Supplementary Data

Functional Siglec lectin domains from soluble expression in the cytoplasm of *Escherichia coli*

Johannes M. Pröpster[a], Fan Yang[b,c], Beat Ernst[b], Frédéric H.-T. Allain[a,*] and Mario Schubert[a,d,*]

[a] Institute of Molecular Biology and Biophysics, ETH Zürich, CH-8093 Zürich, Switzerland

[b] Institute of Molecular Pharmacy, Pharmacenter, University of Basel, CH-4056 Basel, Switzerland

[c] Present address: Department of Molecular & Medical Pharmacology, David Geffen School of Medicine at UCLA, CA 90095, USA

[d] Present address: Department of Molecular Biology, University of Salzburg, A-5020 Salzburg, Austria

Table of contents *page*

Supplementary Data

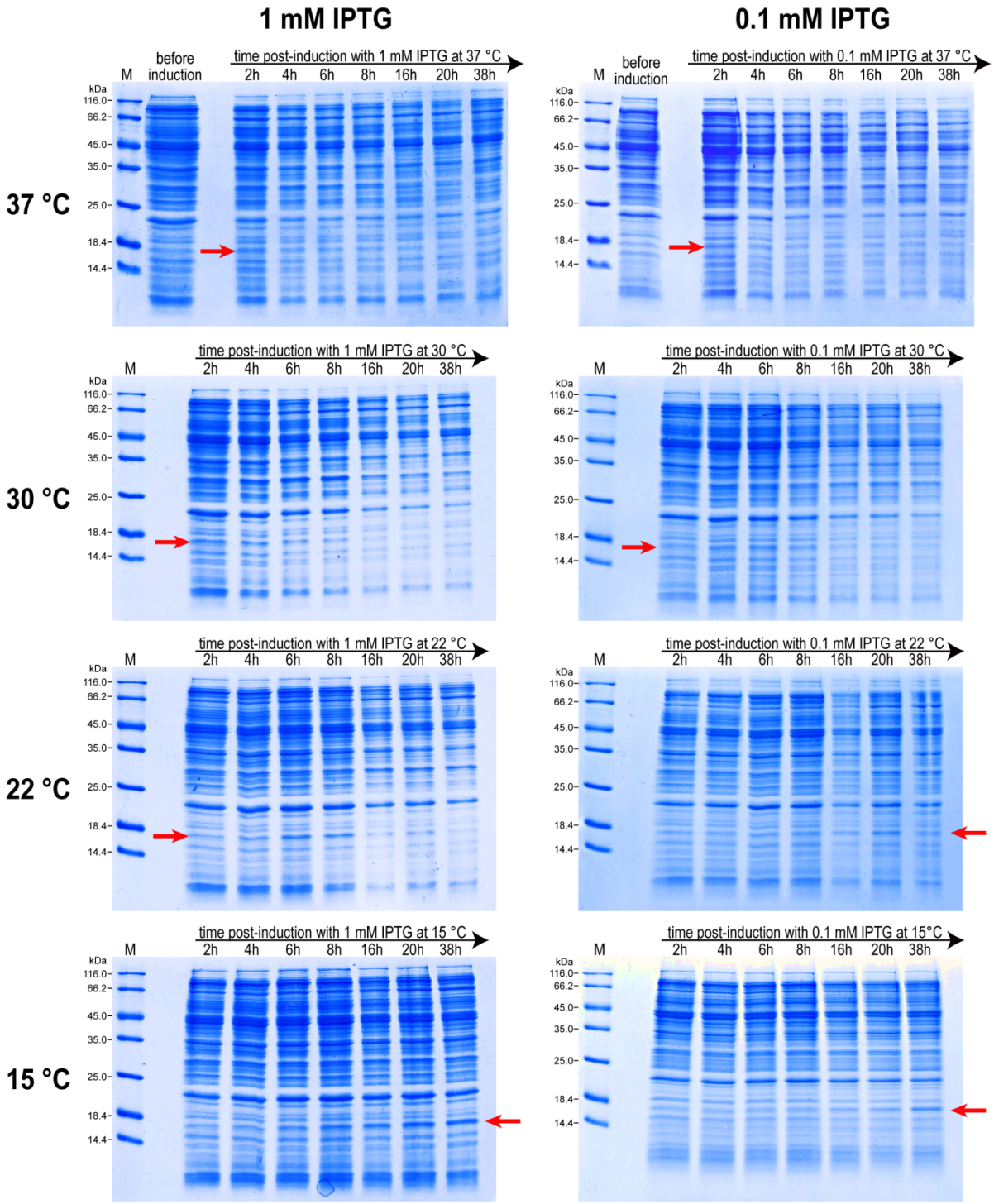

Figure S1. Analytical expressions of soluble Siglec-9 lectin domain in Rosetta-gami B(DE3) as a function of induction temperature, inducer concentration and induction duration. Soluble cell extracts from samples taken before induction and at different time-points after induction with IPTG were separated by SDS-PAGE (15%); gels were stained with Coomassie blue. For better comparability,

approximately equal amounts of samples, corresponding to an original OD_{600} of 0.2 were loaded on each lane. M, molecular weight markers; arrows indicate positions of the Siglec-9 protein band.

Supplementary Data

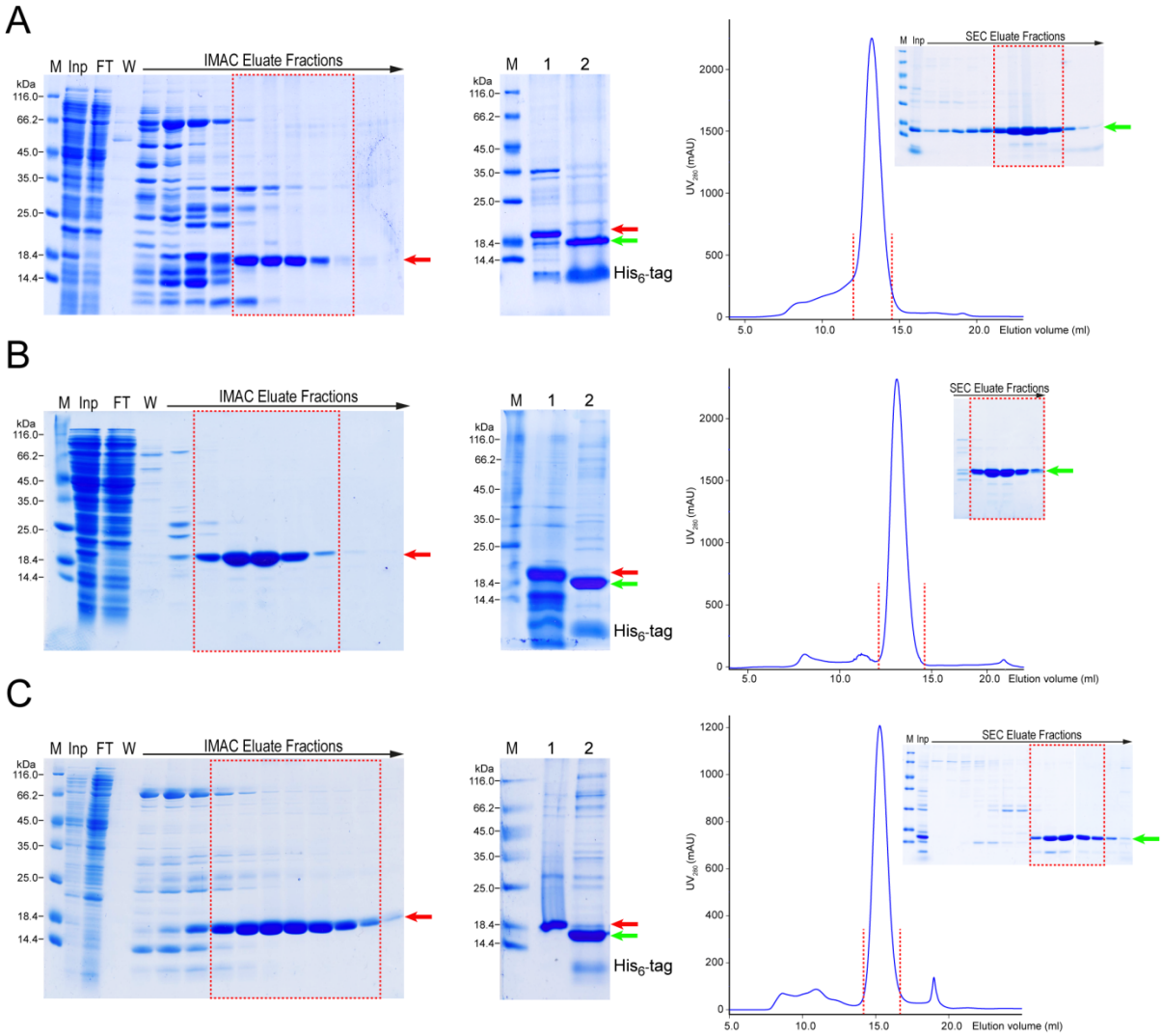

Figure S2. Purification of recombinant (A) Siglec-7, (B) Siglec-8, and (C) Siglec-9 lectin domains. *Left:* Commassie blue-stained SDS-PAGE (15%) gels of the Ni-NTA affinity chromatography (IMAC). *Middle:* Commassie blue-stained SDS-PAGE (15%) gels of pooled and concentrated Ni-NTA affinity chromatography eluate fractions, before (1) and after (2) overnight thrombin cleavage for His_6-tag-removal. *Right:* Size exclusion chromatography (SEC) elution chromatograms and the corresponding Commassie blue-stained SDS-PAGE (15%) gels. M, molecular weight markers; Inp, input; FT, flow-through; W, wash fraction; Arrows indicate the positions of the respective Siglec lectin domain with (*red*) and without (*green*) His_6-tag. Dashed boxes/lines indicate the main peak fractions that were pooled. (SDS-PAGE gels and mass spectra of the final protein preparations are shown in Supplementary Figs. S3 and S4, respectively.)

Supplementary Data

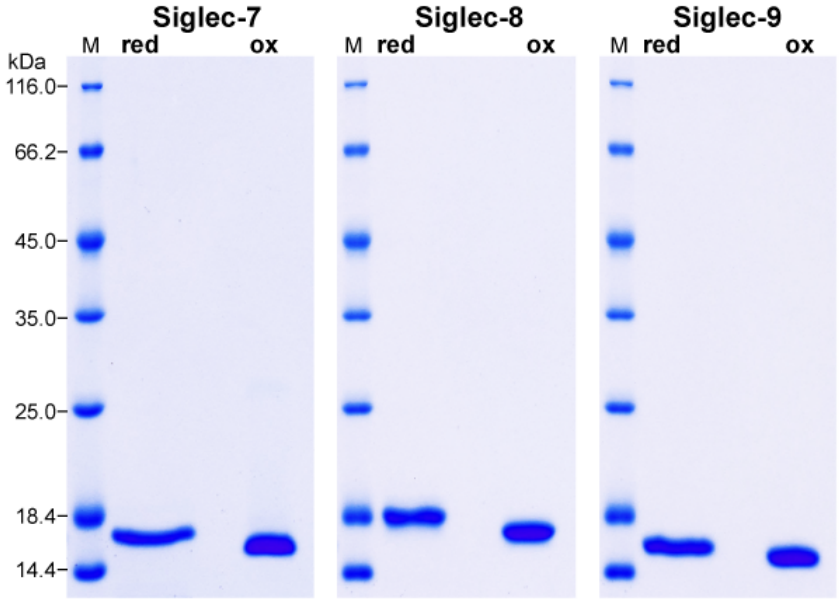

Figure S3. Coomassie blue-stained SDS-PAGE (15%) gels of purified, recombinant Siglec-7, Siglec-8 and Siglec-9 lectin domains under reducing (red) and non-reducing (ox) conditions. M, molecular weight markers. For each lane, samples corresponding to 3 µg of recombinant protein were loaded. Reduced samples were prepared by mixing the protein solutions with reducing 4x SDS sample buffer (16 % v/v β-mercaptoethanol) and subsequent incubation at 98 °C for 10 min. Non-reduced samples were mixed with non-reducing 4x SDS sample buffer (without β-mercaptoethanol) and directly loaded, without prior heating.

Supplementary Data

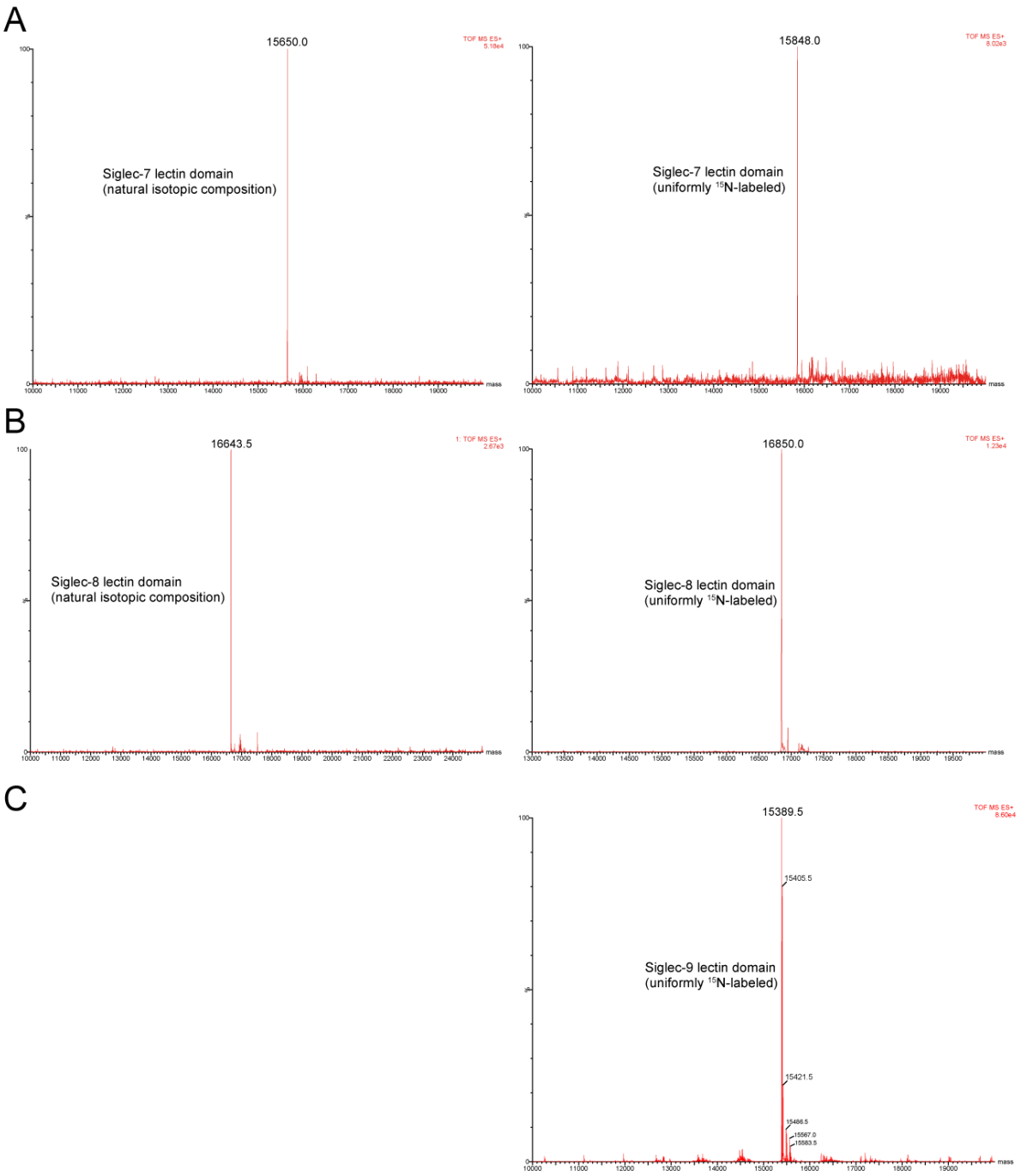

Figure S4. ESI-TOF mass spectra of the purified, recombinant Siglec lectin domains. (A) Siglec-7 with natural isotopic compositions (left panel) and uniform ^{15}N-labeling (right panel); (B) Siglec-8 with natural isotopic compositions (left panel) and uniform ^{15}N-labeling (right panel); (C) Siglec-9 with uniform ^{15}N- isotope-labeling; two minor peaks correspond to oxidized species (+16 Da or +32 Da) formed as side products during MS analysis procedure. Mass spectrum of unlabeled Siglec-9 was not recorded. The theoretical and determined molecular masses for the three recombinant Siglec lectin domains are listed in Supplementary data, Table SIII.

Supplementary Data

Table SI. cDNA clones, bacterial strains, plasmid vectors and oligonucleotides used in this study

cDNA clone, strain, plasmid or oligonucleotide	Relevant characteristics or sequence	Source or reference
cDNA clones		
human Siglec-7 cDNA	I.M.A.G.E. Full Length cDNA clone: IRATp970F1050D Library: 1787/ pCMV-SPORT6 (resistence: ampicillin) Sequence related to NCBI GenBank entry: BC028150	ImaGENES
human Siglec-8 cDNA	I.M.A.G.E. Full Length cDNA clone: IRAUp969G05104D Library: 969/ pOTB7 (resistence: chloramphenicol) Sequence related to NCBI GenBank entry: BC053319	ImaGENES
human Siglec-9 cDNA	I.M.A.G.E. Full Length cDNA clone: IRATp970B1247D Library: 1784/ pCMV-SPORT6 (resistence: ampicillin) Sequence related to NCBI GenBank entry: BC035365 (*corresponding to natural human Siglec-9 variant Lys100Glu*)	ImaGENES
***E.coli* strains**		
TOP10	*Genotype:* F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G λ-	Invitrogen
BL21-CodonPlus(DE3)-RIL	*Genotype: E. coli* B F^- *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^R *gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam^R]	Stratagene
Rosetta-gami B (DE3)	*Genotype:* F^- ompT hsdS_B (r_B^- m_B^-)gal dcm lacY1ahpC (DE3) gor522::Tn10 trxB pRARE (Cam^R, Kan^R, Tet^R)	Novagen
Plasmid vectors		
pET-43.1a(+)	lac operator, T7 promoter, Amp^R, C-terminal His6-tag (*used to construct the Siglec expression vectors for this study*)	Novagen
pET43_Sig7v1_C41S	*(used for recombinant expression of Siglec-7 lectin domain)*	*this study*
pET43_Sig8v1_C42S	*(used for recombinant expression of Siglec-8 lectin domain)*	*this study*
pET43_Sig9v1_C36S	*(used for recombinant expression of Siglec-9 lectin domain)*	*this study*
Oligonucleotides (5 ' to 3 ' sequence)		
PCR-primers used for side-directed-mutagenesis		
S7_C41S_fw	GCAAGAGGGCATGAGCGTCCATGTGCGCTG	Microsynth
S7_C41S_rv	CAGCGCACATGGACGCTCATGCCCTCTTGC	Microsynth
S8_C42S_fw	GTGCAGGAGGGCCTGAGCGTCCATGTGCCCT	Microsynth
S8_C42S_rv	AGGGCACATGGACGCTCAGGCCCTCCTGCAC	Microsynth
S9_C36S_fw	CAGGAAGGCCTGAGCGTCCATGTGCCCTG	Microsynth
S9_C36S_rv	CAGGGCACATGGACGCTCAGGCCTTCCTG	Microsynth
PCR-primers used for insert amplification		
S7_Q19toH148_fw	GGAATTCCATATGGGCCAGAAGAGTAACCGGAAGGATTAC	Microsynth
S7_Q19toH148_rv	TTTCTCGAGGCTACCACGCGGCACCAGGCTGCCGTGGGTCAAGGCTGTCACGTTC	Microsynth
S8_M17toH155_fw	AGGGAATTCCATATGGAGGGAGACAGACAATATGGGGATG	Microsynth
S8_M17toH155_rv	TTTCTCGAGGCTACCACGCGGCACCAGGCTGCCATGGGTCAGGGCTGTCACAAACAC	Microsynth
S9_Q18toH144_fw	GGAATTCCATATGGGCCAGACAAGTAAACTGCTGACGATG	Microsynth
S9_Q18toH144_rv	TTTCTCGAGGCTACCACGCGGCACCAGGCTACCGTGGGTCAAGGCTGTCACATTCAC	Microsynth

Supplementary Data

Table SII. Number of amino acid residues and theoretical molar extinction coefficients of the recombinant Siglec lectin domains

Expression construct	Number of aa residues	Extinction coefficient ε_{280nm}[a] (M^{-1} cm^{-1})
Siglec-7	137	31065
Siglec-8	145	34045
Siglec-9	134	29575

[a]Theoretical molar extinction coefficients were calculated with the ProtParam tool at the ExPASy proteomics server (http://www.expasy.ch/tools/protparam.html).

Table SIII. Rare *E. coli* codons[a] in the nucleotide sequences of the Siglec expression constructs

construct	nucleotide sequence
Siglec-7	atgggccagaagagtaaccggaaggattactcgctgacgatgcagagttccgtgaccgtgcaagagggcatgagcgtc catgtgcgctgctccttctcctacccagtggacagccagactgactctgacccagttcatggctactggttccgggca gggaatgatataagctggaaggctccagtggccacaaacaacccagcttgggcagtgcaggaggaaactcgggaccga ttccacctccttggggacccacagaccaaaaattgcaccctgagcatcagagatgccagaatgagtgatgcgggggaga tacttctttcgtatggagaaaggaaatataaaatggaattataaatatgaccagctctctgtgaacgtgacagccttg acccacggcagcctggtgccgcgtggtagcctcgagcaccaccaccaccaccactaa
Siglec-8	atggagggagacagacaatatggggatggttacttgctgcaagtgcaggagctggtgacggtgcaggagggcctgagc gtccatgtgcccctgctccttctcctaccccaggatggctggactgactctgacccagttcatggctactggttccgg gcaggagacagaccataccaagacgctccagtggccacaaacaacccagacagaagaagtgcaggcagagacccagggc cgattccaactccttggggacatttggagcaacgactgctccctgagcatcagagacgccaggaagagggataagggg tcatatttctttcggctagagagaggaagcatgaaatggagttacaaatcacagttgaattacaaaactaagcagctg tctgtgtttgtgacagccctgacccatggcagcctggtgccgcgtggtagcctcgagcaccaccaccaccaccactaa
Siglec-9	atgggccagacaagtaaactgctgacgatgcagagttccgtgacggtgcaggaaggcctgagcgtccatgtgccctgc tccttctcctaccctcgcatggctggatttaccctggcccagtagttcatggctactggttccgggaaggggccaat acagaccaggatgctccagtggccacaaacaacccagctcgggcagtgtgggaggagactcgggaccgattccacctc cttggggacccacataccgagaattgcaccctgagcatcagagatgccagaagaagtgatgcgggagatacttcttt cgtatggagaaaggaagtataaaatggaattataaacatcaccggctctctgtgaatgtgacagccttgacccacggt agcctggtgccgcgtggtagcctcgagcaccaccaccaccaccactaa

[a]Rare codon analysis was performed using the rare codon calculator tool, RaCC (http://nihserver.mbi.ucla.edu/RACC/). Rare *E. coli* codons are underlined and colored in red (arginine), green (leucine), blue (isoleucine), and orange (proline).

Supplementary Data

Table SIV. Nucleotide and amino acid sequences of the Siglec expression constructs[a]

construct	nucleotide/ amino acid sequence
Siglec-7	atgggccagaagagtaaccggaaggattactcgctgacgatgcagagttccgtgaccgtg (M) G Q K S N R K D Y S L T M Q S S V T V caagagggcatgagcgtccatgtgcgctgctccttctcctacccagtggacagccagact Q E G M S V H V R C S F S Y P V D S Q T gactctgacccagttcatggctactggttccgggcagggaatgatataagctggaaggct D S D P V H G Y W F R A G N D I S W K A ccagtggccacaaacaacccagcttgggcagtgcaggaggaaactcgggaccgattccac P V A T N N P A W A V Q E E T R D R F H ctccttggggacccacagaccaaaaattgcaccctgagcatcagagatgccagaatgagt L L G D P Q T K N C T L S I R D A R M S gatgcggggagatacttctttcgtatggagaaaggaaatataaaatggaattataaatat D A G R Y F F R M E K G N I K W N Y K Y gaccagctctctgtgaacgtgacagccttgacccacggcagcctggtgccgcgtggtagc D Q L S V N V T A L T H G S L V P R (G S ctcgagcaccaccaccaccaccac L E H H H H H H)
Siglec-8	atggagggagacagacaatatggggatggttacttgctgcaagtgcaggagctggtgacg M E G D R Q Y G D G Y L L Q V Q E L V T gtgcaggagggcctgagcgtccatgtgccctgctccttctcctacccccaggatggctgg V Q E G L S V H V P C S F S Y P Q D G W actgactctgacccagttcatggctactggttccgggcaggagacagaccataccaagac T D S D P V H G Y W F R A G D R P Y Q D gctccagtggccacaaacaacccagacagagaagtgcaggcagagacccagggccgattc A P V A T N N P D R E V Q A E T Q G R F caactccttggggacatttggagcaacgactgctccctgagcatcagagacgccaggaag Q L L G D I W S N D C S L S I R D A R K agggataaggggtcatatttctttcggctagagagaggaagcatgaaatggagttacaaa R D K G S Y F F R L E R G S M K W S Y K tcacagttgaattacaaaactaagcagctgtctgtgtttgtgacagccctgacccatggc S Q L N Y K T K Q L S V F V T A L T H G agcctggtgccgcgtggtagcctcgagcaccaccaccaccaccac S L V P R (G S L E H H H H H H)
Siglec-9	atgggccagacaagtaaactgctgacgatgcagagttccgtgacggtgcaggaaggcctg (M) G Q T S K L L T M Q S S V T V Q E G L agcgtccatgtgccctgctccttctcctacccctcgcatggctggatttaccctggccca S V H V P C S F S Y P S H G W I Y P G P gtagttcatggctactggttccgggaaggggccaatacagaccaggatgctccagtggcc V V H G Y W F R E G A N T D Q D A P V A acaaacaacccagctcgggcagtgtgggaggagactcgggaccgattccacctccttggg T N N P A R A V W E E T R D R F H L L G gacccacataccgagaattgcaccctgagcatcagagatgccagaagaagtgatgcgggg D P H T E N C T L S I R D A R R S D A G agatacttctttcgtatggagaaaggaagtataaaatggaattataaacatcaccggctc R Y F F R M E K G S I K W N Y K H H R L tctgtgaatgtgacagccttgacccacggtagcctggtgccgcgtggtagcctcgagcac S V N V T A L T H G S L V P R (G S L E H caccaccaccaccac H H H H H)

[a]N- and C-terminal extensions of the natural amino acid sequences are shaded in grey; amino acid residues not present in the purified proteins are enclosed in parenthesis; the Cys-to-Ser point mutation (replacing the natural Cys involved in the inter-domain disulfide bond to the subsequent C2-set Ig domain) is shaded in yellow.

Supplementary Data

Table SV. Molecular weights (MW) of the recombinant Siglec lectin domains

Recombinant Siglec lectin domain, chemical formula and isotope labeling	**predicted MW[a] (Da)**	**determined MW by ESI-MS[b] (Da)**
Siglec-7 ($C_{685}H_{1047}N_{201}O_{210}S_6$**)**		
natural abundance	15650.3	15650.0
(100% ^{1}H/ 99% ^{15}N)	15848.0	15848.0
Siglec-8 ($C_{738}H_{1115}N_{209}O_{225}S_4$**)**		
natural abundance	16643.3	16643.5
(100% ^{1}H/ 99% ^{15}N)	16848.9	16850.0
Siglec-9 ($C_{673}H_{1021}N_{201}O_{196}S_4$**)**		
natural abundance	15191.8	ND[c]
(100% ^{1}H/ 99% ^{15}N)	15389.5	15389.5

[a]Mass prediction was computed using the BMRB Molecular Mass Calculator (http://www.bmrb.wisc.edu/metabolomics/mol_mass.php).

[b]Mass determination by electrospray-ionization mass spectrometry (ESI-MS) was performed at the protein analysis service platform (B-Fabric) of the Functional Genomics Center Zurich, Switzerland (http://www.fgcz.ethz.ch).

[c]ND: not determined

Supplementary Data

Supplementary Methods

Circular dichroism (CD) spectroscopy

Recorded CD data obtained in millidegrees were normalized to mean residue ellipticity $[\theta_{mrw,\lambda}]$ in degrees · cm^2 · $decimols^{-1}$ using equation 1 (Eq. S1)

$$[\theta_{mrw,\lambda}] = \theta_\lambda / (n_{pb} \cdot c \cdot l \cdot 10) \qquad (\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}) \qquad \textbf{(Eq. S1)}$$

where θ_λ is the observed CD signal at wavelength λ, n_{pb} is the number of peptide bonds in the protein backbone (number of aa residues – 1), c is the molar concentration (mol/l) and l is the cell path length (cm). Since the thermal unfolding transitions for all three Siglec lectin domains proved to follow a two-state irreversible process, any thermodynamic analysis was precluded. As a measure of protein stabilities, apparent melting temperatures, corresponding to the temperature at the transition midpoints (T_m) were obtained by fitting the CD data to equation 2 (Eq. S2), as described by Pace *et al.* [1]

$$\theta_{210} = [(m_F \cdot T + y_F) + (m_U \cdot T + y_U) \cdot e^{(\Delta H_m/RT) \cdot ((T - T_m)/T_m)}] / (1 + e^{(\Delta H_m/RT) \cdot ((T - T_m)/T_m)}) \quad \textbf{(Eq. S2)}$$

where θ_{210} is the observed CD signal at 210 nm wavelength, m_F and m_U are the slopes and y_F and y_U are the intercepts of the pre- and post-transition baselines, respectively, T is the temperature, T_m is the apparent melting temperature, and ΔH_m is the enthalpy change of unfolding at T_m. In order to compare the thermal transition curves of the three different Siglec lectin domains, the ellipticity readings were further normalized to fraction of unfolded protein, f_U, using equation 3 (Eq. S3)

$$f_U = (\theta_{210} - \theta_F) / (\theta_U - \theta_F) = [\theta_{210} - (m_F \cdot T - y_F)] / [(m_U \cdot T - y_U) - (m_F \cdot T - y_F)] \qquad \textbf{(Eq. S3)}$$

where θ_F and θ_U correspond to the CD signal of the protein in the native (folded) and denatured (unfolded) state, respectively. Curve fitting was performed using the software KaleidaGraph v. 4.0 (Synergy Software, Reading, PA, USA).

Supplementary Data

Supplementary References

[1] C.N. Pace, E.J. Hebert, K.L. Shaw, D. Schell, V. Both, D. Krajcikova, J. Sevcik, K.S. Wilson, Z. Dauter, R.W. Hartley, G.R. Grimsley, Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. J Mol Biol 279 (1998) 271-286.