

Genome-Wide Identification of Molecular Mimicry Candidates in Parasites

Philipp Ludin^{1,2,3}, Daniel Nilsson^{1*}, Pascal Mäser^{1,2,3*}

1 Institute of Cell Biology, University of Bern, Bern, Switzerland, **2** Swiss Tropical and Public Health Institute, Basel, Switzerland, **3** University of Basel, Basel, Switzerland

Abstract

Among the many strategies employed by parasites for immune evasion and host manipulation, one of the most fascinating is molecular mimicry. With genome sequences available for host and parasite, mimicry of linear amino acid epitopes can be investigated by comparative genomics. Here we developed an *in silico* pipeline for genome-wide identification of molecular mimicry candidate proteins or epitopes. The predicted proteome of a given parasite was broken down into overlapping fragments, each of which was screened for close hits in the human proteome. Control searches were carried out against unrelated, free-living eukaryotes to eliminate the generally conserved proteins, and with randomized versions of the parasite proteins to get an estimate of statistical significance. This simple but computation-intensive approach yielded interesting candidates from human-pathogenic parasites. From *Plasmodium falciparum*, it returned a 14 amino acid motif in several of the PfEMP1 variants identical to part of the heparin-binding domain in the immunosuppressive serum protein vitronectin. And in *Brugia malayi*, fragments were detected that matched to periphilin-1, a protein of cell-cell junctions involved in barrier formation. All the results are publicly available by means of mimicDB, a searchable online database for molecular mimicry candidates from pathogens. To our knowledge, this is the first genome-wide survey for molecular mimicry proteins in parasites. The strategy can be adopted to any pair of host and pathogen, once appropriate negative control organisms are chosen. MimicDB provides a host of new starting points to gain insights into the molecular nature of host-pathogen interactions.

Citation: Ludin P, Nilsson D, Mäser P (2011) Genome-Wide Identification of Molecular Mimicry Candidates in Parasites. PLoS ONE 6(3): e17546. doi:10.1371/journal.pone.0017546

Editor: Najib El-Sayed, The University of Maryland, United States of America

Received: October 8, 2010; **Accepted:** February 8, 2011; **Published:** March 8, 2011

Copyright: © 2011 Ludin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Daniel Nilsson was supported by a fellowship of the Swedish National Science Foundation, and Pascal Mäser was supported by a research professorship grant of the Swiss National Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pascal.maeser@unibas.ch

‡ Current address: Science for Life Laboratory, Karolinska Institutet, Solna, Sweden

Introduction

Endoparasites are confronted with host defenses at multiple levels: physical barriers, innate immunity, and adaptive immune responses need to be overcome in order to successfully establish an infection and proliferate inside a host. Antigenic variation to escape humoral responses is well documented for the malaria parasites, *Giardia*, African trypanosomes, etc. Further strategies for immune evasion or immune suppression are less well understood. Molecular mimicry as a strategy for immune evasion and host manipulation is well known from viruses [1,2]. While many viruses have a natural propensity to acquire genetic material or proteins from the host cell upon formation of virions, others have by themselves evolved surface proteins for mimicry, e.g. the chemokine receptors of cytomegalovirus [3]. The term molecular mimicry was coined by R. Damian in 1964 and defined as the sharing of antigens between parasite and host [4]. We refer here to molecular mimicry as the display of any structure by the parasite that (i) resembles structures of the host at the molecular level and (ii) confers a benefit to the parasite because of this resemblance. The potential benefits of molecular mimicry include camouflage – as exemplified by the concept of ‘eclipsed antigens’ which are not recognized as such by the host’s immune system due to their similarity to host antigens [5] – and cytoadherence. For intracellular parasites, cytoadherence is a

prerequisite to infection. Trypomastigote *T. cruzi* adhere to fibroblasts via the fibronectin receptor, and exogenous peptides with fibronectin RGD motifs inhibited host cell invasion [6,7]. Cytoadherence of *P. falciparum*-infected erythrocytes to microvascular endothelium contributes to cerebral malaria pathology. *P. falciparum* erythrocyte membrane protein 1 (PfEMP1, encoded by the *var* genes) interacts with adhesion molecules such as ICAM-1, CD36, or thrombospondin via different domains [8,9]. Endothelial adherence prevents the infected erythrocytes from passage to the spleen where they would be eliminated. A third reason why parasites might mimic host molecules is signaling. Parasites may mimic hormone receptors to respond to signals from the host, or mimic hormones to send signals to the host. Functional homologues of the mammalian epidermal growth factor (EGF) receptor were described from trypanosomes [10,11] and helminths [12,13]. *Plasmodium* spp. possess at least two surface proteins with EGF motifs, one (Pf25) expressed in the mosquito [14], the other (MSP1) in the blood-stages where it is critical for erythrocyte invasion [15,16]. Schistosomes send immunosuppressive signals in the form of neuropeptides to both the definite host (man) and the intermediate host (snail) [17]. There are extreme cases of behavioral manipulation of the host by the parasite such as the suicidal diving of grasshoppers infected by hairworms, and there too molecular mimicry is likely to play a role [18].

The first evidence for molecular mimicry between parasite and host came from immunological studies on antisera that cross-reacted with parasite and host. *Ascaris lumbricoides* was found to possess A- and B-like blood group antigens [19]. This was confirmed by more recent studies, which suggested that these antigens had been acquired from host blood [20]. Biosynthesis of human blood group-like antigens was described for *Schistosoma mansoni* [21,22] and *Fasciola hepatica* [23]. However, the function of these antigens produced by the parasite remains to be elucidated. More recently, tools other than antisera were used to address molecular mimicry between parasite and host. Molecular cloning of the involved genes [24,25], elucidation of polysaccharide structures [26], use of monoclonal antibodies [27,28] and synthetic peptides [29] have all contributed to a wealth of evidence that endoparasites take advantage of molecular mimicry to survive in their hosts (see also Table 1). Recurring targets for mimicry by bloodborne pathogens are the components of the complement system, growth hormones and their receptors, and cell adhesion molecules [30]. A parasite's ability to perform molecular mimicry may stem from either having acquired macromolecules from the host (transfer) or from adaptive evolution of the mimicking structures (convergence). Both scenarios are supported by multiple examples from parasites (Table 1). With the rapidly growing number of fully sequenced genomes, direct comparison between host and parasite protein sequences provides a powerful tool to identify molecular mimicry candidates. To our knowledge, however, there has been no systematic approach to study molecular mimicry since parasitology entered the post-genomic era.

Here we develop an *in silico* pipeline to identify molecular mimicry candidates from parasites. In brief, proteome-wide blast surveys were performed with either whole proteins or with overlapping protein fragments to identify similar epitopes in parasite and host. This approach warrants that all linear amino acid epitopes which share significant similarity between parasite and host will be discovered. Searches against control proteomes of free-living eukaryotes served as negative controls to exclude proteins that are generally conserved across phyla, while searches with random sequences allowed to estimate statistical significance. The results are made available by means of an online database for molecular mimicry candidate proteins in pathogens.

Results and Discussion

Molecular mimicry surveys with full length protein sequences

In pilot surveys for molecular mimicry candidates we concentrated on endoparasitic helminths since (i) they are known masters

of immune evasion and host manipulation, and (ii) a convenient negative control is available in the form of the free-living nematode *C. elegans*. In principal, a mimicry candidate is a parasite protein or motif which bears a high degree of resemblance to a protein of the host but not to those of unrelated control species. Such proteins are readily identified by proteome-wide blast surveys. In a first trial, we ran every predicted protein of *Brugia malayi* with blastp against the proteomes of *H. sapiens* and *C. elegans*. As expected, the *B. malayi* proteins returned significantly ($p < 0.0001$, two-tailed Wilcoxon test) higher scores against *C. elegans* than against *H. sapiens*. There were only few *B. malayi* proteins which scored better against the human host (Figure 1, left). The converse picture emerged when the same procedure was carried out with *Schistosoma mansoni* (Figure 1, right) or *S. japonicum* (not shown), where the parasite proteins generally were more similar to human than to *C. elegans* proteins ($p < 0.0001$, two-tailed Wilcoxon test). The systemic nature of the phenomenon (Figure 1, right) speaks against molecular mimicry as the underlying selective force since it involves too many housekeeping proteins that do not interact with the host. *C. elegans* and *S. mansoni* are from different metazoan clades, the ecdysozoa and the lophotrochozoa, respectively [31]. While the *S. mansoni* proteins were also more similar to *D. melanogaster* than to *C. elegans* proteins, the overall similarity to human proteins was still the most pronounced (not shown).

The two-dimensional blastp approach allowed to graphically divide the proteome of *B. malayi* into separate quadrants: parasite-specific proteins (lower left in Figure 1, left), generally conserved proteins such as tubulin or ubiquitin (upper right), nematode-specific proteins (upper left), and mimicry candidates (lower right). However, this rough subdivision is prone to false positives caused by the well documented phenomenon of gene loss in *C. elegans* [32]. In order to eliminate proteins which are generally conserved, the negative control was refined to include – in addition to *C. elegans* – a panel of unrelated, free-living eukaryotes whose genomes have been sequenced: *Saccharomyces pombe*, *Arabidopsis thaliana*, *Ciona intestinalis*, and *Trichoplax adhaerens* (Table 2). For the detection of mimicry candidates we focused on human-pathogenic endoparasites known for their mastery in immune evasion, namely *Brugia malayi*, *Schistosoma mansoni*, *Plasmodium falciparum*, *Leishmania major*, *Cryptosporidium parvum*, *Trichomonas vaginalis* and *Trypanosoma cruzi* (Table 2). The predicted proteomes of the parasites were run as blast queries against the control proteomes and against *H. sapiens*. Molecular mimicry candidates were defined as parasite proteins with (i) a blastp score above 100 to the best hit in the human proteome and (ii) a score in *H. sapiens* at least two-fold higher than the best score achieved in the control proteomes. This

Table 1. Possible mechanism for molecular mimicry and examples from pathogens.

Macromolecule	Mimicry by transfer	Mimicry by convergence
Nucleic acid	<i>Schistosoma mansoni</i> possesses a CRIT gene which shares 98% identical nucleotides with the human orthologue [25].	The 3'UTR of the RNA genome of barley yellow dwarf virus mimics the m ⁷ G cap of eukaryotic mRNA to stimulate translation [59].
Protein	Pathogenic bacteria, <i>E. granulosus</i> and <i>O. volvulus</i> decorate themselves with inhibitors of the complement cascade sequestered from the blood [43,60,61,62].	A 18 aa motif in <i>P. falciparum</i> CSP is nearly identical to the cytoadhesive region of mammalian thrombospondin [49] and was shown to bind to hepatocytes [63].
Sugar	Trans-sialidases transfer sialic acid from host cells to the surface of the parasite. <i>T. cruzi</i> trans-sialidase is a virulence factor in mammals [64]; <i>T. brucei</i> trans-sialidase is required for survival in the tsetse fly [65].	Several pathogenic helminths synthesize the Forssman antigen (globopentosylceramide) [21,66], a glycolipid implicated in cell adhesion and the formation of tight junctions [67].

Mimicry by transfer of nucleic acids or convergence of proteins can be identified *in silico* by comparative genomics (CRIT: Complement C2 receptor inhibitor trispanning (CRIT), C4BP: m⁷G, 7-methyl guanosine; CSP, circumsporozoite protein; Complement-binding protein, CR1: Complement receptor 1, FHL-1: factor-H-like protein-1, fh: factor H, MCP: Membrane cofactor protein, DAF: Decay-accelerating factor).

doi:10.1371/journal.pone.0017546.t001

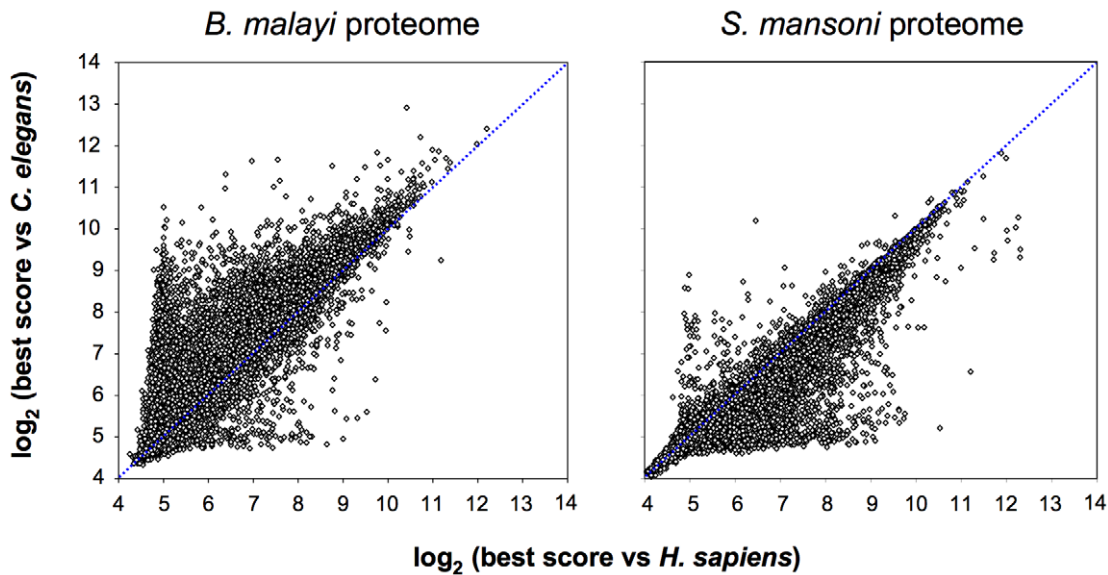


Figure 1. Scatter plot of the blast scores of all proteins from *B. malayi* (left) and *S. mansoni* (right) vs. the host *H. sapiens* (x-axis) and the control *C. elegans* (y-axis). Points below the blue dotted line represent parasite proteins with better scores to *H. sapiens* than to *C. elegans*. doi:10.1371/journal.pone.0017546.g001

search returned 84 hits, most of which from *S. mansoni* (52) and *B. malayi* (15; Table S1). One hit from *B. malayi* was a predicted protein (A8NPN8) with strong similarity to human suppressor of cytokine signaling 5 (SOCS5), in particular to the SH2 domain and the SOCS box (Figure 2). Human SOCS5 was shown to inhibit the IL-4 pathway in T helper cells, promoting T_H1 differentiation [33]. The SH2 domain recognizes the target molecule and the SOCS box recruits the ubiquitin complex that mediates proteosomal degradation of the target [34]. SOCS proteins being crucial regulators of both innate and adaptive immunity, the SOCS5-like protein from *B. malayi* is an interesting candidate. However, it does not carry an export signal and it is therefore not clear how it should interact with host proteins. Possibly, it is released when parasites die.

The known mimicry candidate CRIT (complement C2 receptor inhibitory trispanning, Table 1), which is almost identical between *S. mansoni* and *H. sapiens* [35], was not identified here because human CRIT is not included in the reviewed human proteome from Swissprot (Table 2). Searching against the whole human Uniprot dataset readily returned *S. mansoni* CRIT as the top hit. In the classical complement pathway CRIT blocks the formation of C3 convertase by decreasing the association of C2 with C4b; once C2 is attached to the receptor, it cannot be cleaved by C1 to produce C2a and C2b and thus C3 convertase is no longer formed – the classical pathway is disrupted [25]. It is easy to conceive that a parasite gains an advantage in the human body by exhibiting CRIT and diminishing the proinflammatory response. Based on the high level of DNA similarity *S. mansoni* is thought to have acquired the CRIT gene by horizontal transfer [25,35]. However, while CRIT orthologues are present in all of the sequenced *Schistosoma* species and in *T. cruzi*, the only mammals which possess CRIT are man and rat (Figure S1). This enigmatic distribution can only be explained by multiple instances of gene transfer or gene loss in mammals. Postulating a minimal number of horizontal transfers, a parsimonious interpretation would place the origin of the CRIT gene to schistosomes. The gene could have been acquired (exapted) from the parasites by *H. sapiens* and *R. norvegicus* independently, and finally picked up by *T. cruzi* from a

mammalian host. In this scenario, only the CRIT of *T. cruzi* would be a case of molecular mimicry.

Molecular mimicry surveys with fragmented protein sequences

Several known cases of molecular mimicry from parasites (Table 1) involve shorter peptides, e.g. the thrombospondin motif in *P. falciparum* circumsporozoite protein CSP. Such mimicry

Table 2. Organisms used in this study.

Species	Proteins	Source	Ref.
<i>Brugia malayi</i>	11551	Uniprot	[68]
<i>Cryptosporidium parvum</i>	3805	CryptoDB	[69]
<i>Giardia lamblia</i>	5901	GiardiaDB	[70]
<i>Leishmania major</i>	8406	TritypDB	[71]
<i>Plasmodium falciparum</i>	5479	PlasmoDB	[72]
<i>Schistosoma mansoni</i>	13157	Sanger	[73]
<i>Trichomonas vaginalis</i>	50155	Uniprot	[74]
<i>Trypanosoma cruzi</i>	23031	TritypDB	[75]
<i>Homo sapiens</i>	20298	Uniprot	[76]
<i>Aedes aegypti</i>	16531	Vectorbase	[77]
<i>Anopheles gambiae</i>	14103	Vectorbase	[78]
<i>Arabidopsis thaliana</i>	36671	EBI	[79]
<i>Caenorhabditis elegans</i>	24143	Wormbase	[80]
<i>Ciona intestinalis</i>	15852	JGI	[81]
<i>Schizosaccharomyces pombe</i>	4977	EBI	[82]
<i>Trichoplax adhaerens</i>	11585	Uniprot	[83]

Parasite (top), host (middle), and negative control species (bottom), their predicted number of protein-coding genes, and source of the predicted proteome file (EBI: European Bioinformatics Institute, JGI: Joint Genome Institute).

doi:10.1371/journal.pone.0017546.t002

SOCS5	TQIDYIHCLVPDLLQITGNPCYWGVMRDYEAEEALLEGKPEGTFLLRDSAQEDYLFVSFSFR	420
A8NPN8	TSVHYTNCLVPRLDLIIDSSYWGIMDRYEAEEALLDNKPEGTFLLRDSAQSEYLFVSFSFR	144
	.:. :**** * * . . . ***:*****: .*****:*****	
SOCS5	RYNRSLHARIEQWNHNF SFD AHDPCV FHSSTVTGLLEHYKDPSSCMFFEPLLTISLNRTF	480
A8NPN8	RYKRTLHARIEQKNHRFSFD FSDPSIYSANTITKLISYKDP TKCLFFEPQLSVPLPRNF	204
	.:*** **_**** **.: :.:* *.:*****:.*:**** *.:.* *.*	
SOCS5	PFSLQYICRAVICRCTTYD GIDGLPLPSMLQDFLKEYHYKQKVRVRWLERE-----	531
A8NPN8	VFSLQHLCRARIASLT TYD GVEKLNLPVSLKNFIKEYHYKHPVKTVNYTPD TDLHAYTL	264
	****:*** *. *****: * ** *:*:*****: *.: :	

Figure 2. ClustalW alignment of the candidate mimicry region in A8NPN8 from *B. malayi* to *H. sapiens* SOCS5. The SH2 domain is shaded in yellow, the SOCS box domain in blue. The N-terminal parts of the two proteins do not share any similarity (not shown). doi:10.1371/journal.pone.0017546.g002

candidates would not be detected with the above approach using full-length protein sequences. Thus we refined the systematic survey and developed a peptide-based pipeline for detection of mimicry candidates as outlined in Figure 3. In brief, the parasite proteins were converted to a series of overlapping 14-mers, each of which was searched with ungapped blastp against the control proteomes *C. elegans*, *S. pombe*, *A. thaliana*, *C. intestinalis*, or *T. adhaerens*. The 14-mers with high similarity to any sequence of the controls were filtered out using an empirically developed scheme (Figure S2). The remainder of the 14-mers was screened, again with ungapped blastp, against the *H. sapiens* proteome and those exhibiting strong similarity (Figure S2) to a human sequence were identified as molecular mimicry candidates. For this approach, predicted N-terminal protein export signal sequences were removed since they resemble each other and might produce false positive hits. Parasite 14-mers with 100% identity to a human protein were obtained from *B. malayi* (4), *C. parvum* (1), *P. falciparum* (13) and *S. mansoni* (15). 14-mers with 13 identical residues to a human protein were found in all parasites except *G. lamblia*. The number of hits is summarized in Figure 4. As a control, the same approach (Figure 3) was carried out with versions of the pathogen proteomes where every sequence had been scrambled randomly. This yielded not a single 14-mer of 100% identity to a human protein over all the parasites tested, and only 4 with 13 identities in, underscoring the statistical significance of the identified mimicry candidates. The largest differences between real and randomized proteins were observed for the helminths *B. malayi* and *S. mansoni*, and for *P. falciparum*. Selected mimicry candidates from these parasites are listed in Table 3. The selection was based on number of identical residues, Shannon-entropy of the respective 14-mer as a measure of sequence heterogeneity, and GO terms associated with the hit in the human proteome. An overview of all the high-level GO terms of the human proteins which were matched with mimicry candidates from parasites is shown in Table S2. The mimicry candidates of *P. falciparum* enriched for ‘Cellular component biogenesis’, ‘Localization’, and ‘Growth’, while for the helminths *B. malayi* and *S. mansoni* ‘Biological adhesion’ and ‘Rhythmic process’ were overrepresented in the human hits (compared to the complete human proteome; Table S2).

Among the most interesting of the identified mimicry candidates was a match of 17 identical amino acids from *B. malayi* to human plasma glutamate carboxypeptidase. The *B. malayi* protein (A8QH34) had been previously detected in excretory-secretory products in abundance [36,37]. Moreover, the identified candidate has 67% identity to ES-62 from the rodent filarial nematode

Acanthocheilonema viteae (Uniprot ID O76552), a protein with immunomodulatory impact on different host cells depending on the occurrence of phosphorylcholine [38]. The identified candidate stretch shares 14 identical amino acids with ES-62 of *A. viteae*. Other interesting fragments from *B. malayi* matched to human periphilin-1 (Q8NEY8), a protein of cell-cell junctions in differentiated keratinocytes which was proposed to be involved in barrier formation and epidermal integrity [39], and to plasminogen (P00747), the proenzyme of plasmin which dissolves blood clots and acts as a proteolytic factor in various other processes (Table 3).

In *P. falciparum*, the peptide-based approach significantly enriched for exoproteins ($p < 0.0001$, two-sided chi square test), i.e. proteins with transmembrane domains or export signal predicted by Phobius [40]. The best hit overall was to human vitronectin. Several of the *var* family gene products turned out to share a stretch of 13 to 16 identical amino acids with vitronectin. The candidate mimicry motif lies in the extracellular part of PfEMP1, close to the predicted transmembrane domain (Figure 5, bottom). The corresponding sequence in vitronectin is in the N-terminal half, in the first of the heparin-binding motifs between the somatomedin and the central hemopexin domains (Figure 5, top). Vitronectin is a multifunctional protein that promotes cell adhesion, stabilizes plasminogen activator inhibitor 1, and inhibits the formation of the pore-forming membrane attack complex (MAC) of the complement system. Vitronectin is abundant in the extracellular matrix and in the serum [41]. Pathogenic bacteria such as *Neisseria meningitidis* or *Haemophilus influenzae* decorate themselves with human vitronectin which they acquire from the serum through specific binding partners on their surface [42,43]. Bacteria also exploit human vitronectin for cytoadhesion and host cell invasion [44]. Malaria-infected erythrocytes, however, tested negative for binding to human vitronectin [45]. We identified six PfEMP1 variants possessing the candidate mimicry motif to vitronectin in the *P. falciparum* strain 3D7 and seven in the strain HB3 (Figure 5). The motif is positionally conserved relative to the transmembrane domain of PfEMP1. Searching the non-redundant protein database of GenBank with the corresponding peptide ‘NPEQTPVLKPEEEAP’ returned significant hits (expectancy < 0.001) only from *H. sapiens*, Chimpanzee, Orangutan, and *P. falciparum* (not shown). Interestingly, the genome project of the simian and human malaria parasite *P. knowlesi* had uncovered a candidate molecular mimicry motif to the immunoregulatory host protein CD99 in the extracellular domain of the *kir* gene family products [46].

The fragment-based approach for mimicry candidates in *P. falciparum* also returned a triad between host, vector and parasite. Thrombospondin-related anonymous protein (TRAP, PF13_0201)

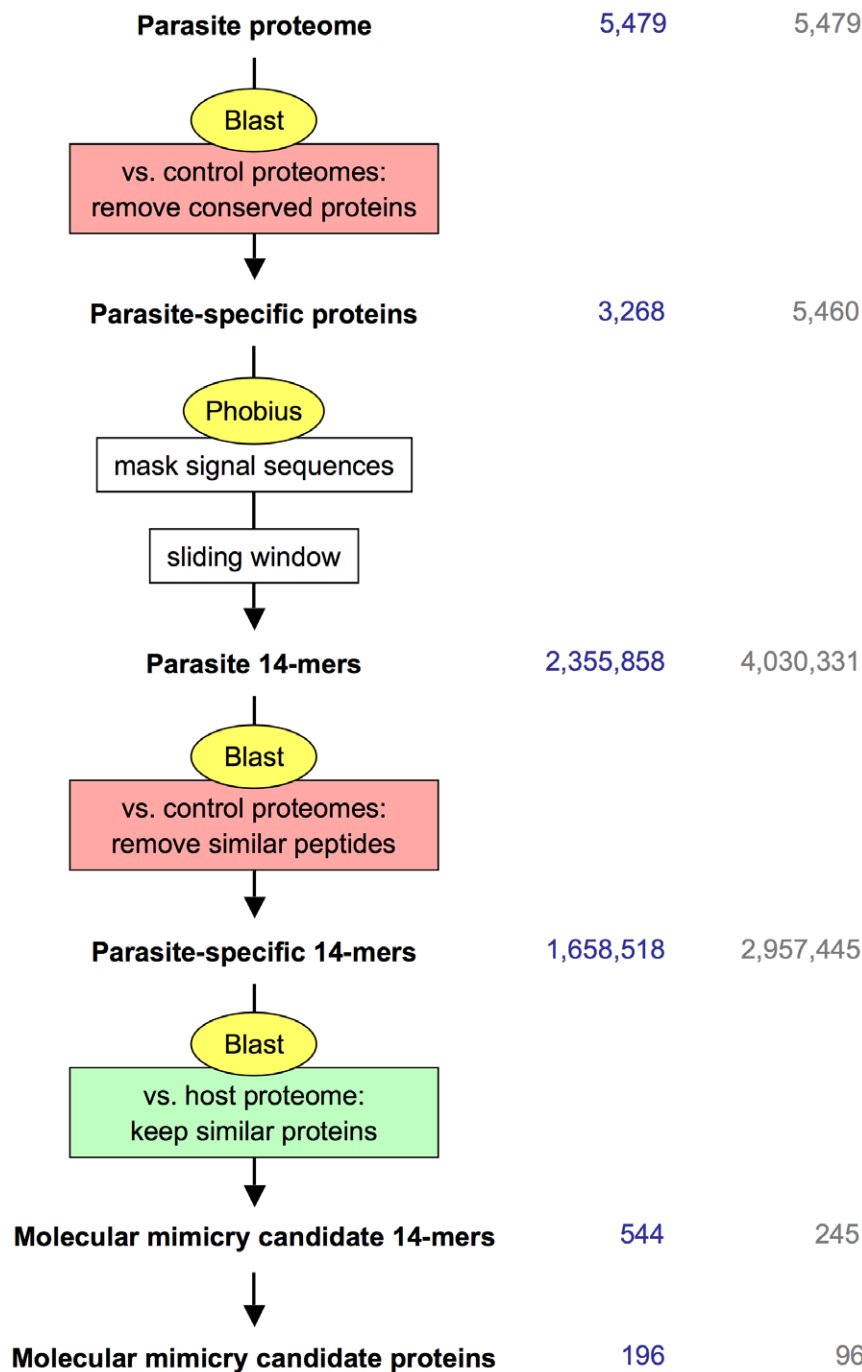


Figure 3. The *in silico* pipeline for identification of molecular mimicry candidates from parasites. See Methods for details. The process is illustrated with the actual numbers from the analysis of the *P. falciparum* proteome in blue, respectively a randomized version of it in grey, vs. the host *H. sapiens*.

doi:10.1371/journal.pone.0017546.g003

of *P. falciparum* matched with the human spondin (Q9HCB6) and a hypothetical protein from *A. gambiae* (AGAP012307, not shown). In the human protein, the region lies in the thrombospondin type-I repeat (TSR) domain which binds to heparin sulphate proteoglycans on hepatocytes [47,48]. This mimicked structure was also found on the circumsporozoite protein (CSP) and has been known for a long time [49]. Whereas CSP mediates the binding of the parasites to the human liver, it is suggested that TRAP is crucial for sporozoite locomotion and cell invasion [50,51]. Interestingly,

the same part of the TSR domain of TRAP has been matched with the *A. gambiae* proteome and it has been demonstrated with loss-of-function mutations that this region is involved in the sporozoite invasion into mosquito salivary glands [52].

mimicDB - Database for molecular mimicry candidates from pathogens

All mimicry candidates from parasites to mammalian and insect hosts (Table 2) were stored in a relational database, mimicDB,

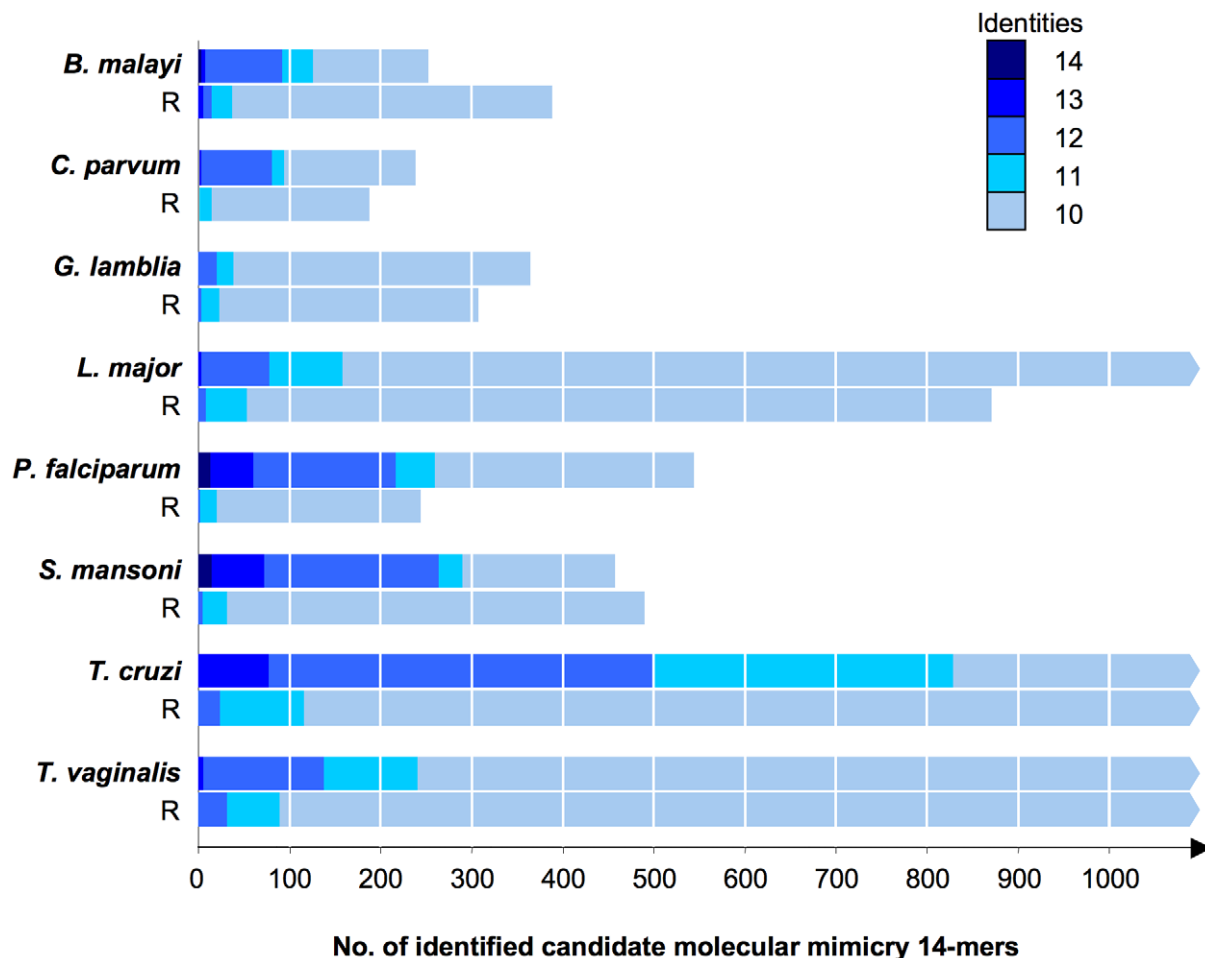


Figure 4. Numbers of identified candidate molecular mimicry 14-mers from parasite proteomes and randomized versions thereof (R). Numbers of amino acid identities between the 14-mers and their best hit in the human proteome are color-coded as indicated. doi:10.1371/journal.pone.0017546.g004

which is publicly accessible via <<http://mimicdb.scilifelab.se>>. The database was designed for ease of community access to the mimicry data (Figure S3). It can be queried using keywords from gene description, different formats of gene and protein accession numbers and names, and in general on free text on the available

data. GO terms are tightly integrated into the database, and queries can be made both on leaf-terms as well as directly onto broader categories higher up in the hierarchy. The queries can be restricted to species using special qualifiers. From the resulting tables, links are provided directly to entries in large public databases (Uniprot,

Table 3. Selected mimicry candidates.

Parasite protein	14-mer motif	Ent.	Id.	Human protein
<i>Bma</i> A8PSR3, uncharacterized protein	TFGFVTKMLIEKDP	3.24	12	Q8NEY8, periphilin-1
<i>Bma</i> A8Q9C9, uncharacterized protein	RKSSQKIRMRDVVL	3.04	12	P00747, plasminogen
<i>Bma</i> A8QH34, leucyl aminopeptidase	HLDSWDVQGAMDD	3.09	14	Q9Y646, plasma glutamate carboxypeptidase
<i>Bma</i> A8PP49, pregnancy-associated plasma protein E	CYIYEGDGECEPFE	2.81	12	Q9BXP8, pappalysin-2
<i>Sma</i> Smp_111120, insulin receptor kinase substrate	SLEKSQAEKKIRR	2.90	14	Q9UHR4, insulin receptor tyrosine kinase substrate
<i>Sma</i> Smp_109770, integrin alpha-4	APNVSMEIMVPNSF	3.09	13	P13612, integrin alpha-4
<i>Pfa</i> PF07_0048, PfEMP1	NPEQTPVLKPEEEA	2.90	14	P04004, vitronectin
<i>Pfa</i> MAL13P1.34, RED-like protein	SKFMGGDEEHTHLV	3.38	12	Q13123, IK cytokine
<i>Pfa</i> PF13_0201, TRAP	WDEWSPCSVTCGKG	3.24	12	Q9HCB6, spondin-1

Hits from *B. malayi* (*Bma*), *S. mansoni* (*Sma*) and *P. falciparum* (*Pfa*) and their human match (Ent, Shannon entropy in bits; Id, number of identities). doi:10.1371/journal.pone.0017546.t003

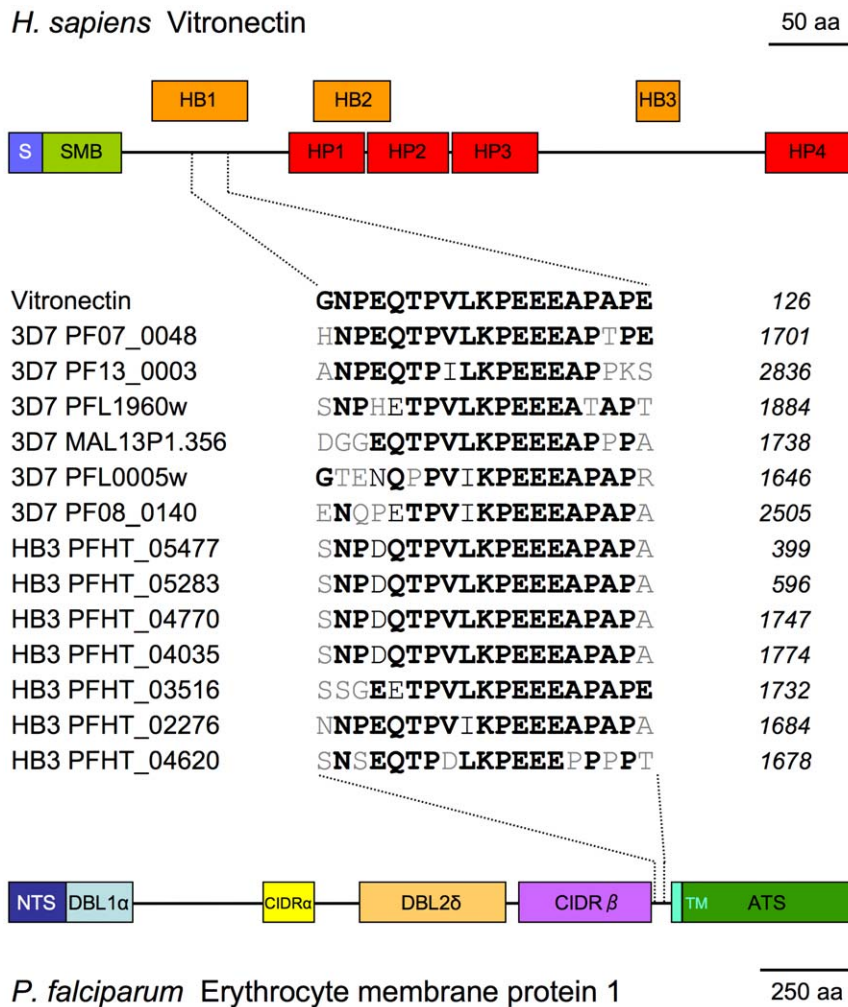


Figure 5. Alignment of human vitronectin (top) and *P. falciparum* PfEMP1 variants (bottom). Identities to vitronectin are printed in bold black, similarities in black. The known vitronectin domains are the signal sequence (blue), somatomedin-B (green), and hemopexin (red). The known PfEMP1 domains are the N-terminal segment (dark blue), Duffy Binding Like α (light blue), cysteine-rich interdomain region α (yellow), Duffy Binding Like 2d (orange), cysteine-rich interdomain region β (purple), transmembrane domain (cyan), acidic terminal segment (green). doi:10.1371/journal.pone.0017546.g005

NCBI) as well as to detailed sequence views. Predicted protein motifs and signal peptides are visualized on the source and target sequences together with the candidate mimicry motifs.

Conclusion

To our knowledge this is the first *in silico* survey for molecular mimicry candidates in parasites. Its systematic, genome-wide nature warrants that all linear amino acid epitopes involved in molecular mimicry between a given parasite and its host are going to be detected. False positive hits can be tracked by including the appropriate controls: proteomes of free-living species to eliminate the proteins which are generally conserved across phyla, and scrambled versions of the parasite proteomes to estimate for random hits resulting from the sheer number of analyzed sequences. False negatives are more problematic; mimicry by non-linear epitopes composed from amino acids of separate folds (or even separate polypeptides) will not be recognized, and neither are glycosylated epitopes (Table 1). Nevertheless, there are examples of molecular mimicry by linear epitopes which are straightforward to detect by comparative genomics as performed here. Proof of concept was obtained from the fact that the known molecular mimicry motif in

TRAP (thrombospondin-related anonymous protein) from *P. falciparum* was detected readily. Many new molecular mimicry candidates were discovered from human parasites, in particular from *B. malayi*, *S. mansoni* and *P. falciparum*, most notably a sequence shared between human vitronectin and several of the *P. falciparum* erythrocyte membrane protein 1 variants. All the identified mimicry candidates are stored in a relational database called mimicDB and searchable on-line. We hope that mimicDB will stimulate research into molecular mimicry of parasites. Given its numerous potential benefits – camouflage, cytoadherence, manipulation of host signaling – molecular mimicry may well be much more common among parasitic microorganisms than currently known.

Materials and Methods

Proteome files

Predicted proteins from completely sequenced genomes (Table 2) were obtained from ftp.ebi.ac.uk (*Arabidopsis thaliana*, *Schizosaccharomyces pombe*), www.tritrypdb.org (*Leishmania major*, *Trypanosoma cruzi*), www.cryptodb.org (*Cryptosporidium parvum*) www.giardiadb.org (*Giardia lamblia*) www.plasmodb.org (*Plasmodium*

falciparum 3D7), www.broadinstitute.org (*Plasmodium falciparum* HB3), ftp.vectorbase.org (*Aedes aegypti*, *Anopheles gambiae*), ftp.wormbase.org (*Caenorhabditis elegans*), ftp.sanger.ac.uk (*Schistosoma mansoni*), ftp.jgi-psf.org (*Ciona intestinalis*), and www.uniprot.org (*Brugia malayi*, *Homo sapiens*, *Trichomonas vaginalis*, *Trichoplax adhaerens*).

Algorithm

BLAST 2.2.17 [53] was obtained from ftp.ncbi.nlm.nih.gov, Phobius 1.01 [40] from <phobius.sbc.su.se>. Automated detection of molecular mimicry candidates as depicted in Figure 3 was performed with Perl scripts, available on request. First, those of the predicted parasite proteins which are generally conserved among eukaryotes were sorted out based on full-length blastp searches against the proteomes of *C. elegans*, *C. intestinalis*, *T. adhaerens*, *S. pombe* and *A. thaliana*. Sequences which returned an e-value $\leq 10^{-10}$ to any sequence of these control proteomes were filtered out. The remaining parasite proteins were run through Phobius and predicted N-terminal export signal sequences were cut off at the predicted cleavage site. Then, the protein sequences were converted to a series of overlapping 14-mers with a sliding window of increment one. The resulting peptides were screened against the five control proteomes with ungapped blastp, and 14-mers above the empirically determined identity threshold (represented by the red line in Figure S2) were removed. With the remaining, parasite-specific 14-mers, an ungapped blastp search was performed against the host proteome and hits above the empirically determined identity threshold (green line in Figure S2) were considered molecular mimicry candidates. Randomized sequences were generated with 'shuffleseq' of the EMBOSS package [54]. All programs were run on the University of Bern Linux cluster, Ubelix <http://ubelix.unibe.ch>. Multiple sequence alignments were performed using ClustalX [55].

Database

The mimicDB database (http://mimicDB.scilifelab.se) uses MySQL as its relational database engine. The database was designed as an extension to the GO term [56] database schema for ease of interrogation on the complete GO hierarchy rather than leaf term only (Figure S3). Protein motif predictions were obtained using hmmer 3.0 [57] with the PFAM database v24.0 [58], and signal peptide predictions using Phobius 1.01 [40]. *Ad hoc* Perl scripts were used to import the mimicry pipeline results, predicted motifs and signals as well as calculate Shannon source entropy for peptides. The interface was constructed using Perl and the Titanium extension to CGI.pm. A package to reconstruct the results and database is available from the authors upon request or can be downloaded from the mimicDB web site.

Supporting Information

Figure S1 ClustalW dendrogram of CRIT orthologues from *Schistosoma mansoni* (Sma), *S. haematobium* (Sha), *S. japonicum* (Sja), *Trypanosoma cruzi* (Tcr), *H. sapiens* (Hsa), and *R. norvegicus* (Rno). The scale bar

indicates changes per site. Bootstrapping numbers (grey) are given as percent positives of 1,000 rounds. (TIF)

Figure S2 The filtering system used in the overlapping fragments approach. Numbers represent identical amino acid residues. Red line: threshold for negative control species. Green line: threshold for molecular mimicry candidate in mammalian host or insect vector. (TIF)

Figure S3 Database schema of mimicDB. The mimicDB database schema centers around *mimic_sequence*, which represents the individual genes. This table has as attribute tables the actual peptide sequences (*mimic_sequence_seq*) and predicted motifs (*mimic_sequence_motif*). Hits between parts of these genes are collected in *mimic_hit*, which stores the coordinates and properties of the hit. A complexity measure, in the form of Shannon source entropy for each peptide hit is stored in *mimic_hit_entropy*. The database connects to the GO consortium GO term database in that *mimic_sequence* entries that have a GO association are referenced by entries in *mimic_sequence_with_go_association*, where the corresponding GO term db gene_product::id is also a foreign key. (TIF)

Table S1 All molecular mimicry candidates identified searching the human proteome with full-length protein sequences from parasites. Scores are from blastp searches using the BLOSUM62 matrix and default parameters. Ratios are of the score against *H. sapiens* divided by the best score achieved against any of the control species *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Ciona intestinalis*, *Schizosaccharomyces pombe*, or *Trichoplax adhaerens*. (XLS)

Table S2 Molecular mimicry candidates identified searching the human proteome with fragmented protein sequences from parasites. Hits are sorted according to GO (gene ontology) process annotation of the respective human target protein. Enrichment ('Enrich') of GO terms in the identified sets of target proteins is expressed in relation to the abundance of the same GO terms in the complete human proteome (last three columns). (XLS)

Acknowledgments

We wish to thank the University of Bern for user time on their Linux cluster Ubelix.

Author Contributions

Conceived and designed the experiments: PL DN PM. Performed the experiments: PL DN. Analyzed the data: PL DN PM. Contributed reagents/materials/analysis tools: PL DN PM. Wrote the manuscript: PL DN PM. Designed mimicDB: DN.

References

- Lambris JD, Ricklin D, Geisbrecht BV (2008) Complement evasion by human pathogens. *Nat Rev Microbiol* 6: 132–142.
- Srinivasappa J, Saegusa J, Prabhakar BS, Gentry MK, Buchmeier MJ, et al. (1986) Molecular mimicry: frequency of reactivity of monoclonal antiviral antibodies with normal tissues. *J Virol* 57: 397–401.
- Michelson S (2004) Consequences of human cytomegalovirus mimicry. *Hum Immunol* 65: 465–475.
- Damian RT (1964) Molecular mimicry: Antigen sharing by parasite and host and its consequences. *American Naturalist* 98: 129–149.
- Damian RT (1962) A theory of immunoselection for eclipsed antigens of parasites and its implications for the problem of antigenic polymorphism in man. *J Parasitol* 48: 16.
- Ouaissi MA, Afchain D, Capron A, Grimaud JA (1984) Fibronectin receptors on *Trypanosoma cruzi* trypomastigotes and their biological function. *Nature* 308: 380–382.
- Ouaissi MA, Cornette J, Afchain D, Capron A, Gras-Masse H, et al. (1986) *Trypanosoma cruzi* infection inhibited by peptides modeled from a fibronectin cell attachment domain. *Science* 234: 603–607.

8. Baruch DI, Gormely JA, Ma C, Howard RJ, Pasloske BL (1996) Plasmodium falciparum erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proc Natl Acad Sci U S A* 93: 3497–3502.
9. Howell DP, Levin EA, Springer AL, Kraemer SM, Phippard DJ, et al. (2008) Mapping a common interaction site used by Plasmodium falciparum Duffy binding-like domains to bind diverse host receptors. *Mol Microbiol* 67: 78–87.
10. Hide G, Gray A, Harrison CM, Tait A (1989) Identification of an epidermal growth factor receptor homologue in trypanosomes. *Mol Biochem Parasitol* 36: 51–59.
11. Ghansah TJ, Ager EC, Freeman-Junior P, Villalta F, Lima MF (2002) Epidermal growth factor binds to a receptor on Trypanosoma cruzi amastigotes inducing signal transduction events and cell proliferation. *J Eukaryot Microbiol* 49: 383–390.
12. Spiliotis M, Kroner A, Brehm K (2003) Identification, molecular characterization and expression of the gene encoding the epidermal growth factor receptor orthologue from the fox-tapeworm Echinococcus multilocularis. *Gene* 323: 57–65.
13. Vicogne J, Cailliau K, Tulasne D, Browaeys E, Yan YT, et al. (2004) Conservation of epidermal growth factor receptor function in the human parasitic helminth Schistosoma mansoni. *J Biol Chem* 279: 37407–37414.
14. Kaslow DC, Quakyi IA, Syin C, Raum MG, Keister DB, et al. (1988) A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains. *Nature* 333: 74–76.
15. Han HJ, Park SG, Kim SH, Hwang SY, Han J, et al. (2004) Epidermal growth factor-like motifs 1 and 2 of Plasmodium vivax merozoite surface protein 1 are critical domains in erythrocyte invasion. *Biochem Biophys Res Commun* 320: 563–570.
16. Blackman MJ, Ling IT, Nicholls SC, Holder AA (1991) Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. *Mol Biochem Parasitol* 49: 29–33.
17. Duvaux-Miret O, Stefano GB, Smith EM, Dissous C, Capron A (1992) Immunosuppression in the definitive and intermediate hosts of the human parasite Schistosoma mansoni by release of immunoreactive neuropeptides. *Proc Natl Acad Sci U S A* 89: 778–781.
18. Biron DG, Marche L, Ponton F, Loxdale HD, Galeotti N, et al. (2005) Behavioural manipulation in a grasshopper harbouring hairworm: a proteomics approach. *Proc Biol Sci* 272: 2117–2126.
19. Oliver-Gonzalez J (1944) Functional antigens in helminths. *J Infect Diseases* 78: 232–237.
20. Ponce de Leon P, Valverde J (2003) ABO System: molecular mimicry of Ascaris lumbricoides. *Rev Inst Med Trop Sao Paulo* 45: 107–108.
21. Oliver-Gonzalez J, Torregrosa MV (1944) A substance in animal parasites related to human isoagglutinogens. *J Infect Diseases* 74: 173–177.
22. Nyame AK, Debose-Boyd R, Long TD, Tsang VC, Cummings RD (1998) Expression of Lex antigen in Schistosoma japonicum and S.haematobium and immune responses to Lex in infected animals: lack of Lex expression in other trematodes and nematodes. *Glycobiology* 8: 615–624.
23. Ben-Ismaïl R, Mulet-Clamagirand C, Carme B, Gentilini M (1982) Biosynthesis of A, H, and Lewis blood group determinants in Fasciola hepatica. *J Parasitol* 68: 402–407.
24. Lu B, PereiraPerrin M (2008) A novel immunoprecipitation strategy identifies a unique functional mimic of the glial cell line-derived neurotrophic factor family ligands in the pathogen Trypanosoma cruzi. *Infect Immun* 76: 3530–3538.
25. Inal JM, Hui KM, Miot S, Lange S, Ramirez MI, et al. (2005) Complement C2 receptor inhibitor trispanning: a novel human complement inhibitory receptor. *J Immunol* 174: 356–366.
26. Lehr T, Geyer H, Maass K, Doenhoff MJ, Geyer R (2007) Structural characterization of N-glycans from the freshwater snail Biomphalaria glabrata cross-reacting with Schistosoma mansoni glycoconjugates. *Glycobiology* 17: 82–103.
27. Holmquist G, Udonsangpetch R, Berzins K, Wigzell H, Perlmann P (1988) Plasmodium chabaudi antigen Pch105, Plasmodium falciparum antigen Pf155, and erythrocyte band 3 share cross-reactive epitopes. *Infect Immun* 56: 1545–1550.
28. Ponce de Leon P, Foresto P, Valverde J (2005) H antigen presence in an Ascaris lumbricoides extract. *Rev Inst Med Trop Sao Paulo* 47: 159–160.
29. Ramos M, Alvarez I, Sesma L, Logean A, Rognan D, et al. (2002) Molecular mimicry of an HLA-B27-derived ligand of arthritis-linked subtypes with chlamydial proteins. *J Biol Chem* 277: 37573–37581.
30. Hall R (1994) Molecular mimicry. *Adv Parasitol* 34: 81–132.
31. Halaných K (2004) The new view of animal phylogeny. *Annu Rev Ecol Evol Syst* 35: 229–256.
32. Gamulin V, Muller I, Muller W (2000) Sponge proteins are more similar to those of Homo sapiens than to Caenorhabditis elegans. *Biol J Linn Soc* 71: 821–828.
33. Seki Y, Hayashi K, Matsumoto A, Seki N, Tsukada J, et al. (2002) Expression of the suppressor of cytokine signaling-5 (SOCS5) negatively regulates IL-4-dependent STAT6 activation and Th2 differentiation. *Proc Natl Acad Sci U S A* 99: 13003–13008.
34. Yoshimura A, Naka T, Kubo M (2007) SOCS proteins, cytokine signalling and immune regulation. *Nat Rev Immunol* 7: 454–465.
35. Inal JM (2005) Complement C2 receptor inhibitor trispanning: from man to schistosome. *Springer Semin Immunopathol* 27: 320–331.
36. Hewitson JP, Harcus YM, Curwen RS, Dowle AA, Atmadja AK, et al. (2008) The secretome of the filarial parasite, Brugia malayi: proteomic profile of adult excretory-secretory products. *Mol Biochem Parasitol* 160: 8–21.
37. Bennuru S, Semnani R, Meng Z, Ribeiro JM, Veenstra TD, et al. (2009) Brugia malayi excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl Trop Dis* 3: e410.
38. Goodridge HS, Stepek G, Harnett W, Harnett MM (2005) Signalling mechanisms underlying subversion of the immune response by the filarial nematode secreted product ES-62. *Immunology* 115: 296–304.
39. Kazerounian S, Aho S (2003) Characterization of periphilin, a widespread, highly insoluble nuclear protein and potential constituent of the keratinocyte cornified envelope. *J Biol Chem* 278: 36707–36717.
40. Käll L, Krogh A, Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 338: 1027–1036.
41. Schwartz I, Seger D, Shaltiel S (1999) Vitronectin. *Int J Biochem Cell Biol* 31: 539–544.
42. Blom AM, Hallstrom T, Riesbeck K (2009) Complement evasion strategies of pathogens-acquisition of inhibitors and beyond. *Mol Immunol* 46: 2808–2817.
43. Singh B, Su YC, Riesbeck K (2010) Vitronectin in bacterial pathogenesis: A host protein used in complement escape and cellular invasion. *Mol Microbiol* accepted article.
44. Bergmann S, Lang A, Rohde M, Agarwal V, Rennemeier C, et al. (2009) Integrin-linked kinase is required for vitronectin-mediated internalization of Streptococcus pneumoniae by host cells. *J Cell Sci* 122: 256–267.
45. Sherwood JA, Roberts DD, Marsh K, Harvey EB, Spitalnik SL, et al. (1987) Thrombospondin binding by parasitized erythrocyte isolates in falciparum malaria. *Am J Trop Med Hyg* 36: 228–233.
46. Pain A, Bohme U, Berry AE, Mungall K, Finn RD, et al. (2008) The genome of the simian and human malaria parasite Plasmodium knowlesi. *Nature* 455: 799–803.
47. Muller HM, Reckmann I, Hollingdale MR, Bujard H, Robson KJ, et al. (1993) Thrombospondin related anonymous protein (TRAP) of Plasmodium falciparum binds specifically to sulfated glycoconjugates and to HepG2 hepatoma cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. *EMBO J* 12: 2881–2889.
48. Robson KJ, Frevort U, Reckmann I, Cowan G, Beier J, et al. (1995) Thrombospondin-related adhesive protein (TRAP) of Plasmodium falciparum: expression during sporozoite ontogeny and binding to human hepatocytes. *EMBO J* 14: 3883–3894.
49. Robson KJ, Hall JR, Jennings MW, Harris TJ, Marsh K, et al. (1988) A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* 335: 79–82.
50. Sultan AA, Thathy V, Frevort U, Robson KJ, Crisanti A, et al. (1997) TRAP is necessary for gliding motility and infectivity of plasmodium sporozoites. *Cell* 90: 511–522.
51. Menard R (2000) The journey of the malaria sporozoite through its hosts: two parasite proteins lead the way. *Microbes Infect* 2: 633–642.
52. Matuschewski K, Nunes AC, Nussenzeig V, Menard R (2002) Plasmodium sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. *EMBO J* 21: 1597–1606.
53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
54. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Biochem Sci* 25: 276–277.
55. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23: 403–405.
56. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat Genet* 25: 25–29.
57. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14: 755–763.
58. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281–288.
59. Wang S, Browning KS, Miller WA (1997) A viral sequence in the 3'-untranslated region mimics a 5' cap in facilitating translation of uncapped mRNA. *EMBO J* 16: 4107–4116.
60. Kraiczky P, Wurzner R (2006) Complement escape of human pathogenic bacteria by acquisition of complement regulators. *Mol Immunol* 43: 31–44.
61. Diaz A, Ferreira A, Sim RB (1997) Complement evasion by Echinococcus granulosus: sequestration of host factor H in the hydatid cyst wall. *J Immunol* 158: 3779–3786.
62. Meri T, Jokiranta TS, Hellwage J, Bialonski A, Zipfel PF, et al. (2002) Onchocerca volvulus microfilariae avoid complement attack by direct binding of factor H. *J Infect Dis* 185: 1786–1793.
63. Cerami C, Frevort U, Sinnis P, Takacs B, Clavijo P, et al. (1992) The basolateral domain of the hepatocyte plasma membrane bears receptors for the circumsporozoite protein of Plasmodium falciparum sporozoites. *Cell* 70: 1021–1033.
64. Rubin-de-Celis SS, Uemura H, Yoshida N, Schenkman S (2006) Expression of trypomastigote trans-sialidase in metacyclic forms of Trypanosoma cruzi increases parasite escape from its parasitophorous vacuole. *Cell Microbiol* 8: 1888–1898.

65. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz-Renggli C, et al. (2004) Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. *J Exp Med* 199: 1445–1450.
66. Mauss EA (1941) Occurrence of Forssman heterogenic antigen in the nematode, *Trichinella spiralis*. *J Immunol* 42: 71–77.
67. Shear HL, Nussenzweig RS, Bianco C (1979) Immune phagocytosis in murine malaria. *J Exp Med* 149: 1288–1298.
68. Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, et al. (2007) Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317: 1756–1760.
69. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, et al. (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304: 441–445.
70. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, et al. (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317: 1921–1926.
71. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, et al. (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309: 436–442.
72. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498–511.
73. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352–358.
74. Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, et al. (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 315: 207–212.
75. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409–415.
76. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. *Science* 291: 1304–1351.
77. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, et al. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316: 1718–1723.
78. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298: 129–149.
79. Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
80. *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282: 2012–2018.
81. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298: 2157–2167.
82. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, et al. (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415: 871–880.
83. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, et al. (2008) The *Trichoplax* genome and the nature of placozoans. *Nature* 454: 955–960.