCM1 (black), gCCM1 (red), and cgCCM1 (blue) with and without ficoll show similar levels of line-broadening, suggesting all three glycoforms have similar populations of dimer.

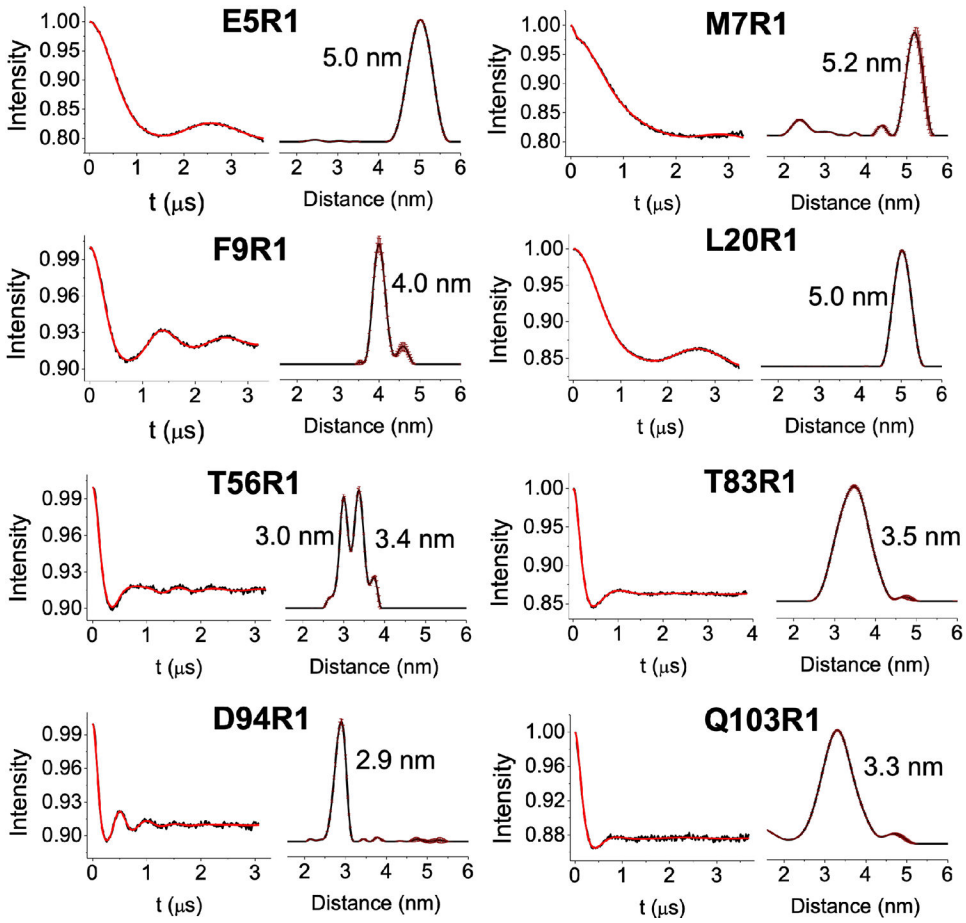


Figure 4. CCM1 DEER distance distributions.

Background-corrected dipolar evolution functions (DEF; left, black trace) fit with Tikhonov regularization (left, red trace) for the indicated spin-labeled CCM1 mutants. Normalized DEER distance distributions (right, black trace) were generated from fits to the DEF for pairs of R1 residues in CCM1 dimer. Error within distance distribution is plotted as maroon bars (right). Average distances in nm.

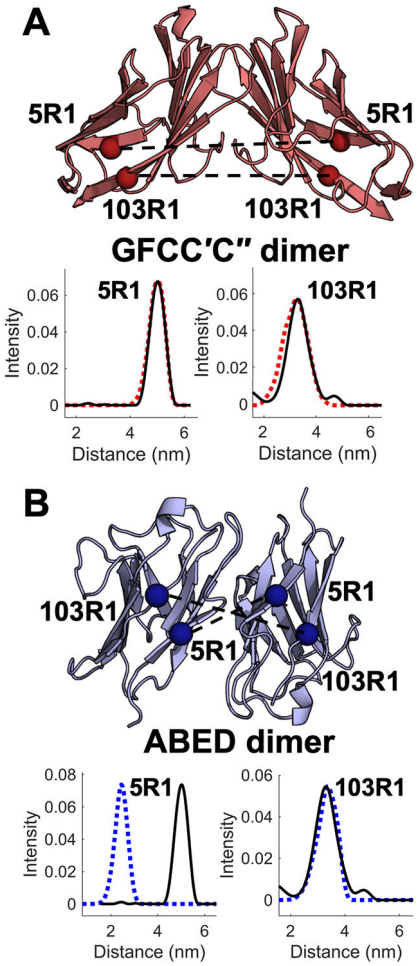


Figure 5. Simulated and experimental DEER distances for 5R1 and 103R1.

Simulated (colored dashed trace) and experimental (solid trace) DEER distance distributions based on GFCC′C″ (A, red) and ABED (B, blue) dimer structures. Black dashed lines indicate separations between labeled sites (spheres). The chosen R1 subset represents a distance which is expected to be different (5R1) and the same (103R1) between conformers.

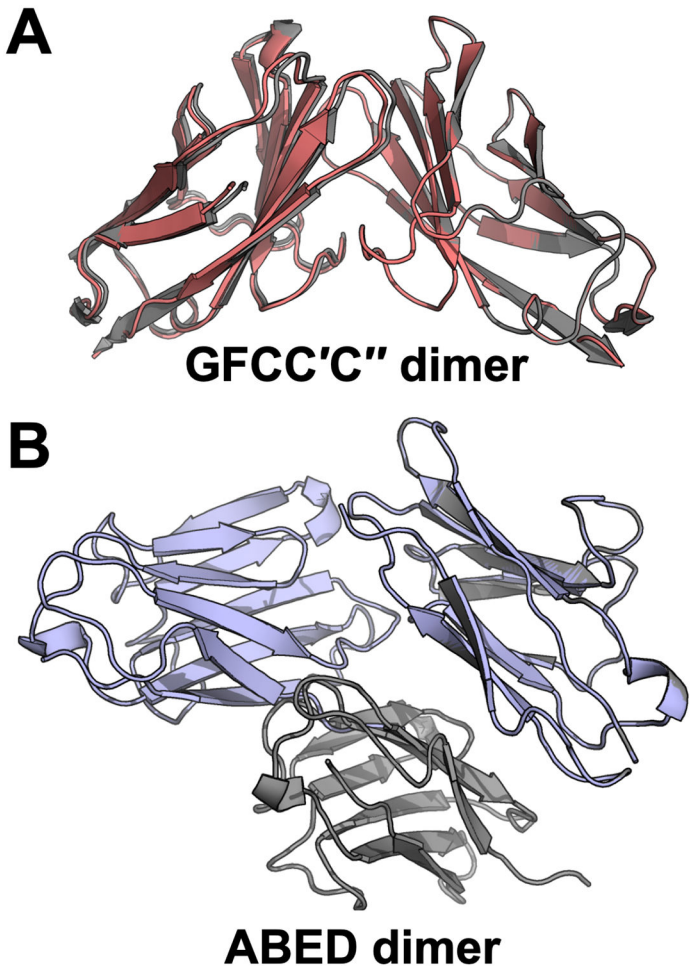


Figure 6. CCM1 dimer refinement using DEER distances.

A simulated annealing refinement of each dimer model (GFCC′C″ in panel A (PDB ID: 4QXW) and ABED in panel B (PDB ID: 2GK2)) was performed using the experimentally determined DEER distances The resulting DEER-refined dimer structures are colored red (A) and blue (B).

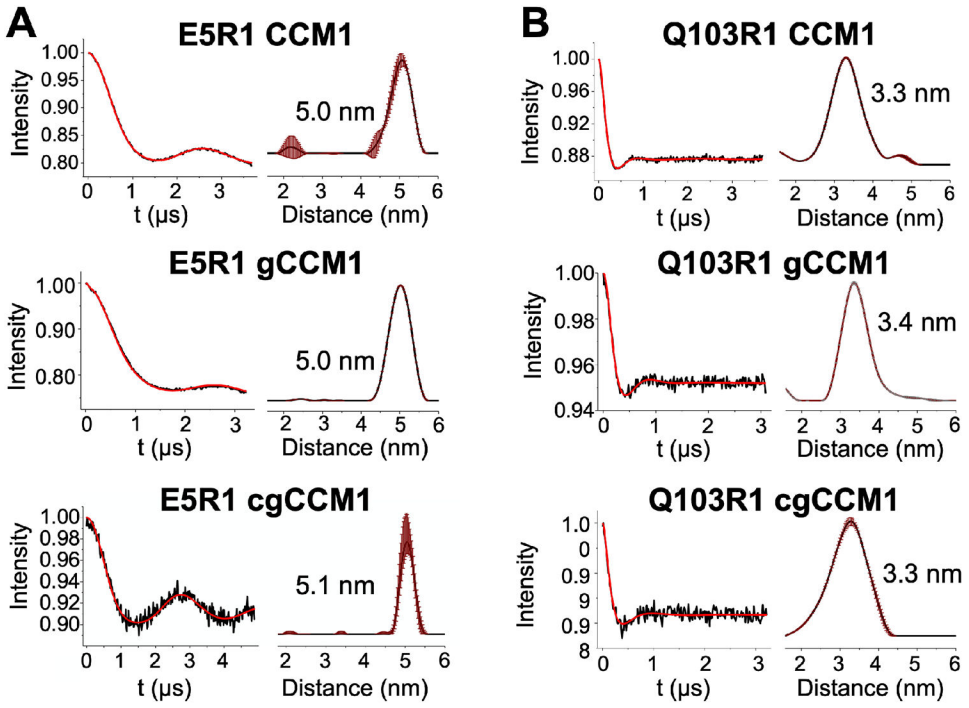


Figure 7. DEER distances of glycosylated CCM1 are similar to non-glycosylated CCM1.

Background-corrected DEFs (left columns, black trace) and resulting normalized distance distributions (right columns, black trace) from the fits to the DEF (left columns, red trace) for 5R1 (A) and 103R1 (B) labeled CCM1, gCCM1, and cgCCM1. Error within the distance distribution is plotted as maroon bars (right columns). Average distances are noted in nm.

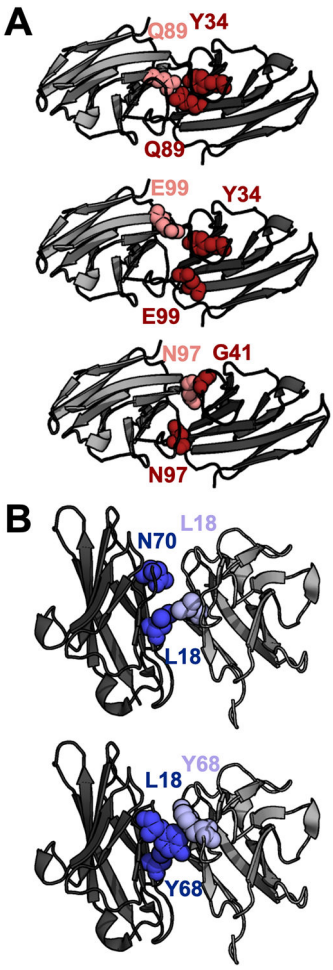


Figure 8. Evolutionarily-coupled (EC) residue pairs in dimer interfaces.

EC residue pairs (spheres) that are within 5 Å across the GFCC′C″ (A, red) and ABED (B, blue) CCM1 dimer interfaces. Three EC pairs were identified across the GFCC′C″ interface (Q89/Y34, E99/Y34, and N97/G41), while two EC pairs were identified across the ABED interface (L18/N70 and L18/Y68). Protomers are represented by dark and light shades, respectively.

Table 1.

Data Collection and Refinement Statistics

Data Collection	**gCCM1**
Wavelength (Å)	1.0
Resolution range	46.73 – 1.5 (1.554-1.5)
Space group	H32
Unit cell dimensions	
a, b, c (Å)	93.455, 93.455, 136.665
α, β, γ (°)	90, 90, 120
Total no. of reflections	408082 (41174)
No. of Unique Reflections	36364 (3588)
Redundancy	11.2 (11.5)
% Completeness	99.80 (99.89)
Mean I/σ(I)	12.88 (1.11)
R_{merge}	0.07249 (2.009)
R_{meas}	0.07597 (2.103)
R_{pim}	0.02249 (0.6181)
Wilson B factor	27.35
$CC_{1/2}$	0.999 (0.494)
Refinement	
Resolution	46.73 - 1.7 (1.761 - 1.7)
No. reflections used	25070 (2496)
Reflections used for R_{free}	1290 (132)
No. of non-hydrogen atoms	1883
protein	1678
ligands	107
solvent	98
Protein residues	214
R_{work}	0.2022 (0.2538)
R_{free}	0.2172 (0.2819)
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (Å)	1.48
Ramachandran plot	
Favored regions	98.10
Allowed regions	1.90
Outliers	0.00
Clashscore	3.41
Average B-factor	33.79

Data Collection	**gCCM1**
protein	32.38
ligands	53.16
solvent	36.82

Statistics for the highest-resolution shell are shown in parentheses.

Table 2:

Xplor-NIH restraints and structural statistics for CCM1 ensembles[a]

Parent Structure	**GFCC′C″**	**GFCC′C″**	**ABED**	**ABED**
Distance Restraint	C_β–C_β	NS1–NS1[b]	C_β–C_β	NS1–NS1[b]
# Distance Restraints	10	9[c]	10	8[c]
# Restraint Violations	2	0	9	6
Maximum Restraint Violation (Å)	0.7	-	19.1	17.6
Mean Ensemble Backbone RMSD (Å)	0.2 ± 0.1	0.2 ± 0.1	0.8 ± 0.7	0.4 ± 0.2
Mean Ensemble Heavy Atom RMSD (Å)	0.9 ± 0.2	0.6 ± 0.2	1.3 ± 0.6	0.6 ± 0.2
Average Backbone RMSD (Å) of Parent to CCM1 ensemble	0.7 ± 0.1	0.4 ± 0.1	19.4 ± 0.1	13.0 ± 0.3

[a] Statistics calculated for the 20 lowest-energy structures

[b] NS1 represents the nitrogen atom of the R1 nitroxide moiety

[c] Restraint(s) removed for sites that were not *in silico* labeled with R1 due to clashing

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli strain MC1061	Coli Genetic Stock Center (CGSC)	CGSC#: 6649
Escherichia coli strain BL21 (DE3)	ThermoFisher Scientific	Cat #: EC0114
Chemicals, peptides, and recombinant proteins		
Ampicillin sodium salt	Fisher	Cat #: BP1760
Streptomycin sulfate salt	Sigma	Cat #: S9137
Difco™ LB Broth	BD	Cat #: 244610
Polyethyleneimine (PEI)	Polysciences, Inc.	Cat #: 23966
Valproic acid sodium salt	Sigma	Cat #: P4543
FreeStyle 293 Expression Medium	ThermoFisher Scientific	Cat #: 12338018
EX-CELL 293 Serum-Free Medium for HEK 293 Cells	Sigma-Aldrich	Cat #: 14571C
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Research Products International	Cat #: I56000
Tris Hydrochloride	Fisher	Cat #: BP153
Sodium chloride (NaCl)	Fisher	Cat #: BP358
Ethylenediamine Tetraacetic Acid, Disodium Salt Dihydrate (EDTA)	Fisher	Cat #: O2793
DL-Dithiothreitol (DTT)	Research Products International	Cat #: D11000
Glycerol	Macron Fine Chemicals	Cat #: 5092-02
Ammonium sulfate	Sigma-Aldrich	Cat #: A5132
Glutathione, reduced	Acros Organics	Cat #: 12000250
TEV protease	In-house	N/A
Pierce Glutathione Agarose	Thermo Scientific	Cat #: 16101
HR Sephacryl S-200 Gel Filtration Column (26/60 mm)	Cytiva Life Sciences	Cat #: 17119501
Cobalt(II) chloride hexahydrate	Sigma-Aldrich	Cat #: 202185
Chelating Sepharose Fast Flow	Cytiva Life Sciences	Cat #: 17057502
HEPES sodium salt	Sigma	Cat #: H7006
Imidazole	Acros Organics	Cat #: 301870025
Endoglycosidase F1	In-house	N/A
MES hydrate	Sigma	Cat #: M2933
PD-10 desalting columns	Cytiva Life Sciences	Cat #: 17085101
Sodium Phosphate Monobasic Monohydrate	Fisher	BP330
Sodium Phosphate Dibasic Anhydrous	Fisher	Cat #: S375
S-(2, 2, 5, 5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL)	Toronto Research Chemicals Inc.	Cat #: O875000
Ficoll PM 70	Sigma	Cat #: F2878
*Pfu*Turbo DNA Polymerase	Agilent	Cat #: 600250
Critical commercial assays		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Nextal Tubes JCSG Core I (commercial crystallization screen)	Qiagen	Cat #: 130724
Nextal Tubes JCSG Core II (commercial crystallization screen)	Qiagen	Cat #: 130725
Nextal Tubes JCSG Core III (commercial crystallization screen)	Qiagen	Cat #: 130726
Nextal Tubes JCSG Core IV (commercial crystallization screen)	Qiagen	Cat #: 130727
PEG/Ion (commercial crystal screen)	Hampton Research	Cat #: HR2-126
Crystal Screen (commercial crystal screen)	Hampton Research	Cat #: HR2-110
Deposited data		
Crystal structure of minimally glycosylated N-terminal CEACAM1 domain	This paper	PDB ID: 7MU8
Crystal structure of unglycosylated CEACAM1 N-terminal domain (GFCC′C′′ dimer)	Huang et al., 2014	PDB ID: 4QXW
Crystal structure of unglycosylated CEACAM1 N-terminal domain (ABED dimer)	Fedarovich et al., 2006	PDB ID: 2GK2
Experimental models: Cell lines		
HEK293S GnTI- cells	ATCC	CRL-3022
FreeStyle™ 293F cells	ThermoFisher Scientific	Cat #: R79007
Oligonucleotides		
PIPE mutagenesis primers	This study (Supplemental Table S6)	N/A
Recombinant DNA		
Plasmid: CEACAM1 N-terminal domain in pGen2	Moremen et al., 2018	N/A
Plasmid: CEACAM1 N-terminal domain in pGEX-2T	Rob Nicholas (University of North Carolina at Chapel Hill)	N/A
Software and algorithms		
XDS	Kabsch, 2010	https://xds.mr.mpg.de
AIMLESS	Evans and Murshudov, 2013	https://www.ccp4.ac.uk
PHASER	McCoy et al., 2007	https://www.ccp4.ac.uk
Coot	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PHENIX	Adams, et al., 2010	https://phenix-online.org
Lab-VIEW EPR Analysis Programs	C. Altenbach, University of California at Los Angeles	http://www.biochemistry.ucla.edu/Faculty/Hubbell/
DEERAnalysis2016	Jeschke et al., 2006	https://epr.ethz.ch/software/older-versions/old_deeranalysis.html
Multiscale Modeling of Macromolecules (MMM)	Kamisetty et al., 2013	https://epr.ethz.ch/software/older-versions/old_mmm.html
XPLOR-NIH	Schwieters et al., 2003	https://nmr.cit.nih.gov/xplor-nih/
MOLMOL	Koradi et al., 1996	https://sourceforge.net/projects/molmol/
GREMLIN	Balakrishnan et al., 2011	http://gremlin.bakerlab.org