C-type lectins CTL4 and CTLMA2: conserved heterodimeric structure and glycan specificity in *Anopheles* mosquitoes.

Ritika Bishnoi[1], Alicia Contet[2], Christopher J Day[3], David Chun-Feng Hou[1], Lauren A. Profitt[4], Deepak Singla[5], Gregory L. Sousa[6], Michael P Jennings[3], Michael Povelones[6], Ann M. Valentine[4], Richard H. G. Baxter[1†]

[1]Dept. of Medical Genetics & Molecular Biochemistry, Lewis Katz School of Medicine at Temple University, 3440 North Broad Street, Philadelphia PA 19140.

[2]Centre d'Immunologie Pierre Fabre, 74160 St. Julien-en-Genevois, France

[3]Institute of Glycomics, Griffith University, Queensland, Australia.

[4]Dept. of Chemistry, Temple University, Philadelphia PA.

[5]Laboratory of Host-Parasite Interaction Studies, National Institute of Malaria Research, Dwarka, India

[6]School of Veterinary Medicine, University of Pennsylvania, Philadelphia PA

[†]Corresponding Author. ORCID: 0000-0001-5140-1726. Email: rbaxter@temple.edu

Abstract

The C-type lectins CTL4 and CTLMA2 cooperatively influence *Plasmodium* infection in the malaria vector *Anopheles*. Here we report the purification and biochemical characterization of CTL4 and CTLMA2 from *An. gambiae* and *An. albimanus*. CTL4 and CTLMA2 are known to form a disulfide-bridged heterodimer via an N-terminal tri-cysteine CXCPC motif. We demonstrate in vitro that CTL4 and CTLMA2 intermolecular disulfide formation is promiscuous within this motif. Furthermore, CTL4 and CTLMA2 exhibit charge complementarity that promotes the formation of higher oligomeric states at physiological pH. Both lectins bind specific sugars, with an apparent preference for glycosaminoglycan motifs comprising β1-3/β1-4

linkages between glucose (Glc), galactose (Gal) and their respective hexosamines. Small-angle x-ray scattering data supports a compact heterodimer between the CTL domains. Recombinant CTL4/CTLMA2 is functional *in vivo*, reversing the enhancement of phenoloxidase activity in ds*CTL4*-treated mosquitoes. We propose these molecular features underline a common function for CTL4/CTLMA2 in mosquitoes, with species and strain-specific variation in degrees of activity in response to *Plasmodium* infection.

Author Summary

Mosquitoes of the genus *Anopheles* are vectors for the single-celled parasite *Plasmodium*, the causative agent of malaria. Mosquitoes, like all insects, utilize the process of melanization for both wound healing and defense against pathogens. CTL4 and CTLMA2 are two proteins found in *Anopheles* mosquitoes that act as inhibitors of melanization, so understanding their molecular function is important to understanding the immune response of *Anopheles* mosquitoes to *Plasmodium* infection. We have purified CTL4 and CTLMA2 from two species of *Anopheles* and studied their molecular properties with a variety of biochemical and biophysical techniques. We also verified that our purified protein is functional by injecting it into mosquitoes. We learned that CTL4 and CTLMA2 are joined together by a disulfide bond between any one of three cysteine residues near the N-terminus of each protein. The CTL4/CTLMA2 complex is compact, but can associate into larger structures in solution, probably because of a loop in each protein that carries an opposite charge. The proteins cooperatively bind calcium and sugars, specifically glycosaminoglycan sugars, which are typically present in the connective tissues of insects. This information will aid in further investigations of the function of CTL4 and CTLMA2.

Introduction

Malaria, the world's most devastating parasitic disease, is transmitted between humans by mosquitoes of the *Anopheles* genus. *An. gambiae* is the principal malaria vector in Sub-Saharan Africa. The innate immune response of *An. gambiae* to malaria parasites (genus *Plasmodium*) in an infectious blood meal is a significant factor influencing the prevalence and intensity of infectious mosquitoes in a population [1, 2]. Understanding the molecular interactions and mechanisms of *Anopheles* immunity is therefore key to comprehending, predicting, and potentially controlling disease transmission.

An. gambiae has a complement-like immune response centered upon thioester-containing protein 1 (TEP1) that effectively targets *Plasmodium* ookinetes following their traversal of the midgut epithelium, prior to their transformation into oocysts [3-7]. The immune response to *Plasmodium* involves additional proteins such as the leucine-rich immune molecule (LRIM) family [8], CLIP proteases [9], and other families. Two LRIM family members, LRIM1 and APL1C, form a heterodimeric complex that directly interacts with TEP1 to regulate its anti-*Plasmodium* activity [10-12]. Yet the extent of interactions between these and other immune factors and mechanistic details of the mosquito immune response remain in large part unknown.

The C-type lectins CTL4 and CTLMA2 were first reported to influence the immune response of *An. gambiae* to *P. berghei* infection coincident with that of the first LRIM family member LRIM1 [13]. RNAi knockdown of TEP1 (ds*TEP1*) [3] and LRIM1 (ds*LRIM1*) [13] results in increased numbers of oocysts in the midgut, indicating a refractory or antagonist phenotype. Knockdown of CTL4 (ds*CTL4*) or CTLMA2 (ds*CTLMA2*) results in significantly reduced numbers of oocysts and increased melanization, implying a susceptible or agonist phenotype.

In the *An. gambiae* L3-5 strain, parasites targeted by TEP1 are killed by lysis, followed by melanization of corpses [14, 15]. In the *An. gambiae* G3 strain, melanization requires silencing of CTL4, in which case it may lead to killing independently of lysis [15]. Melanization in the

absence of CTL4 or CTLMA2 required the function of LRIM1. This suggests that CTL4 and CTLMA2 act to suppress either the targeting of *Plasmodium* ookinetes or the downstream melanization response [15].

CTL4 and CTLMA2 cooperate to protect mosquitoes from infection with Gram-negative bacteria [16]. Either ds*CTL4* or ds*CTLMA2* resulted in decreased survival following infection with *E. coli* but not infection with *S. aureus*. Although ds*CTL4* increased the melanization of *Plasmodium* parasites, it was not observed to increase phenol oxidase activity in the hemolymph following bacterial challenge. Intriguingly, CTL4 and CTLMA2 form a disulfide-bridged heterodimer via an N-terminal CXCXC motif that is necessary for their stability in the hemolymph, analogous to the heterodimer between the LRR proteins LRIM1 and APL1C.

The agonist effect of CTL4/CTLMA2 on *P. berghei* was not initially replicated using the human malaria parasite *P. falciparum* [17]. However, this reflects the naturally lower level of infection intensity for the human parasite; increased infection levels resulted in melanization of *P. falciparum* upon CTL4/CTLMA2 kd [18], and most importantly, strain-specific mosquito parasite interactions and the ability of some parasite strains to evade the mosquito immune response [7, 19]. The phenotype of CTL4/CTLMA2 also depends on the specific host species. In *An. albimanus* CTL4/CTLMA2 kd is antagonistic towards both *P. berghei* and *P. falciparum* [18]. This has led to the hypothesis that the functions of these two proteins or associated cofactors, have diverged and their mechanistic involvement in regulating infection intensity and melanization is unlinked.

These data led us to address the question, what is the molecular structure and function of CTL4/CTLMA2 and to what extent is it conserved throughout the *Anopheles* genus? We report that CTL4/CTLMA2 intermolecular disulfide bond formation can occur via any two cysteines of the CXCXC motif, and that the two proteins can form higher-order oligomers via complementary electrostatic interactions. The solution structure of CTL4 and CTLMA2 was determined by small-

angle x-ray scattering. Analysis of glycan binding of CTL4 and CTLMA2 suggest the heterodimer separately and synergistically recognize β1-3 and β1-4 glucose/galactose linkages. CTL4 knockdown results in increased PO activity following *E. coli* challenge, which is reversed by co-administration of recombinant protein. Taken together, these results suggest a conserved molecular function of these two lectins in the regulation of melanization downstream of immune recognition in anopheline mosquitoes.

Results

Conservation of CTL4/CTLMA2 in Anopheles

CTL4 and CTLMA2 both consist of a signal peptide, a short N-terminal sequence containing the CXCXC motif, and a single CTL domain. A recent report suggested that the function of CTL4 and CTLMA2 or their cofactors have diverged within *Anopheles*, and specifically that *An. albimanus* CTL4 does not contain the N-terminal cysteine residues involved in disulfide linkages [18]. We first reexamined the orthologs of *CTL4* (AGAP005335) and *CTLMA2* (AGAP005334) in Vectorbase [20]. The current gene models (as of February 2019) include proteins with an N-terminal CXCXC motif for 15/16 CTL4 orthologs including *An. albimanus* AALB014534, and 13/14 CTLMA2 orthologs. We performed multiple sequence alignments of both CTL4 and CTLMA2 from 10 Asian, African and New World *Anopheles* species (Fig. 1). The unrooted phylogenetic trees have almost identical topology, and a combined phylogenetic tree has two symmetric branches, except for the position of *An. dirus* with respect to *An. funestus* and *An. maculatus*.

This supports the hypothesis that *CTL4* and *CTLMA2* are wholly conserved within the *Anopheles* genus, with missing orthologs reflecting incomplete annotation of certain genomes. The two genes have a close back-to-back orientation on chromosome 3L. We examined the genomic region of three species with a reported CTL4 ortholog but no *CTLMA2* ortholog: *An. stephensi* ASTE002637, *An. minimus* AMIN007380, and *An. melas* AMEC014491. All

possess a close or overlapping gene in inverse orientation – ASTE002636, AMIN007379 and AMEC088499, comprising two CTL domains. For *An. stephensi* and *An. minimus* the second CTL is preceded by a CXCXC motif, suggesting they may be CTLMA2 orthologs.

An. farauti has a *CTLMA2* ortholog (AFAF005660) but no corresponding *CTL4* ortholog or close, inverted CTL-containing gene. A *CTLMA2* ortholog including an N-terminal CXCXC motif is annotated in *Aedes albopictus*, AALF001196. Furthermore, AALF001196 has a close back-to-back inverted two-exon gene AALF001195, comprising of a serine protease and a CTL domain. The October 2018 gene model for *Ae. aegypti* included a *CTLMA2* ortholog with N-terminal CXCXC motif, AAEL014382, but is absent in the current gene model. A *CTLMA2* ortholog is annotated in *Culex quinquefasciatus,* CPIJ000443, but lacks the N-terminal CXCXC motif. This suggests that *CTLMA2*, and perhaps *CTL4* arose in a common ancestor of *Anopheles* and *Aedes*.

Biochemical Characterization

CTL4, CTLMA2 and the heterodimer were expressed in insect cells using the baculovirus expression vector system (BEVS) and purified to homogeneity. Mature CTL4 (25–177) has a molar mass of 17.3 kDa and pI 7.7. Mature CTLMA2 (18–174) has a molar mass of 17.9 kDa and pI 4.5. The purified heterodimer eluted as a single peak on size-exclusion chromatography (SEC) (Fig. 2A) with an apparent molecular weight of 40 kDa. The purified heterodimer has a molecular weight on non-reducing SDS-PAGE of 35 kDa. Monomeric CTL4 and CTLMA2 appear on reducing SDS-PAGE at 17 kDa and 20 kDa, respectively. As previously noted, the increased apparent molecular weight of CTLMA2 may reflect its unusual (acidic) amino acid distribution [16].

We hypothesized that the CXCXC motif in CTL4/CTLMA2 is a promiscuous disulfide bonding motif. We constructed nine mutants containing a single cysteine in the N-terminal CXCXC motif of CTL4 and CTLMA2, co-expressed the proteins in Sf9 cells, and performed α6xHis Western

Blotting to detect CTL4. A heterodimer was detected on non-reducing SDS-PAGE in all cases (Fig. 2B). CTL4 has an apparent preference for C39 vs. C41, C43 for intermolecular disulfide formation; no similar preference is evident for CTLMA2. This supports our hypothesis that N-terminal intermolecular disulfide formation is promiscuous, which in turn suggests the CXCXC motif does not include an internal disulfide between any two cysteines.

If intermolecular disulfide formation is promiscuous, CTL4/CTLMA2 heterodimers could in principle form multivalent disulfide-bridged oligomers. No disulfide-linked oligomers are detected on non-reducing SDS-PAGE. However, higher-order non-covalent oligomerization is evident as a shoulder of the main peak in SEC (Fig. 2A); no such shoulder is evident for either monomer. We confirmed these species by sedimentation-velocity analytical ultracentrifugation (AUC) (Fig. 2C). At pH 7.5 a series of species of decreasing intensity is observed in the $c(s)$ distribution: s_1 = 3.1 s, s_2 = 4.4 s, s_3 = 5.9 s. We used dynamic light scattering (DLS) to analyze CTL4/CTLMA2 size distribution. We found that self-association is inversely correlated with pH. A trend of increasing apparent molecular weight, radius and polydispersity is observed from pH 9.5 to pH 6.0. Below pH 6.0 significant aggregation and precipitation occurs.

Calcium binding

We measured the calcium binding affinity of *An. gambiae* CTL4, CTLMA2 and the CTL4/CTLMA2 heterodimer of *An. gambiae* and *An. albimanus* by isothermal titration calorimetry (ITC). Under equivalent conditions, binding was observed for CTLMA2 and CTL4/CTLMA2, but not for CTL4 (Fig. 3). Binding constants and thermodynamic parameters were calculated from the results of three independent experiments (Table 1). CTLMA2 Ca^{2+} binding is well fit by a single site model with K_D = 173±27 μM, ΔH = 12±2 kcal/mol. The affinity of *An. gambiae* CTL4/CTLMA2 for calcium is ~40× higher than CTLMA2 with K_D = 4.9±0.5 μM, ΔH = -23±4 kcal/mol. *An. albimanus* CTL4/CTLMA2 has a similar affinity for calcium with K_D = 2.82 μM, ΔH = -12.1 kcal/mol.

Glycan binding

CTL4 and CTLMA2 belong to the lineage of myeloid CTLs - including macrophage mannose receptor (MMR) and DC-SIGN – that form a conserved family of immune receptors in metazoans [21, 22]. Among CTLs with known structure, CTLMA2 has 30% sequence identity to the carbohydrate recognition domain (CRD) of mouse scavenger receptor Crd-4 (SCRL, PDB ID 2OX9) [23] and porcine surfactant protein D (SP-D, PDB ID 4DN8) [24]. CTLMA2 conserves residues associated with Ca^{2+} binding in the glycan binding loop, and the canonical EPN motif associated with D-mannose selectivity (Fig. 4A). In contrast, CTL4 lacks all residues associated with Ca^{2+} binding, consistent with the fact that no binding was observed by ITC. The only CTL of known structure with significant similarity to CTL4 is factor IX/X binding protein (X-bp) from the venom of the Chinese moccasin *Deinagkistrodon acutus* (1IOD). X-bp is a modified CTLD in which the glycan binding domain is replaced by a long loop that mediates dimerization to generate the Factor IX/X binding site. Hence, although some insect CTLs do not require calcium for binding, it is unclear if CTL4 should bind glycans at all.

In order to define their lectin activity, we analyzed CTL4, CTLMA2 and the CTL4/CTLMA2 heterodimer on glycan arrays that display 367 unique glycan structures (see Methods) [25]. These studies demonstrated binding to a range of glycans (Table 2). Array hits were validated by surface plasmon resonance (SPR) (S1 Table). The monomers of CTL4 and CTLMA2 demonstrated significant binding to only four and six glycans, respectively, whereas the heterodimer bound 18 different glycans. There is appreciable difference between ligands bound by the individual monomers and the heterodimer; 3/4 (75%) of glycans recognized by CTL4 and 2/6 (33%) of glycans recognized by CTLMA2 were not recognized by CTL4/CTLMA2.

The CTLs did not recognize mannose-containing glycans, with the exception of manose-6-phosphate by CTL4/CTLMA2, despite the canonical EPN motif present in CTLMA2. Rather, the structures recognized are generally glycosaminoglycan motifs comprising β1-3/β1-4 linkages

between glucose (Glc), galactose (Gal) and their respective hexosamines GlcNac and GalNac. The Galβ1-4Glc linkage was present in 12/23 (52%) of glycans recognized, including 6/23 (26%) containing Galβ1-4GlcNac and 4/23 (17%) containing the keratan motif Galβ1-4GlcNac β1-3Gal or GlcNac β1-3Galβ1-4Glc.

The array also displayed some preference for polymeric and sulfated glycans. Of the four glycans recognized by CTL4, the most significant was globopentaose (Gb5, Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glc); CTL4 also bound HA 160 kDa $(GlcA\beta1\text{-}3Glc\mathit{N}Ac\beta1\text{-}4)_n$ and β1-3Glucan $(Glc\beta1\text{-}4Glc)_n$. Three of five glycans recognized by the CTLMA2 monomer were sulfated, and the CTL4/CTLMA2 heterodimer recognized chondroitin sulfate and chondroitin-6 sulfate, hyaluroninc acid $(GlcA\beta1\text{-}4GlcNAc\beta1\text{-}3)_8$, and HA 160 kDa $(GlcA\beta1\text{-}3Glc\mathit{N}Ac\beta1\text{-}4)_n$. These results suggest there are distinct and synergistic effects of heterodimerization on glycan binding.

Structural Analysis

To further probe the structure of CTL4 and CTLMA2, we generated structural models of CTL4 and CTLMA2 (Fig. 4B) using *MODELLER* [26] with additional manual editing. Both CTL4 and CTLMA2 have an extended loop (loop 1) following the second helix of the CTLD (Fig. 4A) with a high density of complementary charged residues; basic residues for CTL4 and acidic residues for CTLMA2. The complementary electrostatics of these loop 1 residues and their proximity to the N-terminal CXCXC motif suggest this as a potential protein/protein interface within the heterodimer Fig. 4B), which was modeled commencing from a single disulfide bond between CTL4 C43 and CTLMA2 C38. However, the glycan/Ca^{2+} binding loops are a second potential interface, as observed for the two chains of *D. acutus* X-bp.

To validate our model we analyzed the solution structure of CTL4/CTLMA2 by small angle x-ray scattering (SAXS). Experiments were conducted in 0.5 M NaCl, 20 mM CHES pH 9.0, 0.5 mM $CaCl_2$, and 1% glycerol to minimize interparticle interactions. In these conditions, the protein displayed a linear relation between extrapolated intensity at zero angle and concentration (I_0 vs.

c) up to a concentration of 3.1 mg/ml. The buffer-subtracted curve of intensity *I* vs. *q* (Fig. 4C) was submitted to *SAXSMoW2* which yields R_G = 23.0 Å (I_0=0.61) and molecular weight MW = 39 kDa, only 11% greater than the expected heterodimer MW of 35 kDa [27]. However, the calculated Guinier plot is based on only seven data points; fitting over an extended range (Fig. 4C, inset) yields R_G = 24.5 Å (I_0 = 0.64), while fitting the pairwise distribution function *P*(*r*) (Fig. 4D) yields an R_G = 25.4 Å (I_0 = 0.65). The *P*(*r*) distribution fit by *DAMMIF* [28] with 20 *ab initio* bead models with an average normalized structural discrepancy NSD = 1.0 ± 0.2. The main body of the *ab initio* models are similar in shape to the expected CTL4/CTLMA2 heterodimer.

Additional density extends from the main body of the bead models that can reflect either the N-terminal sequence of both proteins including the CXCXC motif, or a minor population of CTL4/CTLMA2 with a higher radius gyration. To test this hypothesis we performed multi-state modeling of the SAXS profile with the program *MULTIFOXS* [29]. We generated a complete model for CTL4-6xHis/CTLMA2 with an N-terminal coiled-coil terminated by an intermolecular disulfide between CTL4 C39 and CTLMA2 C34. MULTIFOXS generated an ensemble of 10,000 variants of the model to compare to the experimental scattering curve. The only flexible residues were CTL4 40–45 and 178–183 (6xHis) and CTLMA2 35-39, with the N-terminal coiled-coil serving as a rigid body connecting the two chains. The best one-state model fit the data with χ^2 = 1.13 and had R_G = 23.7 Å. The minimum χ^2 = 1.07 was achieved with a 3-state model, in which 80% of the scattering is contributed by two compact models with R_G = 22.0 Å And R_G = 23.6 Å. This data is consistent with formation of a compact heterodimer between CTL4 and CTLMA2.

CTL4/CTLMA2 inhibit phenol oxidase activation in response to E. coli

CTL4 and CTLMA2 function as inhibitors of the mosquito melanization response to infection. It was previously reported that CTL4 knockdown did not lead to increased phenol oxidase (PO)

activity in response to infection with a mixture of *E. coli* and *S. aureus* [16]. The same study however, found that ds*CTL4* and ds*CTLMA2* mosquitoes were specifically susceptible to Gram-negative bacteria. Hence, we re-examined the effect of CTL4 and CTLMA2 knockdown on hemolymph PO activity in response to only *E. coli* infection (Fig. 5A).

At 4 h post-infection with *E. coli*, ds*CTL4* and ds*CTLMA2* mosquitoes had significantly enhanced PO activity compared to *dsLacZ* controls (Fig. 5B). The average knockdown efficiency for *CTL4* and *CTLMA2* was 93±3% and 89±3%, respectively, with >80% in any single experiment. Co-administration of recombinant CTL4/CTLMA2 with *E. coli* reversed the enhancement of PO activity in ds*CTL4* mosquitoes compared to BSA controls (Fig. 5C). These results demonstrate that CTL4/CTLMA2 is directly involved as a negative regulator of PO activity. However, co-administration of CTL4/CTLMA2 had no significant effect on the level of *E. coli*-induced PO activity in ds*LacZ* mosquitoes, and levels of *E. coli*-induced PO activity was significantly higher in ds*CTL4* mosquitoes co-injected with recombinant CTL4/CTLMA2 were higher than that of ds*LacZ* mosquitoes co-injected with BSA. Hence, the rescue effect of injecting recombinant CTL4/CTLMA2 is incomplete compared to the presence of endogenous protein.

Discussion

Here we describe the biochemical characterization of two C-type lectins, CTL4 and CTLMA2, that play key roles as negative regulators of the melanization cascade in the malaria vector *An. gambiae*. This is relevant to malaria transmission as knockdown of CTL4 and CTLMA2 results in reduced susceptibility to *P. berghei* and, with comparable infection levels, *P. falciparum*. The effect of CTL4 and CTLMA2 on *Plasmodium* is species-dependent, suggesting divergent evolution of the trait during speciation within the *Anopheles* genus. Here we focus on the likely common molecular features of these proteins conserved throughout *Anopheles*.

First, heterodimerization of CTL4 and CTLMA2 is mediated by promiscuous intermolecular disulfide bonding of an N-terminal CXCXC motif conserved throughout *Anopheles* and, for CTLMA2, the *Aedes* genus. Aside from Zn-binding sites in DNA/RNA binding proteins, the CXCXC motif is uncommon in proteins of known structure. At time of publication the motif is present in only 74 representative protein-only structures at <4.0 Å resolution in the Protein Databank. The sequences are generally linear with disulfide bonds (if any) to different strands.

Angepoietin-1 has an internal disulfide bond between the first and second cysteines of a CXCXC motif, with the backbone carbonyls serving as ligands to a calcium binding site. This is not observed for CTL4/CTLMA2 since we only observe a single Ca^{2+} binding site by ITC corresponding to the CTLMA2 Ca^{2+}/glycan loop. Steric hindrance tends to limit the formation of multiple disulfide between adjacent strands of closely spaced cysteine residues. Hence, it is not surprising that although any two cysteines between CTL4 and CTLMA2 can mediate heterodimerization, there is no evidence of appreciable higher order oligomerization mediated by disulfide bonding.

However, we see clear evidence that CTL4 and CTLMA2 form non-covalent higher-order oligomers by SEC, AUC and DLS. The constant ratio of sedimentation coefficients measured by AUC $s_2/s_1 = s_3/s_2 = 1.4$ is consistent with a 1, 2, 4, … series (CTL4/CTLMA2, $CTL4_2CTLMA2_2$, $CTL4_4CTLMA2_4$, etc.) of oligomerization. Within the CRD we hypothesize that two loops are potential mediators of higher order interaction. First, the extended loop 1 with charge complementarity is a feature of both CTLs that (i) distinguishes them from structural homologs such as Crd-4, and (ii) is conserved throughout *Anopheles* (but, interestingly, not *Aedes*). A further argument suggesting this loop is responsible is the pH dependence of oligomerization seen by DLS, because at high pH the basic residues in CTL4 loop 1 become progressively less charged.

Second, the Ca^{2+}/glycan binding loop, which (i) for CTL4 does not bind Ca^{2+} but is homologous to the dimerization loop of X-bp, (ii) for CTLMA2 Ca^{2+} affinity is enhanced in the presence of CTL4 suggesting a physical interaction either directly or allosterically alters its structure. The CTL4/CTLMA2 Ca^{2+} binding curve is best fit by substoichiometric binding (N=0.36). Partial occupancy of the binding site by residual calcium (present in buffers or incompletely removed by dialysis against 5 mM EDTA) or partial occlusion of the binding site on CTLMA2 (by interaction with CTL4 or perhaps residual glycan carried over from conditioned media during purification) are possible sources of substoichiometric binding. The observed oligomerization of the heterodimer may contribute to both these experimental factors, just as it may serve to enhance the recognition of polymeric glycans through avidity.

The results of glycan array analysis suggest both CTL4 and CTLMA2 specifically recognize glycosaminoglycans comprising β1-3/β1-4 linkages between glucose (Glc), galactose (Gal) and their respective hexosamines. For CTLMA2 this is consistent with its structural homology to mouse scavenger receptor Crd-4 (PDB ID 2OX9), which specifically recognizes the Lewis^{x} trisaccharide Galβ1-4(Fucα1-3)GlcNAc [30]. Our prior hypothesis for glycan binding was that CTL4, having a mutated Ca^{2+}/glycan binding loop, would not recognize glycans. This specificity is consistent with the functional role of the CTL4/CTLMA2 in melanization. Insect connective tissues are rich in acidic and neutral glycosaminoglycans (a.k.a. mucopolysaccharides, mucins), and serve to cement hemocytes together to encapsulate parasites and foreign bodies in the hemocoel [31]. CTL4/CTLMA2 may tend to line connective tissues in mosquitoes, thereby inhibiting self-melanization in response to injury or infection, analogous to the association of vertebrate Factor H with sialylated surfaces, which serves to inhibit complement activation on self surfaces.

A number of questions remain unanswered. Does CTL4/CTLMA2 interact with specific proteins as well as glycans, e.g. TEPs, LRIM proteins, CLIPs or other factors in the mosquito? Does

CTL4/CTLMA2 inhibit melanization even after PO is activated, or does it simply inhibit the proteolytic activation of PO or an upstream factor? Is the binding of protein cofactors dependent on their conformation or proteolytic activation (i.e. TEPs, CLIPs)? Do putative cofactors require CTL4/CTLMA2 bind glycan? Is the self-association of CTL4/CTLMA2 relevant or affected by these factors? The ability to reverse enhanced PO activity in ds*CTL4* mosquitoes with co-administration of CTL4/CTLMA2 allows these questions to be addressed by site-specific alterations to the recombinant protein. Elucidating the mechanism of CTL4/CTLMA2 in the melanization response to infection can ultimately address its role in the natural susceptibility or refractoriness of different *Anopheles* species to *Plasmodium* infection.

Materials and Methods

Protein Expression and Purification

Full-length *An. gambiae* CTL4 (AGAP005335) and CTLMA2 (AGAP005334), *An. albimanus* CTL4 (AALB014534) and CTLMA2 (AALB005905) were obtained by total gene synthesis and subcloned into pFastbac1 with C-terminal 6×His tag for the production of recombinant baculovirus. For heterodimeric CTL4/CTLMA2 the genes were cloned into pFastbac-Dual with a C-terminal 6×His tag on CTL4. Additional constructs were cloned into pFB-GP67-Hta with a TEV-cleavable 6×His tag. Proteins were expressed in T.ni cells cultured in ESF-921 media (Expression Systems LLC). Conditioned media was harvested at 48 h post-infection (hpi) concentrated and diafiltrated with tangential flow filtration, followed by affinity chromatography using Co Talon resin (Clontech). Further purification was accomplished by ion exchange and size-exclusion chromatography on an AKTA PURE system.

Western Blotting

Sf9 cells were infected at MOI of 0.1 and conditioned media collected 96 hpi. For non-reducing SDS-PAGE DTT was excluded from the 6× loading buffer; samples were not heated prior to

electrophoresis on 4–20% mini protean TGX precast gels (BioRad). Gels were transferred to nitrocellulose membrane at 30 V overnight in Towbin buffer with 20% methanol; but 1% SDS in the electrophoresis buffer was necessary for efficient transfer of the heterodimer in non-reducing conditions. Western blotting was performed with α6×His mouse mAb (Clontech) and secondary IRDye 800 CW Goat anti-mouse (LI-COR).

Analytical ultracentrifugation (AUC)

Sedimentation velocity AUC was performed on a Beckman XL-I ultracentrifuge with absorbance optics set at 280 nm. CTL4/CTLMA2 was concentrated to 1 mg/ml in 0.2 M NaCl, 20 mM HEPES pH 7.5 and centrifuged at 45,000 rpm. Data analysis was performed with SEDFIT, using values for partial specific volume and buffer viscosity calculated with SEDNTERP.

Dynamic Light Scattering (DLS)

DLS experiments were performed on a DynaPro Plate Reader (Wyatt Technologies) using the Solution and Stability Screen 2 (Hampton Research). CTL4/CTLMA2 in 0.2 M NaCl, 20 mM HEPES pH 7.5 was diluted to 1 mg/ml in a final volume of 20 µl and 4× dilution of the reagents A2–12, H2–H12 resulting in 50 mM buffer ± 1.0 M NaCl.

Isothermal Titration Calorimetry

Binding studies were performed at 25 °C with a MicroCal PEAQ-ITC (Malvern Analytical). Purified proteins were dialyzed against TBS buffer (0.15 M Nacl, 20 mM Tris, pH 8.0) containing 5 mM EDTA to remove bound calcium. Proteins were subsequently diialyzed in Chelex (Biorad) treated TBS buffer (pH 7.5) to remove both EDTA and calcium. Final protein and ligand solutions were made in the same Chelex-treated TBS (pH 7.5) and filtered through 0.2 µm before loading. The cell was loaded with 300 µl of 100 µM protein. The syringe contained 2.5 mM $CaCl_2$ for CTL4 and CTLMA2, 1.25 mM $CaCl_2$ for CTL4/CTLMA2. Titrations comprised 19 injections – 2 µl first injection followed by 18× 4 µl injections – with 150 s between each injection. A TBS buffer control

was titrated against ligand and used as reference. Integrated heats were fit with a single-site binding model – fixed N=1 for CTLMA2, floating N for CTL4/CTLMA2 – to determine dissociation constant K_D and thermodynamic parameters. The results for *An. gambiae* samples arise from three independent experiments. For *An. albimanus* CTL4/CTLMA2 three experiments were performed but under different conditions, with a range of resulting K_D from 0–5 μM and ΔH from 10–27 kcal/mol; The result reported corresponds to equivalent conditions as for *An. gambiae* CTL4/CTLM2 and with median calculated values for K_D, ΔH.

Analysis of glycan binding

Glycan arrays were produced and the glycan library used for screening as previously described [25]. Glycan arrays were probed with 1 μg of 6×His-tagged CTL4, CTLMA2 and CTL4/CTLMA2 in PBS containing 1 mM $CaCl_2$ and 1 mM $MgCl_2$. A mouse anti-His primary antibody was added (1:1 molar ratio with protein) followed by rabbit anti-mouse Alexa647 (1:0.5 ratio with primary), and goat anti-rabbit Alexa647 (1:0.5 ratio with secondary). The final volume applied to the array was 300 μl incubated for 15 min. Arrays were washed 3 times in 1 mM $CaCl_2$ and 1 mM $MgCl_2$. Arrays were scanned on an Innoscan 1100AL using the 632 nm laser. Binding was considered to be positive when values were 3 standard deviations above the background of the array. Experimental details of SPR validation are provided as Supplementary Information (S1 Methods).

Small-Angle X-ray Scattering

SAXS data was collected in-house on a BioSAXS-2000 (Rigaku Corp.) with a DECTRIS PILATUS 100K detector. Samples were prepared as a dilution series from 1–4.5 mg/ml in 0.5 M NaCl, 20 mM HEPES pH 9.0, 0.5 mM $CaCl_2$, 1% glycerol. Data acquisition was 30 min collected as 5 min exposures to ensure no measurable radiation-induced changes within the aquisition period. Following conversion to I vs. q curves, primary data analysis using the *ATSAS* software package [32]. Buffer subtraction and Guinier analysis was performed with *PRIMUS*, $P(r)$

calculation with *GNOM*. Construction of *ab initio* bead models were derived for each structure was performed with *DAMMIF* using D_{max}=80 Å, P1 symmetry.

Mosquito rearing and gene silencing by RNAi

The *An. gambiae* G3 strain was obtained through BEI Resources, NIAID, NIH: *An gambiae*, Strain G3, MRA-112, contributed by Mark Q. Benedict. Mosquitoes were reared on a 12 hr light/dark cycle at 28°C and 75% relative humidity. Mosquitoes were maintained on 10% sucrose and fed sheep blood (HemoStat Laboratories, #SBH100) for egg production. T7 promoter-tagged templates for dsRNA synthesis were generated from a clone in plasmid pIB (*LacZ*), a clone in plasmid pIEx-10 (*CTLMA2*), and from mosquito cDNA (*CTL4*) using the iProof High-Fidelity PCR Kit (Bio-Rad, #1725331) and purified with the GeneJET PCR Purification Kit (Thermo Fisher Scientific, #K0701) according to the manufacturer provided instructions. Double stranded RNA reactions were performed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB, #E2040S) and purified with the GeneJET RNA Purification Kit (Thermo Fisher Scientific, #K0731) according to the manufacturer's instructions. Gene silencing efficiencies were performed as described previously [15] with the following modifications: cDNA was synthesized from 1 μg total RNA using the iScript cDNA Synthesis Kit (Bio-Rad, #1708891) following the manufacturer's instructions and qRT-PCR was performed with a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems) using PerfeCTa SYBR Green SuperMix, Low ROX (Quanta BioSciences, Inc. #95056-500). T7 template and qPCR primer pair sequences are provided as Supplementary Information (S2 Table).

Phenol oxidase (PO) activity assay

PO activity was measured in mosquito cohorts 4 h after the injection of *E. coli* strain DH10B or a mixture of *E. coli* and protein. Mid-log phase *E. coli* was rinsed and resuspended in PBS to OD 0.8 (Average dose: 124,063 CFU/μl). For PO assays involving the co-administration of *E. coli* and protein, bovine serum albumin (BSA, Sigma #B2518) or recombinant *An. gambiae*

CTL4/CTLMA2 was combined with *E. coli* to a final concentration of 2.5 µg/µl just prior to injection. Mosquito injections, hemolymph collection, protein quantification, and the PO activity assay were performed as described previously [16] with the following modifications: (1) For PO assays following the injection of *E. coli* or a mixture of *E. coli* and protein, PO activity was assessed using 4-5 µg of hemolymph protein or the total hemolymph protein obtained from 100 mosquitoes, respectively. (2) Absorbance at 492 nm was recorded every 10 mins for 1 hour in a Molecular Devices SpectraMax 190 plate reader.

Acknowledgements

This work was supported in part by the National Institute of General Medical Sciences R01GM114358 of the National Institutes of Health to RHGB, National Institute for Allergy and Infectious Diseases R01AI139060 of the National Institutes of Health to MP, and National Health and Medical Research Council (NHMRC; Australia) Program Grant 1071659 and Principal Research Fellowship 1138466 to MPJ.

Author Contributions

Conceptualization, R.B., A.C., G.L.S., M.J., M.P. and R.H.G.B.; Methodology, R.B., A.C., C.J.D., D.C.F.H., L.A.P., G.L.S., M.P.J., M.P., A.M.V. and R.H.G.B.; Investigation, R.B., A.C., C.J.D., D.C.F.H., L.A.P., D.P. and G.L.S.; Visualization, R.B., A.C., C.J.D., L.A.P., G.L.S. and R.H.G.B.; Writing – Original Draft, R.H.G.B.; Writing – Review and Editing, C.J.D., G.L.S., M.J.P., A.M.V., M.P. and R.H.G.B.; Funding Acquisition, M.P.J., M.P., A.M.V. and R.H.G.B.; Supervision, M.P.J., M.P., A.M.V. and R.H.G.B.

Conflicts of Interests

The authors declare no conflicts of interests.

References

1. Clayton AM, Dong Y, Dimopoulos G. The *Anopheles* innate immune system in the defense against malaria infection. Journal of Innate Immunity. 2014;6(2):169-81. Epub 2013/08/31. doi: 10.1159/000353602. PubMed PMID: 23988482; PubMed Central PMCID: PMCPMC3939431.

2. Baxter RH, Contet A, Krueger K. Arthropod innate immune systems and vector-borne diseases. Biochemistry. 2017;56(7):907-18. Epub 2017/01/11. doi: 10.1021/acs.biochem.6b00870. PubMed PMID: 28072517; PubMed Central PMCID: PMCPMC5741185.

3. Blandin S, Shiao S-H, Moita LF, Janse CJ, Waters AP, Kafatos FC, et al. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. Cell. 2004;116(5):661-70. doi: 10.1016/s0092-8674(04)00173-4. PubMed PMID: 15006349.

4. Dong Y, Aguilar R, Xi Z, Warr E, Mongin E, Dimopoulos G. *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. PLoS Pathogens. 2006;2(6):e52. doi: 10.1371/journal.ppat.0020052. PubMed PMID: 16789837.

5. Blandin SA, Wang-Sattler R, Lamacchia M, Gagneur J, Lycett G, Ning Y, et al. Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. Science. 2009;326(5949):147-50. Epub 2009/10/03. doi: 10.1126/science.1175241. PubMed PMID: 19797663; PubMed Central PMCID: PMC2959166.

6. White BJ, Lawniczak MKN, Cheng C, Coulibaly MB, Wilson MD, Sagnon NF, et al. Adaptive divergence between incipient species of *Anopheles gambiae* increases resistance to *Plasmodium*. Proceedings of the National Academy of Sciences of the United States of

America. 2011;108(1):244-9. Epub 2010/12/22. doi: 10.1073/pnas.1013648108. PubMed PMID: 21173248; PubMed Central PMCID: PMC3017163.

7. Molina-Cruz A, DeJong RJ, Ortega C, Haile A, Abban E, Rodrigues J, et al. Some strains of *Plasmodium falciparum*, a human malaria parasite, evade the complement-like system of *Anopheles gambiae* mosquitoes. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(28):E1957-62. Epub 2012/05/25. doi: 10.1073/pnas.1121183109. PubMed PMID: 22623529; PubMed Central PMCID: PMC3396512.

8. Waterhouse RM, Povelones M, Christophides GK. Sequence-structure-function relations of the mosquito leucine-rich repeat immune proteins. BMC Genomics. 2010;11:531. Epub 2010/10/06. doi: 10.1186/1471-2164-11-531. PubMed PMID: 20920294; PubMed Central PMCID: PMC3020904.

9. Kanost MR, Jiang H. Clip-domain serine proteases as immune factors in insect hemolymph. Current Opinion in Insect Science. 2015;11:47-55. doi: 10.1016/j.cois.2015.09.003. PubMed PMID: 26688791; PubMed Central PMCID: PMCPMC4680995.

10. Fraiture M, Baxter RHG, Steinert S, Chelliah Y, Frolet C, Quispe-Tintaya W, et al. Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of *Plasmodium*. Cell Host and Microbe. 2009;5(3):273-84. doi: 10.1016/j.chom.2009.01.005. PubMed PMID: 19286136.

11. Povelones M, Waterhouse RM, Kafatos FC, Christophides GK. Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. Science. 2009;324(5924):258-61. doi: 10.1126/science.1171400. PubMed PMID: 19264986.

12. Baxter RHG, Steinert S, Chelliah Y, Volohonsky G, Levashina EA, Deisenhofer J. A heterodimeric complex of the LRR proteins LRIM1 and APL1C regulates complement-like immunity in *Anopheles gambiae*. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(39):16817-22. Epub 2010/09/10. doi: 10.1073/pnas.1010575107. PubMed PMID: 20826443; PubMed Central PMCID: PMCPMC2947905.

13. Osta MA, Christophides GK, Kafatos FC. Effects of mosquito genes on *Plasmodium* development. Science. 2004;303(5666):2030-2. doi: 10.1126/science.1091789. PubMed PMID: 15044804.

14. Shiao S-H, Whitten MMA, Zachary D, Hoffmann JA, Levashina EA. Fz2 and cdc42 mediate melanization and actin polymerization but are dispensable for *Plasmodium* killing in the mosquito midgut. PLoS Pathogens. 2006;2(12):e133. doi: 10.1371/journal.ppat.0020133. PubMed PMID: 17196037; PubMed Central PMCID: PMCPMC1757202.

15. Volz J, Müller H-M, Zdanowicz A, Kafatos FC, Osta MA. A genetic module regulates the melanization response of *Anopheles* to *Plasmodium*. Cellular Microbiology. 2006;8(9):1392-405. Epub 2006/08/23. doi: 10.1111/j.1462-5822.2006.00718.x. PubMed PMID: 16922859.

16. Schnitger AKD, Yassine H, Kafatos FC, Osta MA. Two C-type lectins cooperate to defend *Anopheles gambiae* against Gram-negative bacteria. Journal of Biological Chemistry. 2009;284(26):17616-24. doi: 10.1074/jbc.M808298200. PubMed PMID: 19380589; PubMed Central PMCID: PMCPMC2719400.

17. Cohuet A, Osta MA, Morlais I, Awono-Ambene PH, Michel K, Simard F, et al. *Anopheles* and *Plasmodium*: from laboratory models to natural systems in the field. EMBO Reports. 2006;7(12):1285-9. PubMed PMID: 17099691.

18. Simões ML, Mlambo G, Tripathi A, Dong Y, Dimopoulos G. Immune regulation of *Plasmodium* is *Anopheles* species specific and infection intensity dependent. mBio. 2017;8(5). Epub 2017/10/19. doi: 10.1128/mBio.01631-17. PubMed PMID: 29042500; PubMed Central PMCID: PMCPMC5646253.

19. Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, Winikor J, et al. The human malaria parasite *Pfs47* gene mediates evasion of the mosquito immune system. Science. 2013;340(6135):984-7. Epub 2013/05/11. doi: 10.1126/science.1235264. PubMed PMID: 23661646.

20. Giraldo-Calderon G, Emrich S, MacCallum R, Maslen G, Dialynas E, Topalis P, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Research. 2015;43(D1):D707-D13. doi: 10.1093/nar/gku1117. PubMed PMID: WOS:000350210400103.

21. Mayer S, Raulf M-K, Lepenies B. C-type lectins: their network and roles in pathogen recognition and immunity. Histochemistry and Cell Biology. 2017;147(2):223-37. Epub 2016/12/22. doi: 10.1007/s00418-016-1523-7. PubMed PMID: 27999992.

22. Pees B, Yang W, Zárate-Potes A, Schulenburg H, Dierking K. High innate immune specificity through diversified C-type lectin-like domain proteins in invertebrates. Journal of Innate Immunity. 2016;8(2):129-42. Epub 2015/11/19. doi: 10.1159/000441475. PubMed PMID: 26580547.

23. Feinberg H, Taylor ME, Weis WI. Scavenger receptor C-type lectin binds to the leukocyte cell surface glycan Lewis(x) by a novel mechanism. J Biol Chem. 2007;282(23):17250-8. Epub 2007/04/09. doi: 10.1074/jbc.M701624200. PubMed PMID: 17420244; PubMed Central PMCID: PMCPMC2289868.

24. Mizuno H, Fujimoto Z, Atoda H, Morita T. Crystal structure of an anticoagulant protein in complex with the Gla domain of factor X. Proc Natl Acad Sci U S A. 2001;98(13):7230-4. Epub 2001/06/12. doi: 10.1073/pnas.131179698. PubMed PMID: 11404471; PubMed Central PMCID: PMCPMC34651.

25. Waespy M, Gbem TT, Elenschneider L, Jeck AP, Day CJ, Hartley-Tassell L, et al. Carbohydrate Recognition Specificity of Trans-sialidase Lectin Domain from *Trypanosoma congolense*. PLoS Negl Trop Dis. 2015;9(10):e0004120. Epub 2015/10/16. doi: 10.1371/journal.pntd.0004120. PubMed PMID: 26474304; PubMed Central PMCID: PMCPMC4608562.

26. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics. 2016;54:5.6.1-5.6.37. Epub 2016/06/20. doi: 10.1002/cpbi.3. PubMed PMID: 27322406; PubMed Central PMCID: PMCPMC5031415.

27. Piiadov V, Ares de Araújo E, Oliveira Neto M, Craievich AF, Polikarpov I. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. Protein Sci. 2019;28(2):454-63. Epub 2018/12/13. doi: 10.1002/pro.3528. PubMed PMID: 30371978; PubMed Central PMCID: PMCPMC6319763.

28. Franke DA, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography. 2009;42:342-6.

29. Schneidman-Duhovny D, Hammel M, Tainer J, Sali A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Research. 2016;44(W1):W424-W9. doi: 10.1093/nar/gkw389. PubMed PMID: WOS:000379786800070.

30. Coombs PJ, Graham SA, Drickamer K, Taylor ME. Selective binding of the scavenger receptor C-type lectin to Lewis^{x} trisaccharide and related glycan ligands. J Biol Chem. 2005;280(24):22993-9. Epub 2005/04/21. doi: 10.1074/jbc.M504197200. PubMed PMID: 15845541.

31. Salt GW. The cellular defence reactions of insects. Cambridge [England]: University Press; 1970.

32. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Crystallogr. 2017;50(Pt 4):1212-25. Epub 2017/06/26. doi: 10.1107/S1600576717007786. PubMed PMID: 28808438; PubMed Central PMCID: PMCPMC5541357.

Table 1: Calcium binding of CTL4 and CTLMA2 by ITC

	An. gambiae			*An. albimanus*
	CTL4	CTLMA2	CTL4/CTLMA2	CTL4/CTLMA2
N (sites)	1.00	1 (fixed)	0.36 (0.07)	0.497 (0.003)
K_D (μM)	N/A	173 (27)	4.9 (0.5)	2.82 (0.01)
Δ*H* (kcal/mol)	N/A	-12 (2)	-23 (4)	-12.1 (0.1)
Δ*G* (kcal/mol)	N/A	-5.1 (0.1)	-7.2 (0.1)	-7.57
-*T*Δ*S* (kcal/mol)	N/A	7 (2)	16 (4)	4.57

Table 2:Glycan array results for CTL4 and CTLMA2*

ID#	Glycans	CTL4	CTLMA2	CTL4/ CTLMA2
6	GalNAcβ-sp3			1.14
47	6-H2PO3Manα-sp3			1.32
113	GlcNAcβ1-3GalNAcα-sp3			1.16
163	6-O-Su-Galβ1-4GlcNAcβ-sp3	1.96		
199	4,6-O-Su2-GalNAcβ1-4-(3-O-Ac)GlcNAcβ-sp3		1.42	
251	GlcNAcβ1-4Galβ1-4GlcNAcβ-sp2		1.04	
287	3-O-Su-Galβ1-3(Fucα1-4)GlcNAcβ-sp3		1.12	1.14
368	Fucα1-2(GalNAcα1-3)Galβ1-4GlcNAcβ-sp3			1.42
383	Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ-sp2			1.47
421	Neu5Acα2-3(GalNAcβ1-4)Galβ1-4Glcβ-sp2			1.94
428	Fucα1-3(Neu5Acα2-3Galβ1-4)6-O-Su-GlcNAcβ-sp3		1.01	1.38
479	Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-sp4			1.92
488	Galβ1-4GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)GalNAcα-sp3		1	1.53
505	(GN-M)2-3,6-M-GN-GNβ-sp4		1.18	1.74
529	Neu5Acα2-6(Galβ1-3)GlcNAcβ1-3Galβ1-4Glcβ-sp4		1.09	1.49
533	(Neu5Acα2-8)2Neu5Acα2-3(GalNAcβ1-4) Galβ1-4Glc-sp2			1.58
536	Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-sp4			2.13
625	(GlcAβ1-4GlcNAcβ1-3)8-NH2-ol			2.14
18B	Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glc	3.65		
13K	(GlcA/IdoAβ1-3(±4/6S)GalNAcβ1-4)n (n<250)			2.16
13M	(GlcA/IdoAβ1-3(±6S) GalNAcβ1-4)n (n<250)			2.94
14I	HA 1600000 da 2.5 mg/ml	1.48		1.52
14K	β1-3Glucan	2.32		

*Values indicate fold above background plus 3 standard errors of the background.

Figure Legends

Figure 1: Unrooted phylogenetic trees for CTL4 (blue), CTLMA2 (red) in ten *Anopheles* species including *An. gambiae* (purple) and *An. albimanus* (cyan). The N-terminal portion of both multi-sequence alignments is shown with conserved cysteine residues highlighted yellow, other conserved residues highlighted in gray. The CXCXC motif is conserved in all sequences except those truncated at the N-terminus.

Figure 2: (A) Size-exclusion chromatogram (Superdex75 16/60) of CTL4/CTLMA2 heterodimer with non-reducing (NR) and reducing (R) SDS-PAGE of peak fractions. (B) α6×His Western blotting for nine cysteine mutants of the CTL4/CTLMA2 heterodimer. The 6×His tag is located on the C-terminus of CTL4. In all samples formation of a heterodimer is evident. (C) Plot of $c(s)$ vs. s for sedimentation velocity analytical ultracentrifugation of 1 mg/ml CTL4/CTLMA2, 200 mM NaCl, 20 mM HEPES pH 7.5. Three peaks of increasing sedimentation coefficient are observed, s_1 = 3.1 s, s_2 = 4.4 s, s_3 = 5.9 s. (D) Dynamic Light Scattering of CTL4/CTLMA2 vs. pH (50 mM buffer). An increasing trend is apparent in apparent MW (■), radius in nm (●) and polydispersity of the sample (▲), as the pH decreases from 9.5 to 6.0.

Figure 3. ITC binding isotherm for calcium binding to (A) *An. gambiae* CTL4, (B) *An. gambiae* CTLMA2, (C) *An. gambiae* CTL4/CTLMA2, and (D) *An. albimanus* CTL4/CTLMA2. Protein concentration in cell was 100 μM, Ca^{2+} ($CaCl_2$) concentration in syringe was 2.5 mM for monomers, 1.25 mM *An. gambiae* CTL4/CTLMA2, 0.8 mM for *An. albimanus* CTL4/CTLMA2. No binding is evident for CTL4, weak binding for CTLMA2 and tight binding for CTL4/CTLMA2.

Figure 4. (A) Structure-based sequence alignment (NCBI VAST) for CTL4 and CTLMA2 homology models (*MODELLER*) with mouse scavenger receptor CTLD (2OX9) and Surfactant Neck protein D (4DN8). Identically conserved residues are shown in red. Residues conserved in 75% of structures are in grey. Residues of the Ca^{2+}/glycan binding loop are highlighted in pink. (B) Molecular model for CTL4 (green) and CTLMA2 (orange). CTL4 basic loop 1 residues are

highlighted in blue, CTLMA2 acidic loop 1 residues in red. Ca^{2+}/glycan binding loop highlighted in pink. Cysteines and predicted disulfide bonds shown as sticks. The heterodimer model for CTL4/CTLMA2 heterodimer. Note the favorable orientation of CTL4 and CTLMA2 complementary charged loop1. (C) Experimental scattering curve with inset Guinier plot (*PRIMUS*), R_G=24.5 Å. (D) $P(r)$ distribution (*GNOM*) with fit to experimental data, R_G=25.4 Å, D_{max}=80 Å. (E) Superposition of five *ab initio* bead models (*DAMMIF*) with first of a three-state model from *MULTIFOXS*.

Figure 5. (A) Experimental design for measuring hemolymph PO activity 4 h after *E. coli* challenge (OD 0.8). PO activity was determined by combining hemolymph with the substrate L-3,4-dihydroxyphenylalanine (L-DOPA) as described in the Materials & Methods. (B) Enhanced *E. coli*-induced hemolymph PO activity in CTL4/CTLMA2 knockdowns (n = 3–5, Tukey's multiple comparisons test). * = 0.017, ** = 0.0035. (C) Co-administration of recombinant CTL4/CTLMA2 reverses the enhancement of *E. coli*-induced hemolymph PO activity in ds*CTL4* mosquitoes (n = 3, Tukey's multiple comparisons test). Bars represent mean ± SD with $p \leq 0.05$ considered significant.

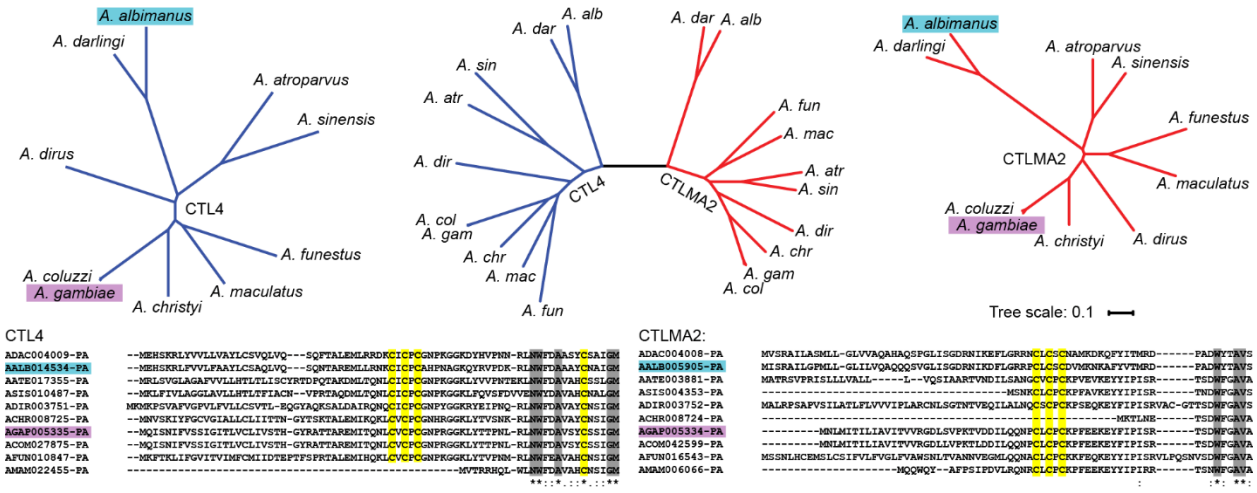
A. albimanus
A. darlingi
A. atroparvus
A. sinensis
A. dirus
CTL4
A. funestus
A. coluzzi
A. gambiae
A. maculatus
A. christyi
A. alb
A. dar
A. sin
A. atr
A. dir
A. col
A. gam
A. chr
A. mac
A. fun
CTL4
CTLMA2
A. dar
A. alb
A. fun
A. mac
A. atr
A. sin
A. dir
A. chr
A. gam
A. col
A. albimanus
A. darlingi
A. atroparvus
A. sinensis
A. funestus
CTLMA2
A. maculatus
A. coluzzi
A. gambiae
A. christyi
A. dirus
Tree scale: 0.1
CTL4
ADAC004009-PA --MEHSKRLYVVLLVAYLCSVQLVQ---SQFTALEMLRRDKCICPCGNFKGGKDYHVFNN-RLNWFDAASYCSAIGM
AALB014534-PA --MEHSKRLFVVLFAAYLCSAQLVQ---SQNTAREMLLRNKCICPCAHFNAGKQYRVPDK-RLNWFDAAAYCNAIGM
AATE017355-PA --MRLSVGLAGAFVVLLHTLTLISCYRTDPQTAKDMLTQNLCICPCGNFKGGKLYVVFNTEKLNWFDAVAHCSSLGM
ASIS010487-PA --MKLFIVLAGGLAVLLHTLTFIACN--VFRTAQDMLTQNLCICPCGNFKGGKLFQVSFDVVENWYDAVAHCNALGM
ADIR003751-PA MKMKPSVAFVGFVLFVLLCSVTL-EQGYAQKSALDAIRQNQCICPCGNFYGGKRYEIPNQ-RLNWFDAVAYCNSIGM
ACHR008725-PA --MNVSKIYFGCVGIALLCLIITTN-GYTSKTALEMIKQKLCVCPCGNHRGGKLYTVSNK-RLNWFDAVAYCNSIGM
AGAF005335-PA --MQISNIFVSSIGITLVCLIVSTH-GYRATTAREMITQNLCVCPCGNFRGGKLYTTFNL-RLNWFDAVSYCSSIGM
ACOM027875-PA --MQISNIFVSSIGITLVCLIVSTH-GYRATTAREMITQNLCVCPCGNFRGGKLYTTFNL-RLNWFDAVSYCSSIGM
AFUN010847-PA --MKFTKLIFGVITVIMFCMIIDTEPTFSPRTALEMIHQKLCVCPCGNFKGGKLYTVFNM-RLNWFEAVAHCNSIGM
AMAM022455-PA ----------MVTRRHQL-WLNWFDAVAHCNSIGM
::*.::*.::
CTLMA2:
ADAC004008-PA MVSRAILASMLL-GLVVAQAHAQSPGLISGDRNIKEFLGRRNCLCSCNAMKDKQFYITMRD------PADWYTAVS
AALB005905-PA MISRAILGPMLL-GLILVQAQQQSVGLISGDRNIKDFLGRRPCLCSCDVMKNKAFYVTMRD------PADWYTAVS
AATE003881-PA MATRSVPRISLLLVALL-----L--VQSIAARTVNDILSANGCVCPCKPVEVKEYYIPISR------TSDWFGAVA
ASIS004353-PA ----------MSNKCLCPCKPFAVKEYYIPISR------TNDWFGAVS
ADIR003752-PA MALRPSAPVSILATLFLVVVIPLARCNLSGTNTVEQILALNQCSCPCKPSEQKEYFIPISRVAC-GTTSDWFGAVS
ACHR008724-PA ----------MKTLNE------TSDWFGAVA
AGAF005334-PA --------MNLMITILIAVITVVRGDLSVPKTVDDILQQNPCLCPCKPFEEKEYFIPISR------TSDWFGAVA
ACOM042599-PA --------MNLMITILIAVITVVRGDLLVPKTLDDILQQNPCLCPCKPFEEKEYFIPISR------TSDWFGAVA
AFUN016543-PA MSSNLHCEMSLCSIFVLFVGLFVAWSNLTVANNVBGMLQQNACLCPCKKFEQKEYYIPISRVLPQSNVSDWFGAVA
AMAM006066-PA ----------MQQWQY--AFPSIPDVLRQNRCLCPCKPFEEKEYYIPIRR------TSNWFGAVA
: :*: **:

Figure 1

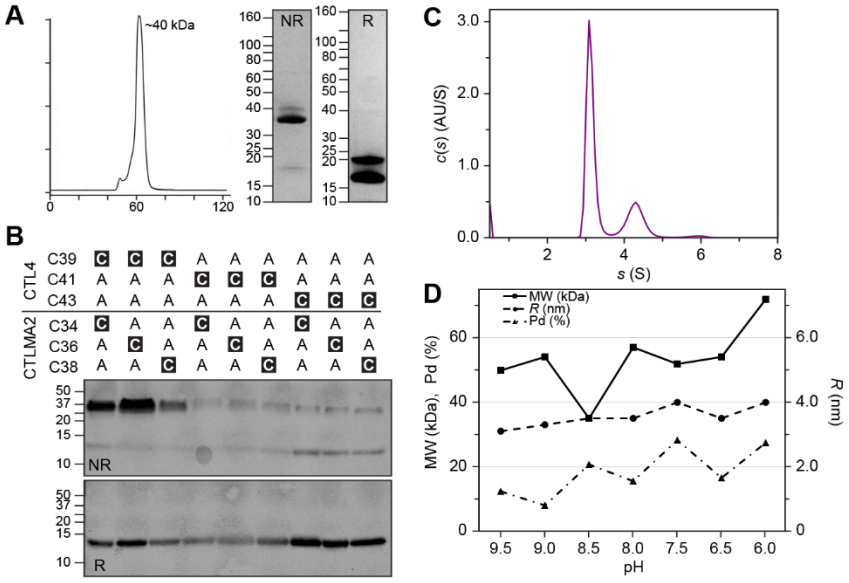
A
~40 kDa
NR
R
B
CTL4
C39
C41
C43
CTLMA2
C34
C36
C38
NR
R
C
c(s) (AU/S)
s (S)
D
MW (kDa)
R (nm)
Pd (%)
MW (kDa), Pd (%)
R (nm)
pH

Figure 2

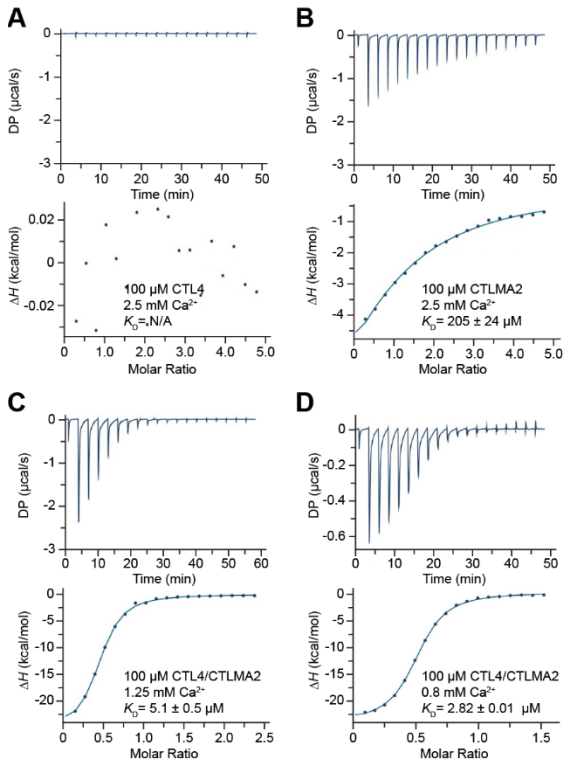
A
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
100 µM CTL4
2.5 mM Ca2+
KD=.N/A
Molar Ratio
B
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
100 µM CTLMA2
2.5 mM Ca2+
KD= 205 ± 24 µM
Molar Ratio
C
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
100 µM CTL4/CTLMA2
1.25 mM Ca2+
KD= 5.1 ± 0.5 µM
Molar Ratio
D
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
100 µM CTL4/CTLMA2
0.8 mM Ca2+
KD= 2.82 ± 0.01 µM
Molar Ratio

Figure 3

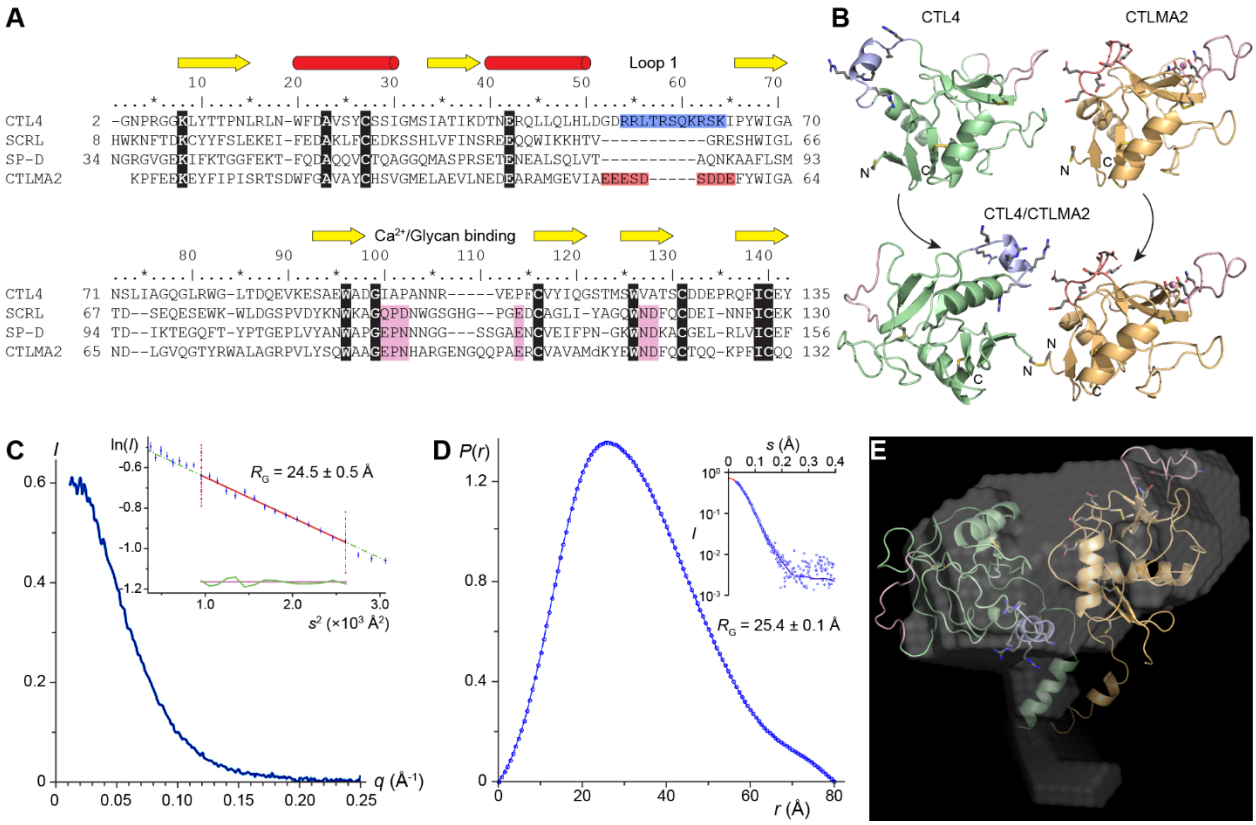
A
Loop 1
10 20 30 40 50 60 70
CTL4 2 -GNPRGGKLYTTPNLRLN-WFDAVSYCSSIGMSIATIKDTNERQLLQLHLDGDRRLTRSQKRSKIPYWIGA 70
SCRL 8 HWKNFTDKCYYFSLEKEI-FEDAKLFCEDKSSHLVFINSREEQQWIKKHTV----------GRESHWIGL 66
SP-D 34 NGRGVGEKIFKTGGFEKT-FQDAQQVCTQAGGQMASPRSETENEALSQLVT----------AQNKAAFLSM 93
CTLMA2 KPFEEKEYFIPISRTSDWFGAVAYCHSVGMELAEVLNEDEARAMGEVIAEEESD-----SDDEFYWIGA 64
Ca2+/Glycan binding
80 90 100 110 120 130 140
CTL4 71 NSLIAGQGLRWG-LTDQEVKESAEWADGIAPANNR-----VEPHCVYIQGSTMSWVATSCDDEPRQFICEY 135
SCRL 67 TD--SEQESEWK-WLDGSPVDYKNWKAGQPDNWGSGHG--PGEDCAGLI-YAGQWNDFQCDEI-NNFICEK 130
SP-D 94 TD--IKTEGQFT-YPTGEPLVYANWAPGEPNNNGG---SSGAENCVEIFPN-GKWNDKACGEL-RLVICEF 156
CTLMA2 65 ND--LGVQGTYRWALAGRPVLYSQWAAGEPNHARGENGQQPAEPCVAVAMdKYPWNDFQCTQQ-KPFICQQ 132
B
CTL4
CTLMA2
N
C
CTL4/CTLMA2
C
D
P(r)
1.2
0.8
0.4
0
0 20 40 60 80
r (Å)
s (Å)
0 0.1 0.2 0.3 0.4
I
RG = 25.4 ± 0.1 Å
E

Figure 4

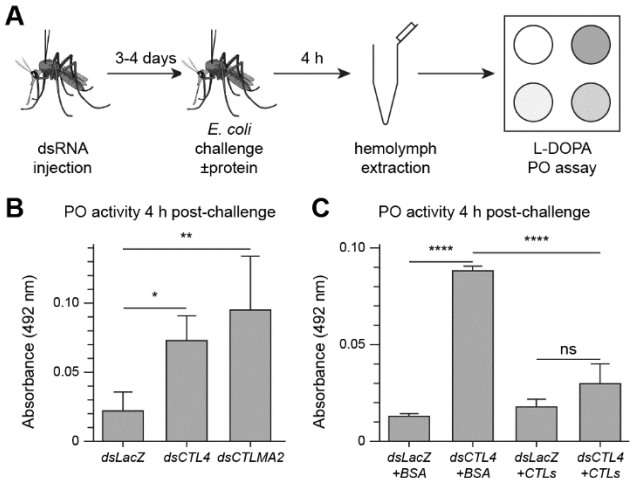
A
3-4 days
4 h
dsRNA
injection
E. coli
challenge
±protein
hemolymph
extraction
L-DOPA
PO assay
B
PO activity 4 h post-challenge
Absorbance (492 nm)
0.10
0.05
0
**
*
dsLacZ
dsCTL4
dsCTLMA2
C
PO activity 4 h post-challenge
Absorbance (492 nm)
0.10
0.05
0
****
****
ns
dsLacZ +BSA
dsCTL4 +BSA
dsLacZ +CTLs
dsCTL4 +CTLs

Figure 5

S1 Methods. Surface plasmon resonance (SPR) of CTL4/CTLMA2 with glycans.

Methods

SPR was performed on a Biacore T200 using a CM5 senor chip. Flow cell one was left blank with ethanolamine only blocking the NHS activated carboxydextran. The proteins were immobilised onto flow cell 2-4 (CTL4, CTLMA2, CTL4/MA2 respectively). Glycans were flowed from 160 nM to 100 μM across a 1:5 dilution. Analysis was performed using the Biacore T200 evaluation software using the surface bound menu; Affinity; Steady state affinity. Glycan were run in triplicate with glycans run in single replicates with the program repeated three times.

Results

To confirm the glycan array results and to determine binding preferences for the CTLs SPR analysis was performed. In almost all interactions the glycan array and SPR were in agreement for the presence of interactions with the SPR indicating the glycan array had four false negative results: CTLMA2 monomer with H-antigen, chondroitin sulfate and chondroitin-6-sulfate, and CTL4 monomer with chondoitin-6-sulfate. However, all of the false negative results showed binding on the array with the complexed CTL4/CTLMA2 (Table 2 and S1 Table).

The highest affinity binding was observed to sulfo-Lewis A with binding to the CTLMA2 (106 nM) and CTL4/CTLMA2 (95 nM) complex at a K_D of ~100 nM (S2 Table). The highest affinity binding observed for CTL4 was also to a sulfated glycan, sulfo-lactosamine (292 nM).

S1 Table: Surface plasmon resonance (SPR) of CTL4/CTLMA2 with glycans

Glycan	Glycan array ID	CTL4 (μM)	CTLMA2 (μM)	CTL4/CTLMA2 (μM)
A blood group trisaccharide	386	49.27±12.2	1.77±0.19	5.56±2.9
B blood group type 5	N/A	NCDI	NCDI	NCDI
H disaccharide	479	NCDI	10.39±3.26	26.56±6.81
asialo-GM1	18B	15.4±3.99	NCDI	NCDI
LNT	536	NCDI	NCDI	8.22±1.71
LNnT	383,488	NCDI	NCDI	17.62±2.99
GlcNAc monosaccharide	113,251	27.87±11.4	29.69±7.91	21.91±2.8
Chondroitin sulfate	13K	NCDI	0.417±0.1	0.404±0.19
Chondroitin-6-sulfate	13M	33.54±14.2	8.47±2.42	5.71±1.2
Sialyl-Lewis A	N/A	NCDI	NCDI	NCDI
Sulfo-Lewis A	287	NCDI	0.106±0.06	0.095±0.02
Sialyl-Lewis X	428	NCDI	0.619±0.1	3.81±1.0
Sulfo-lactosamine	163	0.292±0.15	NCDI	NCDI

S2 Table: dsRNA and qPCR primers used

Name	Sequence
T7 Primers	
LacZ-T7 F	taatacgactcactatagggAGAATCCGACGGGTTGTTACT
LacZ-T7 R	taatacgactcactatagggCACCACGCTCATCGATAATTT
AgCTL4-T7 F	taatacgactcactatagggTGGTTTGATGCCGTGTCCT
AgCTL4-T7 R	taatacgactcactatagggAATAAATTGTCTCGGTTCATCATC
AgCTLMA2-T7 F	taatacgactcactatagggGCCTTTGCCCGTGCAAACCGTTC
AgCTLMA2-T7 R	taatacgactcactatagggTTGACAGATGAACGGTTTCTGCTG
qPCR primers	
AgS7-qPCR F	GTGCGCGAGTTGGAGAAGA
AgS7-qPCR R	ATCGGTTTGGGCAGAATGC
AgCTL4-qPCR F	GCACGGGTACAGGGCTACTA
AgCTL4-qPCR R	GCGTGGTGTACAGCTTTCCT
AgCTLMA2-qPCR F	GCTGTCATCACAGTGGTTCG
AgCTLMA2-qPCR R	GGGTTTTGTTGAAGAATATCATCC