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.

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, *Plasmodium*, a protozoan that infects red blood cells of both human and non-human primate species. The infected red blood cells stick to the inner surface of blood vessels, which are then irreversibly damaged, leading to cerebral malaria. After invasion, the parasite differentiates quickly into a form called the trophozoite, which is the primary form of the parasite within the host and is responsible for further growth and division within the human body. The trophozoite stage is then followed by a mature, elongated form called the schizont, which facilitates the maturation and release of new blood parasites. The parasites can then continue the cycle of infection by releasing new blood parasites into the bloodstream, allowing them to infect new red blood cells. During this cycle, the parasites undergo rapid cell division, producing large numbers of new parasites, which can rapidly spread throughout the body, potentially causing widespread infection and illness. The infection is typically asymptomatic in most cases, but can progress to a more severe, life-threatening condition known as cerebral malaria, which is characterized by high fever, severe headaches, confusion, and occasionally seizures. In African countries, cerebral malaria is a leading cause of death in children under the age of five.

The parasite's ability to evade host immune responses is critical for its survival and transmission. To achieve this, the parasite may use strategies such as antigenic variation, whereby specific antigens are rapidly replaced with new forms that are not recognized by the host's immune system. This is particularly evident in *Plasmodium* infections, where the parasite undergoes rapid antigenic variation, allowing it to evade the host's immune response. In addition to antigenic variation, the parasite may also use other strategies, such as the production of molecular mimics, to evade host immune responses. Molecular mimics are structures that mimic the appearance of host proteins, allowing the parasite to escape detection by the host's immune system. 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. Trypomastigotes of *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 (MSPI) in the blood-stage where it is critical for erythrocyte invasion [15,16]. Schistosomes send immunosuppressory signals in the form of neuropetides to both the definite host [15] and the intermediate host [16]. 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].

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
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

| Macromolecule | Mimicry by transfer | Mimicry by convergence |
|---------------|---------------------|------------------------|
| **Nucleic acid** | **Schistosoma mansoni 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 mG cap of eukaryotic mRNA to stimulate translation [59]. |
| **Protein** | Pathogenic bacteria, E. granulosus and O. volvulus decorate themselves with inhibitors of the complement cascade sequenced from the blood [43,60,61,62]. | A 16 aa motif in P. falciparum 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. T. cruzi trans-sialidase is a virulence factor in mammals [64]; T. brucei trans-sialidase is required for survival in the tsetse fly [65]. | Several pathogenic helminths synthesize the Forssman antigen (globosperosylceramide) [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: mG, 7-methyl guanosine; CSP: circumsporozoite protein; Complement-binding protein, CR1: Complement receptor 1, FHL-1: factor-H-like protein-1, fht: factor H, MCP: Membrane cofactor protein, DAF: Decay-accelerating factor).

doi:10.1371/journal.pone.0017546.t001
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 TH1 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 parsimonial 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.  |
|--------------------------|----------|----------|-------|
| *Brugia malayi*          | 11551    | Uniprot  | [68]  |
| *Cryptosporidium parvum* | 3805     | CryptoDB | [69]  |
| *Giardia lamblia*        | 5901     | GiardiaDB| [70]  |
| *Leishmania major*       | 8406     | TritypDB | [71]  |
| *Plasmodium falciparum*  | 5479     | PlasmoDB | [72]  |
| *Schistosoma mansoni*    | 13157    | Sanger   | [73]  |
| *Trichomonas vaginalis*  | 50155    | Uniprot  | [74]  |
| *Trypanosoma cruzi*      | 23031    | TritypDB | [75]  |
| *Homo sapiens*           | 20298    | Uniprot  | [76]  |
| *Aedes aegypti*          | 16531    | Vectorbase| [77] |
| *Anopheles gambiae*      | 14103    | Vectorbase| [78] |
| *Arabidopsis thaliana*   | 36671    | EBI      | [79]  |
| *Caenorhabditis elegans* | 24143    | Wormbase | [80]  |
| *Ciona intestinalis*     | 15852    | JGI      | [81]  |
| *Schizosaccharomyces pombe* | 4977   | EBI      | [82]  |
| *Trichoplax adhaerens*   | 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
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 with transmembrane domains or export signal predicted by Phobius 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 PIEMP1, 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 Nisseria meningitides 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 PIEMP1 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 positioned conserved relative to the transmembrane domain of PIEMP1. Searching the non-redundant protein database of GenBank with the corresponding peptide ‘NPEQITPVLKPEEEAP’ returned significant hits (expectancy <0.001) only from H. sapiens, Chlamydiae, 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 immune-regulatory 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)}
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,
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,

![Diagram of identified candidate molecular mimicry 14-mers from parasite proteomes and randomized versions thereof](image)

**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

| Parasite protein | 14-mer motif | Ent. | Id. | Human protein |
|------------------|--------------|------|-----|---------------|
| B. malayi        | TFGFVTMLIEKDP | 3.24 | 12  | Q8NEY8, periphilin-1 |
| B. malayi        | RKSSQKIRMVDVL | 3.04 | 12  | P00747, plasminogen |
| S. mansoni       | CYYEGDGECEPFE | 2.81 | 12  | Q9UHR4, insulin receptor tyrosine kinase substrate |
| P. falciparum    | APVSMMEVNPNSF | 3.09 | 13  | P13612, integrin alpha-4 |
| P. falciparum    | NPEQTPVLPKREEE | 2.90 | 14  | Q90004, vitronectin |
| T. cruzi         | SKFMGDEEHTHLV | 3.38 | 12  | Q13123, IK cytokine |
| T. vaginalis     | WDEWSPCSVTCGKG | 3.24 | 12  | Q9HCB6, spondin-1 |

**Table 3. Selected mimicry candidates.**

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
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...
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. thalasina. Sequences which returned an e-value ≤ 10−10 to any sequence of these control proteomes were filtered out. The remaining parasite proteins were run through Phobius and predicted N-terminal sequence signal signals 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 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 [50], 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)

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. Designed mimicDB: DN.

References
1. Lambris JD, Ricklin D, Geibrechts BV (2008) Complement evasion by human pathogens. Nat Rev Microbiol 6: 132–142.
2. 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.
3. Michelson S (2004) Consequences of human cytomegalovirus mimicry. Hum Immunol 65: 463–475.
4. Damian RT (1962) Molecular mimicry: Antigen sharing by parasite and host and its consequences. American Naturalist 98: 129–149.
5. 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.
6. Ouaissi MA, Afchain D, Capron A, Grimaud JA (1984) Fibronectin receptors on Trypanosoma cruzi trypomastigotes and their biological function. Nature 308: 389–392.
7. Ouaissi MA, Cornette J, Afchain D, Capron A, Grav-Masse H, et al. (1986) Trypanosoma cruzi infection inhibited by peptides modeled from a fibronectin cell attachment domain. Science 234: 603–607.
32. Gamulin V, Muller I, Muller W (2000) Sponge proteins are more similar to
31. Halanych K (2004) The new view of animal phylogeny. Annu Rev Ecol Evol
30. Hall R (1994) Molecular mimicry. Adv Parasitol 34: 81–132.
28. Ponce de Leon P, Foresto P, Valverde J (2005) H antigen presence in an Ascaris
27. Holmquist G, Udomsangpetch R, Berzins K, Wigzell H, Perlmann P (1988)
26. Lehr T, Geyer H, Maass K, Doenhoff MJ, Geyer R (2007) Structural
22. Nyame AK, Debose-Boyd R, Long TD, Tsang VC, Cummings RD (1998)
21. Oliver-Gonzalez J, Torregrosa MV (1994) A substance in animal parasites
20. Ponce de Leon P, Valverde J (2003) ABO System: molecular mimicry of Ascaris
19. Oliver-Gonzalez J (1944) Functional antigens in helminths. J Infect Diseases 78:
18. Biron DG, Marche L, Ponton F, Loxdale HD, Galeotti N, et al. (2005)
16. Blackman MJ, Ling IT, Nicholls SC, Holder AA (1991) Proteolytic processing of
15. Ghansah TJ, Ager EC, Freeman-Junior P, Villalta F, Lima MF (2002)
14. Hide G, Gray A, Harrison CM, Tait A (1989) Identification of an epidermal growth factor receptor homolog in trypanosomes. Mol Biochem Parasitol 40: 37–47
13. Finlay BB (1997) Molecular mimicry and the parasite proteome. Curr Opin Microbiol 10: 563–570.
12. Ponce de Leon P, Valverde J (2002) Molecular mimicry in the human parasite Schistosoma mansoni by release of immunomodulatory peptides. Nat Acad Sci U S A 99: 778–781.
11. Biron DG, Marche L, Ponton F, Loxdale HD, Galeotti N, et al. (2005) Immunomodulatory activity in a grasshopper haemolymph hormone: a proteinomic approach. Proc Biochem 41: 2117–2126.
10. Blackman MJ, Ling IT, Nicholls SC, Holder AA (1991) Proteolytic processing of the Plasmodium falciparum microneme surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. Mol Biochem Parasitol 49: 29–33.
9. Baruch DI, Gormely JA, Ma C, Howard RJ, Pasloske BL (1996) Plasmodium
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.
7. Haug J, Jorgensen K, Olesen P, Vang H, Madsen E (1995) Differential binding of a human anti-Plasmodium falciparum antibody to placental trophozoites and schizonts. J Exp Med 181: 1295–1307.
6. 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.
5. 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.
4. Finlay BB (1997) Molecular mimicry and the parasite proteome. Curr Opin Microbiol 10: 563–570.
3. Finlay BB (1997) Molecular mimicry and the parasite proteome. Curr Opin Microbiol 10: 563–570.
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. Mansa EA (1941) Occurrence of Forssman heterogenic antigen in the nematode, Trichinella spiralis. J Immunol 42: 71–77.
67. Shear HL, Nussenzeit 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: 441–443.
70. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adams 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, Fang E, White O, Berriman M, et al. (2002) The genome sequence of the human malaria parasite Plasmodium falciparum. Science 297: 171–175.

73. 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: 409–413.
74. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304–1351.
75. Nene V, Worthman JR, Lawson D, Haas B, Koehler C, et al. (2007) Genome sequence of Aedes aegypti, a major arbovirus vector. Science 316: 1718–1723.
76. Holt RA, Suhramanian GM, Halbera A, Sutton GG, Charlah R, et al. (2002) The genome sequence of the malaria mosquito Anopheles gambiae. Science 298: 129–149.
77. Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796–815.
78. C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282: 2012–2018.
79. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, et al. (2002) The draft genome of Caenorhabditis elegans: insights into chordate and vertebrate origins. Science 298: 2157–2167.
80. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, et al. (2002) The genome sequence of Schizosaccharomyces pombe. Nature 415: 871–880.
81. Srivastava M, Begovic E, Chapman J, Putnam N, Hellsten U, et al. (2006) The Trichoplax genome and the nature of placozoans. Nature 454: 955–960.