SI Appendix for "The core trisaccharide of an N-linked glycoprotein accelerates folding and enhances stability"

Sarah R. Hanson, Elizabeth K. Culyba, Tsui-Ling Hsu, Chi-Huey Wong, Jeffery W. Kelly, and Evan T. Powers

Supporting Methods

1. General

Unless otherwise noted, chemicals and products were purchased from Fisher Scientific or Sigma-Aldrich. Phosphate buffered saline (PBS) was prepared from PBS tablets (Sigma P-4417) and maintained at pH 7.2. All buffer solutions were filtered (Millipore 0.2 μm) and autoclaved before use. Protein was concentrated using Amicon centrifugation devices, MWCO 10 kDa (Millipore). Final CD2 concentrations were determined by either BCA assay or evaluation of absorbance at 280 nm using calculated extinction coefficients (Vector NTI, Informax).

1.1 hCD2 structure

Fig. 1*B* shows the hCD2ad NMR structure. Structural coordinates were obtained from the PDB (accession code 1GYA.pdb). Coordinates were manipulated and rendered using PyMOL. hCD2ad is a member of the immunoglobulin superfamily (IgSF, a β-sandwich fold) that mediates immunological synapse formation and immune cell activation (1). The protein has no prolines or cysteines to complicate its folding kinetics and it has a single N-glycosylation site at Asn-65, which is in a type I β-turn flanked by strands D and E. The innermost $GlcNAc_2$ disaccharide of the oligomannose N-glycan at Asn-65 makes contacts with strands B and E of hCD2ad (2) which are likely important, as strands B, E, F and G form a structural core common to IgSF folds that is thought to nucleate folding (3), although the effect of N-glycans on the fold have yet to be considered. The N-glycan in human CD2ad is necessary for both acquiring and maintaining proper structure (2, 4, 5); in contrast, rat CD2ad does not require an N-glycan (6). Introducing elements of the rat sequence (Lys61Glu/Phe63Leu, Fig. 1*B*) into human CD2ad near its glycosylation site enables hCD2ad to fold and maintain structure without its N-glycan (2), allowing for a comparison of the influence of mutation(s) vs. glycosylation.

1.2 hCD2ad amino acid sequence

The sequence of the glycosylated hCD2ad variants is (variants **3**-**7**):

DYKDDDDKLIEGR KEITNALETW GALGQDINLD IPSFQMSDDI DDIKWEKTSD KKKIAQFRKE KETFKEKDTY KLFK**N**GTLKI KHLKTDDQDI YKVSIYDTKG KNVLEKIFDL KIQER

The first 9 residues (in blue) are a FLAG-tag (with an extra Leu residue), which was included for ease of purification. The next four residues (in red) are a Factor Xa cleavage site (FXas), which was included so that the FLAG-tag could be removed. It was found, however, that the FLAG-tag did not influence hCD2ad folding (*vide infra*), so it was not removed from the variants before they were characterized. The glycosylation site (Asn-65) is shown in bold and underlined. Note that the residues are numbered starting from the first residue (Lys) after the FXas.

The sequence of the nonglycosylated, wild-type hCD2ad is (variant **1**):

DYKDDDDKIEGR KEITNALETW GALGQDINLD IPSFQMSDDI DDIKWEKTSD KKKIAQFRKE KETFKEKDTY KLFK**N**GTLKI KHLKTDDQDI YKVSIYDTKG KNVLEKIFDL KIQER

This sequence is identical to that above except that it lacks the Leu residue at the C-terminal end of the FLAG-tag. Note that, since the FLAG-tag itself does not affect folding, this minor difference between the two FLAG-tags should not either.

The sequence of the nonglycosylated hCD2ad mutants (variant **2**):

KEITNALETW GALGQDINLD IPSFQMSDDI DDIKWEKTSD KKKIAQFRKE KETFKEKDTY **E**LL**K**<u>**N**</u>GTLKI KHLKTDDQDI YKVSIYDTKG KNVLEKIFDL KIQER

This protein lacks the FLAG-tag, which as noted above does not affect protein folding kinetics or thermodynamics, **2a** has the double mutation K61E and F63L (highlighted in bold), while **2b** has only the K61E mutation.

1.3 Molecular biology

All PCR was performed using Hot Start KOD polymerase (EMD Biosciences) using recommended conditions. Restriction enzymes were obtained New England Biolabs and applied as indicated. DNA fragments were ligated with standard conditions supplied for T4 ligase (Roche). Amplified and digested DNA was purified using 1% agarose (molecular biology grade) gel prepared in TAE buffer. DNA isolation/purification steps, including genomic isolation, plasmid isolation, restriction digestion clean-up, and PCR purification were performed with Qiagen kits. Clones were transformed, amplified, and maintained in DH10B *E. coli*. All clones were verified for accuracy by sequencing.

1.4 Protein purification steps on FPLC and HPLC

All FPLC procedures were carried out on an AKTA FPLC from GE Healthcare. A HiLoad Superdex 75 gel filtration column (320 mL) was run in PBS with a flow rate of 2.5 mL/min at room temperature (hCD2ad retention time of 70 min). Gel Filtration steps by HPLC (Hitachi 7000) were carried out on a BioSep Sec-S 3000 (300 × 21.2 mm; Phenomenex) column with PBS mobile phase, run at a flow rate of 1.2 mL/min (hCD2ad retention time of 56 min).

1.5 Liquid chromatography mass spectrometry (LCMS)

MS was performed using an Agilent 1100 LC coupled to an Agilent 1100 single quad ESI mass spectrometer. LC was performed with a 4.6 mm × 50 mm Zorbax C8 column (Agilent) with mobile phases of H_2O/0.1% formic acid (A) and ACN/0.1% formic acid (B). The hCD2ad gradient ramped from 5% A to 85% B over 15 minutes (retention time = 8.5 min). ESI data were collected by positive scan in the 1,000- to 2,000-kDa mass range. Deconvolution Reporting Software (DRS, Agilent) was used to render protein masses from multiply charged states.

1.6 Circular dichroism (CD) spectrometry

Circular dichroism spectra were recorded on an Aviv Model 202SF CD spectrometer, as described in *Methods*.

1.7 Fluorescence spectrometry

hCD2ad has a two tryptophan residues, one of which is buried in the hydrophobic core, allowing for an intrinsic fluorescence that depends on the folding status. Fluorescence

measurements for hCD2ad variants were obtained using either a CARY Eclipse (Varian) or an ATF-105 (Aviv) fluorescence spectrometer, as described in Methods.

2. Preparation of hCD2ad variants

2.1 Nonglycosylated variants

2.1.1 Construction of nonglycosylated variant genes

Genes for nonglycosylated versions of hCD2ad (variants **1** and **2**) were assembled using the thermodynamically balanced inside out (TBIO) method from DNAWorks (7). These bacterial hCD2ad genes were designed with optimized codons for heterologous expression in *E. coli*. Listed below are the 12 primers (ng_1-ng_12) TBIO generated to create the variant **1** gene flanked at the 5' end with a BamHI site and an encoding region for a Factor Xa protease cleavage site (FXas, primer ng_1) and the 3' end with a stop codon followed by an EcoRI site (primer ng_12). The stabilized mutant hCD2ad (variant **2**) gene was constructed similarly, replacing primers ng_7 and ng_8 with mutng_7 and mutng_8, respectively. These primers introduce mutations K61E and F63L into the expressed protein. TBIO was carried out, as prescribed (7), in two successive rounds to generate the desired 348kb PCR product.

2.1.2 TBIO DNA Works Primers for bacterial expression of non-glycosylated hCD2ad

```
ng_1  CGCGGATCCATCGAAGGTCGAAAAGAAATT
ng_2  ATCCATCGAAGGTCGAAAAGAAATTACCAATGCGCTTGAGACCTGGGGTGCCT
ng_3  CTTGAGACCTGGGGTGCCTTAGGCCAAGACATTAACCTGGACATCCCGTCGTT
ng_4  TAACCTGGACATCCCGTCGTTCCAAATGAGCGATGACATTGACGACATCAAGT
ng_5  CGATGACATTGACGACATCAAGTGGGAGAAGACCAGCGACAAGAAGAAGATCG
ng_6  ACCAGCGACAAGAAGAAGATCGCGCAGTTCCGCAAGGAGAAGGAGACGTTCAA
ng_7  GTTCTTGAAAAGCTTGTAGGTGTCCTTCTCCTTGAACGTCTCCTTCTCCTTGC
ng_8  TCTTCAGGTGCTTGATTTTCAGCGTGCCGTTCTTGAAAAGCTTGTAGGTGTCC
ng_9  TGCTCACCTTGTAGATGTCCTGGTCGTCGGTCTTCAGGTGCTTGATTTTCAGC
ng_10 TCCAGCACGTTTTTACCCTTGGTGTCGTAGATGCTCACCTTGTAGATGTCCTG
ng_11 ACGCTCTTGAATTTTCAAATCAAATATCTTTTCCAGCACGTTTTTACCCTTGG
ng_12 CCGGAATTCCTAACGCTCTTGAATTTTCAAATCAAATATCT
```

```
mutng_7  GTTCTT**CAG**AAG**TTC**GTAGGTGTCCTTCTCCTTGAACGTCTCCTTCTCCTTGC
mutng_8  TCTTCAGGTGCTTGATTTTCAGCGTGCCGTTCTT**CAG**AAG**TTC**GTAGGTGTCC
```

2.1.3 Cloning of nonglycosylated variants gene as GST-fusions

TBIO PCR products were subcloned into the expression vector pGEX-2T by double digestion (BamHI and EcoRI) and ligation of vector and PCR products. The resulting clones,

pGST-hCD2ad**1** and pGST-hCD2ad**2** contained a coding region from N- to C-terminus of: GST-FXas-hCD2ad**1** or **2**. The GST fusion was engineered for robust expression and purification of soluble N-terminal GST-CD2 fusions.

2.1.4 Cloning of nonglycosylated variants as 6×His-FLAGtag-fusions

Inserts from pGST-hCD2ad**1** and pGST-hCD2ad**2** were subcloned into pMH4 expression vectors using the PIPES method, to create pHisFLAG-hCD2ad**1** and pHisFLAG-hCD2ad**2**, respectively. These constructs are equipped with a dual N-terminal tag consisting of a 6thio-6His and a FLAG-tag, which are separated by a region encoding for a tobacco etch virus (TEV) protease cleavage site (TEVs). The total N- to C-protein coding region is: 6thio-6His-TEVs-FLAGtag-FXas-hCD2ad**1** or **2**.

2.1.5 Site-irected mutagenesis:

All other variants were engineered from these constructs using site directed mutagenesis.

2.1.6 Expression of nonglycosylated variants in *E. coli* (rich medium)

Bacterial hCD2ad expression vectors were transformed into *E. coli* expression strains (pGST-hCD2ad**1** or **2** constructs into DH10β cells and pHisFLAG-hCD2ad**1** or **2** into BL21(DE3) cells) and maintained in LB growth medium supplemented with appropriate antibiotic (100 μg/mL ampicillin). All cultures were grown at constant temperature (37 °C) and liquid cultures were aerated by agitation at 250 rpm. Protein expression was induced by adding 0.5 M IPTG to liquid cultures in log phase growth (OD_{600} of 0.8), which were subsequently expanded for 2 hours. Cells were pelleted by centrifugation (6,00 rpm, 30 in, 4 °C), washed with PBS buffer, repelleted, and frozen (–20 °C).

2.1.7 Lysate preparation

Cells were washed and reconstituted in an appropriate purification buffer at 1/50th of the original growth volume. After passage through a French press three times (15,00 psi; for GST fusions) or sonication (twice, 2 min; for His-tagged constructs), crude lysates were separated via centrifugation (10,000 rpm, 30 min) into soluble (supernatant) and insoluble fractions (pellet). To solubilize protein in the pelleted fraction, insoluble material was treated with 6 M guanidine hydrochloride (GdHCl) in the appropriate binding buffer. Supernatants from 6 M GdHCl treatment were subjected to denaturing purification.

2.1.8 Purification of GST-hCD2ad**2** fusion protein

Soluble GST-hCD2ad**2** lysates were prepared in PBS buffer containing 1% tritonX and passed over glutathione Sepharose 4B (GE Healthcare) by gravity. The gel was washed with 15 column volumes of PBS before applying 4 column volumes of elution buffer (50 mM Tris, 10 mM glutathione, pH 8.0). Typical GST-hCD2ad**2** yields were about 65 mg of protein per L of growth medium. The final yield of hCD2ad**2** was approximately 25 mg/L after cleavage with Factor Xa (ESI MS found: 12,396).

2.1.9 Purification of 6×His tagged proteins

Superflow Ni-NTA resin was used to affinity purify proteins bearing 6thio-6His tags, using conditions described in the Qiagen manual. Briefly, for native purification, soluble fractions were applied to a gravity Ni-NTA column in lysis buffer + 5 mM imidazole, and washed with 25 column volumes of lysis buffer and washing buffer (20 mM imidazole). Bound protein was removed with 5 column volumes of elution buffer (150 mM imidizole). Denaturing purification was performed similarly with the addition of 6 M GdHCl to all solutions. Eluted fractions were dialyzed into PBS and concentrated before TEV protease cleavage.

2.1.10 TEV cleavage of N-terminal tags from non-glycosylated proteins

Protein was dialyzed into TEV protease buffer (phosphate buffer with 1 mM DTT). TEV protease, produced in house as an N-terminal 6×His tagged fusion, was added at a ratio of 1 mg of TEV protease to 5 mg of hCD2ad substrate at room temperature for 3 h. Cleaved TEV peptides and TEV protease were removed by native Ni-NTA conditions. hCD2ad protein was then passed over an FPLC gel filtration column (Superdex 75) and concentrated before further use (hCD2ad variant **1** ESI MS found: 13880; hCD2ad variant **2a** ESI MS found: 13880; hCD2ad variant **2b** ESI MS found: 13847).

2.1.11 Factor Xa cleavage of N-terminal tags from non-glycosylated proteins

Protein was dialyzed into 50 mM Tris, pH 8.0 (10,000 MWCO membrane, Pierce), before Factor Xa (EMD) treatment (1U Factor Xa: 1.2 mg of hCD2ad substrate, 4°C, 12 h) in Tris buffer (50 mM Tris-HCl, 100 mM NaCl, 5 mM $CaCl_2$, pH 8.0). The cleavage mixture was quenched with 100μM PMSF for 10 min and separated by FPLC (Superdex 75). The hCD2ad protein fractions were concentrated and purified by an additional HPLC gel filtration step (FLAG-free hCD2ad variant **2b** ESI MS found: 12396).

2.2 Glycosylated variants

2.2.1 Cloning of hCD2ad into a mammalian expression vector

The hCD2ad was amplified from cDNA prepared from human blood mononuclear cells. The 5' primer (CGA CAA GCT TAT CGA AGG TCG **TAA AGA GAT TAC GAA TGC CTT GG**) contained an overhanging HindIII restriction site and FXas coding region contiguous with the gene annealing sequence (in bold). The 3' primer (CGC GGA TCC CTA CCT CTC TTG AAT CTT CAA ATC AAA) contained a stop codon followed by a BamHI restriction site. The 349 bp PCR product (hCD2ad**3**) was purified, doubly digested with HindIII and BamHI, and ligated into the mammalian expression vector pFLAG-CMV-3, treated under the same conditions, to create pFLAG-CD2**3**. The total engineered coding region includes from the N- to C-terminus: a preprotrypsin leader sequence (PLS, for excretion into the medium); a FLAG-tag flanked with a FXas, and then the hCD2ad sequence (PLS-FLAGtag-FXas-hCD2ad**3**)

2.2.2 Subcloning of hCD2ad into an insect shuttle vector

Primers were designed to amplify the entire PLS-FLAGtag-FXas-CD2**3** sequence flanked at the 5' end with an engineered KpnI site and the 3' end with an endogenous SacI site (5'_CD2h: GGT ACC CGG GAT CCC TAC CT CTC TTG, and 3'_CMV: AAT GTC GTA

ATA ACC CCG CCC CGT TGA CGC). Digestion (KpnI and SacI) and ligation of the 446 bp product (FLAG-CD2**4**) and the insect shuttle vector pFastBac (Invitrogen), yielded clone pFLAG-CD2**4**.

2.2.3 Expression of hCD2ad in HEK cells (hCD2ad variant **3**)

Mammalian expression was performed using human embryonic kidney (HEK), FreeStyle 293F, cells using the FreeStyle 293 Expression System Kit (Invitrogen). Briefly, cells were maintained between 3 and 5×10^5 cells/mL in Freestyle 293 Serum Free Expression Medium at 37 °C in a humidified atmosphere with 8% CO_2, under agitation (125 rpm). Transiently expressed protein was obtained by treating freshly grown cells (3×10^7/30 mL of fresh medium) with 293fectin-DNA complexes (30 µg of pFLAG-CD2**3**, 1 mL of OPTI-MEM, and 30 µL of 293fectin) and growing for 2 days. After which cells were pelleted and PBS (0.2×), protease inhibitors (1 tablet per 200 mL; Roche), and sodium azide (0.02%) were added to the growth medium extract. The mixture was filtered through a 0.2 µm membrane before proceeding to purification.

2.2.4 Expression of hCD2ad in Sf9 (insect) cells (hCD2ad variant **4**)

Expression in insect cells was carried out using Sf9 and Tn5 cell cultures maintained in sf900II-SFM (Invitrogen) and Ex-cell 405 SFM (Sigma-Aldrich), respectively, at constant temperature (27 °C) in the dark. Agitation (150 rpm) was used to aerate liquid cultures. Cultures were sustained by standard protocols, with cell numbers maintained in shaking cultures between $0.7 – 4.0\times10^6$ cells/mL of medium and in adherent cultures at $0.7–2.0\times10^6$ cells/25-cm^2 adherent surface area. Recombinant baculovirus stocks containing the FLAG-CD2**3** genes were produced using the BactoBac Baculovirus Expression System (Invitrogen), beginning from the pFastBac1 shuttle constructs pFLAG-CD2**3**. P1 viral stocks were verified to contain the hCD2ad insert by PCR (80% insert to blank found) and were frozen at –80 ˚C for long-term storage. Working baculovirus stocks were maintained by standard amplification rounds (growth to 70% cell death after multiplicity of infection, moi, ranging from 0.1-0.3), removal of cells and debris (4,000 rpm, 35 min), and storage of the growth supernatant in 10% FBS at 4 ˚C. For hCD2ad expression, virus was freshly amplified by infecting an adherent Sf9 culture (T150 flask, 1.2×10^7 cells, 45 mL of sf900II-SFM) with stock virus at a final moi of 1 for three days. The infection supernatant was subsequently added to 200 mL of log phase shaking culture cells (2.5×10^6 cells/mL). After 2 days of growth (approximately 30% cell death), the cells and cellular debris was removed. PBS (0.2×), protease inhibitors (1 tablet per 200 mL; Roche), and sodium azide (0.02%) were added to the growth medium extract and the mixture was filtered through a 0.2-µm membrane (Millipore) before proceeding to purification.

2.2.5 Purification of glycosylated variants **3** and **4** by N-terminal FLAG-tag

FLAG-tagged hCD2ad proteins from mammalian (**3**) and insect (**4**) cultures were purified using anti-FLAG M2 affinity gel (Sigma). Growth medium extracts treated with protease inhibitors (Roche) and sodium azide (0.02% final volume) were passed through the affinity gel three times by gravity. The affinity gel was then washed with ten column volumes of PBS, PBS-T (PBS + 0.05% Triton X), and PBS, before eluting the bound protein with ten column volumes of 100 mM glycine (pH 3.0). The eluted fractions were immediately neutralized by adding 1/10th volume of 1 M Tris (pH 9.0), before being combined, exchanged into an appropriate buffer, and

concentrated in Amicon centrifugation devices. If desired, the FLAG-tag, hCD2ad protein (0.5 mg/mL) could be removed by treatment with Factor Xa (1 U:2.5 μg of CD2ad) in Tris buffer (50 mM Tris-HCl, 100 mM NaCl, 5 mM $CaCl_2$, pH 8.0) for 2.5 h at 37 °C, after which the reaction was stopped with 100 μM PMSF for 10 min. Final protein was subjected to gel filtration (HPLC) and concentrated.

2.2.6 LCMS analysis of the composition of glycosylated hCD2ad variants

In addition to providing higher yields of glycoproteins, insect cells were chosen for expression of hCD2ad because they typically produce simple oligomannose- and core-type glycoforms, which are less heterogeneous than mammalian glycans and easier to remodel. Insect cells have a less sophisticated N-glycan processing pathway, where trimming of mannose residues down to the core structure, $Man_3GlcNAc_2$, known as paucimannose (pm), and adding fucose residues to the bridgehead GlcNAc in either α1,6 or α1,3 fashion, typically represents the highest form or N-glycan processing (8). In mammalian cells, processing down to core structures is always followed by further elaboration into hybrid and complex types, typically leading to a high degree of heterogeneity (9). Such mammalian and insect glycosylation trends were reflected in our analysis of glycan heterogeneity by LCMS.

hCD2ad variant **3** from HEK cells showed a heterogeneous distribution of glycans the major components of which were oligomannose, hybrid, and complex types (no detected sialylation). Table S1 lists the most abundant glycoforms in hCD2ad variant **3** at 50% relative abundance cutoff. The sample complexity is certainly much greater than that shown in the table, as witnessed by broad distribution smaller fragmentation peaks grouped around the major forms.

hCD2ad variant **4** from insect cell expression showed varying amounts of heterogeneity depending on the insect cell line used and on the individual preparation. All preparations had narrower mass distribution than those produced in HEK cells, with the most abundant glycoform being fucosylated paucimannose—$Man_3GlcNAc_2$(Fuc) (Table S2). Other glycan structures included non-fucosylated paucimannose—$Man_3GlcNAc_2$—and oligomannose—$Man_6GlcNAc_2$ and $Man_7GlcNAc_2$. Consistent with literature precedent, we found that protein from Sf9 cells had a more pronounced paucimannose majority (around 70%), than protein produced in Tn5 cells. Thus, to keep populations more homogenous, Sf9 cells were used.

2.3 Enzymatic remodeling

The N-glycan structure was remodeled so that the individual and collective contributions to protein folding energetics of the different saccharide residues could be examined. The enzymes used in our approach are outlined in Fig. S1. The procedures are described in Methods. Analysis of the steps of Endo D and F remodeling are shown in Figure S2.

To remove the N-glycan in its entirety, hCD2ad variants **3** and **4** were treated with PNGase F (New England Biolabs) or PNGase A (EMD Biosciences). hCD2ad variants **3** and **4** (0.5 mg/mL) were treated with 15,000 U/mL PNGase F in phosphate buffer (50 mM, pH 7.5) at 37 °C overnight. PNGase A (10) or F (11, 12) both were capable of completely removing the N-glycan from hCD2ad variants **3** and **4** (mass of PNGase-treated hCD2ad found: 13994; Table S3), which adds some additional insight into the N-glycan structures. For hCD2ad variant **3**, this verifies that the hybrid and complex glycoforms did not bear any terminal sialic acid residues (since PNGase A is inhibited by sialylated glycans), which might go undetected during MS

analysis from suppression or fragmentation. For hCD2ad variant **4**, this indicates that the predominant Fuc linkage is α1,6 in nature, since PNGase F is inhibited by the α1,3Fuc linkage that can also occur in insect cells. Notably, PNGase-mediated release of N-glycans occurs through hydrolytic cleavage of the amide bond between Asn and GlcNAc, which introduces an Asp peptide mutation at the protein glycosylation site. In CD2 the effect of the Asn65Asp mutation is unknown and may contribute to folding kinetics. Therefore, we produced the nonglycosylated wild type sequence with Asn65 intact by expression in *E. coli* as described above.

3.0 Additional experimental details and data for the characterization of hCD2ad folding kinetics and thermodynamics

3.1 General

PBS buffer (1×, pH 7.2) was made fresh daily from 10× stock and filtered. Urea solutions were prepared fresh daily in 1× PBS, filtered, and concentrations were confirmed by index of refraction (ior). Subsequent dilutions of urea were made in into 1× PBS and with concentrations checked by ior. Constants defined in equations include the universal gas constant (R) and temperature (T). The value of RT at 25 °C was taken to be 0.592 kcal/mol. Data were imported and fit in Mathematica 5.2 or 6.0 (Wolfram Research).

3.2 Kinetic Data

Kinetics measurements were taken on an AVIV ATF-105 stopped-flow fluorometer with two syringes (syringe 1: 1 mL, syringe 2: 2 mL) that allow up to a 16-fold dilution of the components of syringe 1 with that of syringe 2, in a minimum volume of 80 μL, to fill a 50-μL flow cell. The dead time was estimated to be 50-100 ms; generally, data after the first 200 ms were used for fitting. Excitation was set to 280 nm (bandwidth: 2 nm); emission, 330 nm (bandwidth: 8 nm); and photomultiplier voltage, 1,000 V. Data were recorded for one minute, at intervals of 10ms for the first 20 s, and then every 100 ms. For unfolding experiments, the decrease in fluorescence intensity at 330 nm (F_{330}) was monitored after native protein in PBS (syringe 1) was mixed with larger volumes of concentrated urea solutions (syringe 2). For refolding experiments, the increase in F_{330} was monitored after denatured protein in urea solution (syringe 1) was diluted with larger volumes of PBS (syringe 2). All shots were repeated at least 5 times. Continuous irradiation of hCD2ad at 280 nm led to a decrease in fluorescence intensity over time that correlated with the excitation bandwidth, indicating that photobleaching was taking place. Data were fit as described in the *Methods*.

The data for the chevron plots shown in Fig. 3 are provided as a supplementary Excel file. Note that our kinetically derived $\Delta G_{f,0}$ and m_{eq} values match well with those obtained directly from thermodynamic experiments (see below), which is a good indication of two-state behavior.

3.3 Thermodynamic stability of stable hCD2ad variants using chaotrope denaturation

All fluorescence measurements for equilibrium chaotrope denaturation experiments were taken on a Cary Eclipse fluorescence spectrophotometer. The temperature at reading was kept constant at 25 °C using a Cary single-cell Peltier accessory. The thermodynamic stability ($\Delta G_{f,0}$)

of a protein is equal to the difference in free energy between its native and unfolded states. For two-state globular proteins, equilibrium measurements in various concentrations of denaturant can be used to determine the free energy of folding under native conditions ($\Delta G_{f,0}$) by the linear extrapolation method (LEM), the principle of which states that the free energy of denaturation at a given denaturant concentration, ΔG_f, depends linearly on the denaturant concentration (13):

$$\Delta G_f = \Delta G_{f,0} + m_{eq}[\text{denaturant}] \quad \textbf{[S1]}$$

For equilibrium denaturation studies, solutions of hCD2ad variants were prepared in PBS supplemented with various concentrations of urea from 0 to 5 M, and were allowed to equilibrate for 2 minutes before fluorescence emission spectra were scanned. Notably, hCD2ad unfolding in response to increasing urea concentration causes a shift and intensity change in fluorescence spectrum. Thus, plots of fluorescence intensity at single wavelengths (F_λ), or the ratio of intensities at two wavelengths ($F_{\lambda 1}/F_{\lambda 2}$), versus urea concentration were plotted to demonstrate unfolding. $\Delta G_{f,0}$ and m_{eq} values for hCD2ad variants were estimated by fitting fluorescence intensity at 320 nm (F_{320}) vs. urea concentration data to the equation below:

$$F_{320} = \left(F_{N,320} + b_{N,320}[\text{urea}]\right) + \frac{\left(\Delta F_{D-N,320} + \Delta b_{D-N,320}[\text{urea}]\right)}{1 + e^{-\left(\Delta G_{f,0} + m_{eq}[\text{urea}]\right)/RT}} \quad \textbf{[S2]}$$

where $F_{N,320}$ is the fluorescence intensity of the native protein at 320 nm and 0 M urea, $b_{N,320}$ is the slope of the pre-transition baseline, $\Delta F_{D-N,320}$ is the difference in the fluorescence intensities of the denatured and native proteins at 320 nm and 0 M urea, $\Delta b_{D-N,320}$ is the difference between the slopes of the post- and pre-transition baselines, and [urea] is the urea concentration.

The stability curves of fully glycosylated hCD2ad variants display apparent two-state behavior, revealing a highly reversible and cooperative denaturation curve in which the protein transitions from native to denatured with increasing urea concentration, as depicted in Figure S4. As with kinetic studies, fits of Eq. [S2 to the data yield similar parameters for hCD2ad variants **4** and **7**, suggesting that heterogeneity and fucosylation do not significantly affect folding energetics. Also, variant **5**, which has a single GlcNAc, is less stable than the variants with extended glycans, and has an unfolding curve characterized by a weakly defined pre-transition region. This behavior is consistent with the full contribution of N-glycans to protein stability being due to a glycan structure larger than a single GlcNAc and smaller than the paucimannose structure that predominates in the hCD2ad variant **4** population, likely $ManGlcNAc_2$ as indicated by kinetic studies. The stabilized aglycone (variant **2b**) also shows two-state behavior, albeit with reduced stability and a limited pre-transition base line. Despite the fact that short and sloping pre-transition base lines introduced error into the data fits shown in Figure S4, the equilibrium denaturation-derived stabilities agree well with their kinetically derived counterparts. Such agreement is considered to be a good check for two-state behavior (14).

We verified that the FLAG-tag, appended to hCD2ad for facile purification, does not affect the protein denaturation profile. To do so, variant **4** was treated with Factor Xa and purified, to produce variant **4a**, which does not have a FLAG-tag but is otherwise identical to variant **4**, for side-by-side comparison. The curves shown in Figure S4 are virtually identical for variants **4** and **4a**, demonstrating that the FLAG-tag does not contribute to or interfere with stability.

3.4 Osmolyte renaturation/urea denaturation.

Nonglycosylated wild type hCD2ad (**1**) is not fully folded under "native" conditions (i.e., PBS buffer). It can, however, be induced to fold by adding osmolytes, small molecules (e.g., trimethylamine N-oxide, sucrose, sarcosine, glycine, and proline) that stabilize the native states of proteins (15). Variant **1** can be stabilized in proline as seen in Fig. 2C. Analogously to denaturant-induced destabilization, the extent of osmolyte-induced stabilization increases linearly with osmolyte concentration; however, note that m_{eq} for osmolytes is negative, whereas it is positive for denaturants. Osmolytes and denaturants have to be used in combination to determine $\Delta G_{f,0}$ for partially folded proteins, due to lack of pre-transition baselines. The effects of osmolytes and denaturants on ΔG_f are additive (16) so that:

$$\Delta G_f = \Delta G_{f,0} + m_{eq,os}[\text{osmolyte}] + m_{eq,den}[\text{denaturant}] \qquad \textbf{[S3]}$$

All phases of the folding transition were captured by measuring foldedness (F_{320}) at a range of osmolyte *and* denaturant concentrations. Fluorescence spectra of hCD2ad variant **1** were measured at 25 °C on an Aviv model ATF-105 fluorometer at 85 combinations of urea and proline concentrations, where the urea concentration varied from 0 to 1.5 M and the proline concentration varied from 0 to 5.6 M. Values of F_{320} were then plotted as a surface vs. [urea] and [proline] and the data were fit by the equation:

$$F_{320} = \left(F_{N,320} + b_{N,320,urea}[\text{urea}] + b_{N,320,pro}[\text{proline}]\right) + \frac{\left(\Delta F_{D-N,320} + \Delta b_{D-N,320,urea}[\text{urea}] + \Delta b_{D-N,320,pro}[\text{proline}]\right)}{1 + e^{-\left(\Delta G_{f,0} + m_{eq,urea}[\text{urea}] + m_{eq,pro}[\text{proline}]\right)/RT}} \qquad \textbf{[S4]}$$

Eq. [S4] is analogous to the equation for F_{320} for simple urea denaturations (Eq. **S2**) except that it contains terms that account for the dependence of F_{320} and ΔG_f on proline: $b_{N,320,pro}$ is the slope of the native state's F_{320} on with respect to [proline], $\Delta b_{D-N,320,pro}$ is the difference between the slopes of the denatured and native states' F_{320} with respect to [proline], and $m_{eq,pro}$ is the slope of ΔG_f with respect to [proline]. Fitting this equation to the F_{320} data yielded the following parameter estimates: $\Delta G_{f,0}$ = 0.4 ± 0.4 kcal/mol, $m_{eq,pro}$= –0.6 ± 0.3 kcal/mol M, $m_{eq,urea}$= 1.9 ± 1.0 kcal/mol M. The fit to the data is shown in Fig. S5.

4.0 References

1. Tangye SG, Phillips JH, Lanier LL (2000) The CD2-subset of the Ig superfamily of cell surface molecules: receptor-ligand pairs expressed by NK cells and other immune cells. *Semin Immunol* 12:149-157.
2. Wyss DF*, et al.* (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* 269:1273-1278.
3. Clarke J, Cota E, Fowler SB, Hamill SJ (1999) Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Struct Fold Des* 7:1145-1153.
4. Recny MA, *et al.* (1992) N-Glycosylation is required for human CD2 immunoadhesion functions. *J Biol Chem* 267:22428-22434.
5. Arulanandam ARN*, et al.* (1993) The CD58 (LFA-3) binding site is a localized and highly charged surface area on the AGFCC'C" face of the human CD2 adhesion domain. *Proc Natl Acad Sci USA* 90:11613-11617.
6. Reid JD, Parker CE, Borchers CH (2007) Protein arrays for biomarker discovery. *Curr Opin Mol Ther* 9:216-221.
7. Hoover DM, Lubkowski J (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* 30**:**e43-.
8. Altmann F, Staudacher E, Wilson IBH, März L (1999) Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconj J* **16,** 109-123.
9. Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 54:631-664.
10. Li JSS, Li JY (2005) Characterization of N-linked oligosaccharides in chorion peroxidase of *Aedes aegypti* mosquito. *Protein Sci* 14:2370-2386.
11. Norris GE, Flaus AJ, Moore CH, Baker EN (1994) Purification and crystallization of the endoglycosidase PNGase F, a peptide:N-glycosidase from *Flavobacterium meningosepticum*. *J Mol Biol* 241:624-626.
12. Tarentino AL, Plummer TH, Jr (1994) Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. *Methods Enzymol* 230:44-57.
13. Fersht AR (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York).
14. Maxwell KL*, et al.* (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* 14:602-616.
15. Bolen DW, Baskakov IV (2001) The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol* 310:955-963.

16. Holthauzen LMF, Bolen DW (2007) Mixed osmolytes: the degree to which one osmolyte affects the protein stabilizing ability of another. *Protein Sci* 16:293-298.