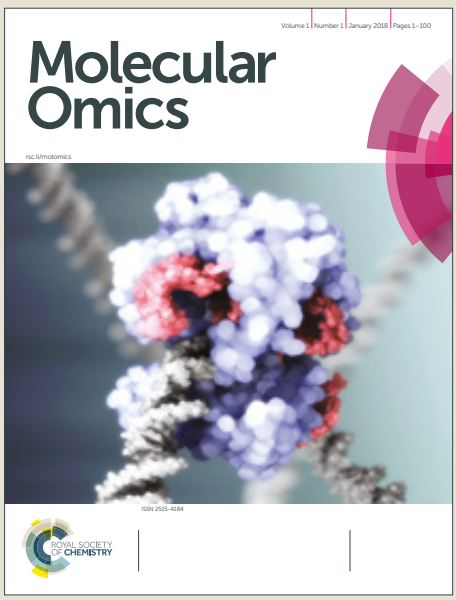
Molecular Omics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: E. Bencurova, S. K. Gupta, E. Oskoueian, M. Bhide and T. Dandekar, *Mol. Omics*, 2018, DOI: 10.1039/C8MO00130H.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/molomics

View Article Online
DOI: 10.1039/C8MO00130H

Omics and bioinformatics applied to vaccine development against *Borrelia*

Elena Bencurova*[#1], Shishir K. Gupta*[1], Ehsan Oskoueian[2], Mangesh Bhide[3,4], Thomas Dandekar[1,5#]

[1]Department of Bioinformatics, Biocenter, Am Hubland, D-97074 Wuerzburg, Germany

[2]Mashhad Branch, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education, and Extension Organization (AREEO), Mashhad, Iran

[3]University of veterinary medicine and pharmacy, Laboratory of biomedical microbiology and immunology, Komenskeho 73, 04181 Kosice, Slovakia

[4] Institute of Neuroimmunology, Slovak Academy of Sciences, Dubravska cesta 9, Bratislava.

[5] EMBL Heidelberg, BioComputing Unit, Meyerhofstr. 1, 69117 Heidelberg, Germany

*These authors contributed equally to this work

Corresponding authors: bencurova.elena@gmail.com, dandekar@biozentrum.uni-wuerzburg.de; Tel.: +49-931-3184551; Fax: +49-931-3184552

Keywords: *Borrelia*, network analysis, vaccine development, epitope prediction, computational vaccinology

Abstract

Borrelia burgdorferi is an extracellular spirochete that causes Lyme disease. Currently, no effective vaccine is available for human and animals except for dogs. In the present study, an extensive bioinformatics pipeline was established to predict new candidates that can serve for vaccine development including building the protein-protein interaction network based on orthologues of experimentally verified protein-protein interaction networks, elucidation of the proteins involved in the immune response, selection of the topologically-interesting proteins and their prioritization based on their antigenicity. Proteomic network analysis did yield an interactome network with 120 nodes with 97 interactions. Proteins were selected to obtain a subnet containing only the borrelial membrane proteins and immune-related host proteins.

View Article Online
DOI: 10.1039/C8MO00130H

This strategy resulted in the selection of 15 borrelial targets, which were subjected to the extensive bioinformatics analysis to predict their antigenic properties. Based on the strategy applied in this study the proteins encoded by *erpX* (ErpX proteins, UniProt ID: H7C7L6) and *erpL* (ErpL protein, UniProt ID: H7C7M3) and *erpY* (ErpY protein, UniProt ID: Q9S0D9) are suggested as novel set of vaccine targets to control Lyme disease. Moreover, five different tools were used to validate their antigenicity regarding B-cells. The combination of all these proteins in a vaccine should allow improved protection against *Borrelia* infection.

1. Introduction

Lyme disease (LD) is one of the most rapidly spreading vector-borne disease caused by spirochetal bacteria *Borrelia* that may cause serious infection in human and animal species. The *Borreliae* possess several mechanisms that enable them to invade the host and escape from the host immune system including downregulation of immunogenic surface proteins, degradation of extracellular matrix and utilization of the fibrinogenic system of host in order to invade tissue and persist in several hosts. Compared to other Gram-negative bacteria, borrelial outer membrane lacks the immunogenic surface glycolipid lipopolysaccharides (LPS). However, *Borrelia* express on their surface a variety of the low molecular outer surface proteins (Osp´s), which are necessary for the persistence of spirochete in various host (1, 2). Osp´s interact with several complement regulatory proteins (CRP) and protect *Borrelia* against complement-mediated killing (3, 4). It was reported earlier that *Borrelia* possess several proteins for the interaction with CRP factor H such as ErpA, ErpC, ErpP (5), CRASP-1 (3), OspE (6), FhbA (7) and others (Table 1). Apart from CRP-binding proteins, spirochetes express a variety of other surface proteins, which are able to interact with adaptive immune system (8, 9).

The severity of the disease, high treatment costs and significant increase of LD cases, result in a strong demand for vaccination as the most effective measure for prevention against LD. Nowadays, no licensed vaccine is available for immunization against Lyme disease in human (10). In the past, human vaccine LYMErix™ (SmithKline Beecham, Pittsburgh, PA, USA) was available, however, the low efficiency, late vaccination effect and insufficient protection against the full spectrum of the borrelial species withdrew the LYMErix™ from

View Article Online
DOI: 10.1039/C8MO00130H

the market (11). Currently, only the vaccines for dogs Recombitek® (Merial) and Novibac®Lyme (Merck Animal Health) are licensed and available for use in veterinary medicine.

Based on the slow progress in vaccine development against LD new strategies based on *in silico* methods were successfully implemented and advocated for the development of new vaccines (12-14). Although bioinformatics analysis of genetic, proteomic and biochemical data reveal valuable information about the interactions between *Borrelia* and its hosts, protein-protein analysis can be boosted further by novel bioinformatics-based approaches that allow analyzing the data in a high-throughput manner. In our new study, multiple databases and a homology-based pipeline were used in order to obtain a comprehensive view of all interactions between the borrelial and host proteins involved in immune defence processes.

To improve the search for putative vaccine candidates against *Borrelia*, we applied bioinformatics approaches after an extensive literature review to achieve a well-curated and detailed interactome network between *Borrelia burgdorferi* sensu lato (*B.b*.s.l) and its human host. The best candidates from the human-*Borrelia* network were further analyzed to identify optimal vaccine targets based on their antigenicity.

2. Methods

2.1 Building of an orthology-based network

To establish the human-*Borrelia* interaction network we used the stringent interolog based method (15, 16). The reference proteome of human (proteome ID: UP000005640, 70615 protein entries) and *B. burgdorferi* (strain ATCC 35210 / B31 / CIP 102532 / DSM 4680, proteome ID:UP000001807, 1290 protein entries) were retrieved from UniProt database (17). Experimentally verified PPIs of human and other mammalian hosts and *Borrelia* were collected from PUBMED, PHISTO, PATRIC and HPIDB databases (18-20) as well as from manual curation (Table 1) and used as a template to demarcate the orthology based human:*B. burgdorferi* interaction networks. All interolog calculations were performed with InParanoid 8 (21) and only the seed orthologs were considered. Interactions were constrained by presence

View Article Online
DOI: 10.1039/C8MO00130H

and absences of functional domains between the interacting proteins. To increase the confidence only the predictions were considered true in which the interacting pair contains at least one combination of truly interacting domain as mentioned in three domain interaction databases DOMINE, DIMA or IDDI (22-24). Domains in the interactors were annotated by PFAM database (25). The list of the appropriate protein-protein interactions was further filtered according to the localization of the protein using the SwissProt platform and only the membrane proteins of *Borrelia* were selected as targeted proteins for the novel vaccine candidates. Identification of non-homologous borrelial proteins was performed with BlastP analysis. The expectation value (e value) inclusion threshold was set to 0.001 with the minimum bit score of 100 and identity <30%. The functional importance of the proteins found was evaluated using sequence similarity to entries in the Database of Essential Genes (DEG) with following criteria: analysis BlastP (26) with an E-value cut-off of 1E-05, matrix BLOSUM62 and bitscore 100. Moreover, to complete the interaction map it was also extensively hand-curated taking data from databases and literature into account.

2.2 Gene Ontology (GO) analysis

All the protein sequences involved in host interactions were further analyzed concerning their function. GO analysis was performed by DAVID version 6.7 bioinformatics platform (27) and subsequently manually curated and categorized. Additionally, based on Gene Ontology (GO) annotations we extracted the interactions where protein sequence of human was marked as complement system or related protein.

2.3 Network analysis and visualization

The networks were graphically visualized using Network Analyzer plugin version 2.7 of Cytoscape version 2.8.1 (28). The number of the nodes and edges, as well as connectivity of nodes, were determined by graph theoretical analysis.

2.4 Screening for antigenicity properties of potential vaccine targets

View Article Online
DOI: 10.1039/C8MO00130H

In order to assess the most prospective vaccine targets based on their antigenicity, three different analysis were employed: a) identification of linear B-cell epitopes, b) estimation of the antigenic score, and c) BLAST-based filtering of human homologs. Four servers were employed for identification of epitopes. BCPREDS (29) was used to predict the B-cell epitopes with an epitope length of 15 and a specificity of 75% ; LBtope (30) with the identical conditions; Antibody Epitope Prediction at IEDB (31) using the length 15/16 amino acids and Bepipred Linear Epitope Prediction method, and; BcePred (32) was used for prediction of continuous B-cell epitopes using the default threshold and with the length at least 10 amino acids. The antigenicity of the most expected proteins was predicted using the VaxiJen program (http://www.jenner.ac.uk/VaxiJen) with a threshold cut-off of 0.5. The MHC-class I epitopes were also analyzed as shown in our previous work (33) using the matrices HLA DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*1101 and DRB1*1501. Subsequently, we performed a BlastP search of each epitope against human sequences with an E-value cut-off of 2E-05, matrix PAM30 and word size 2.

3 Results and Discussion

3.1 Identification of interactome network

Protein-protein interaction networks allow high-throughput protein screening in considerably less time compared to conventional experimental techniques and facilitate straightforward exploring of novel perspective vaccine and drug targets. Over the last decade, the data from the whole-genome sequencing platforms became more accessible and large datasets of proteins resulted in development of databases, such as STRING (34), Host-pathogen interaction database (20), Database of interacting proteins (35) and Microbial protein interaction (36). However, the disadvantage of those databases is the limited number of host and pathogen species, which include usually the model organism and most important medical pathogens, such as influenza virus, *Helicobacter pylori, Arabidopsis thaliana* and *Drosophila melanogaster* (35, 20, 36).

Using data mining and computational analysis based on the orthologous proteins of a man and *Borrelia* we generated a list of potential human:*Borrelia* interactions. Information

View Article Online
DOI: 10.1039/C8MO00130H

about the protein interactions was collected using literature data mining (Table 1) and the map of interactome was constructed (Fig. 1). The complete protein-protein interaction network consisted of 856 putative interaction pairs between 895 proteins, from which 666 belongs to human and 229 to borrelial proteins (Table S1). The proteins with most interactions among the borrelial proteins was DNA-directed RNA polymerase subunit beta (UniProt ID: Q59191, 45 interactions) followed by flagellin protein FliN (UniProt ID: Q44903, 18 interactions), Gln aminotransferase (UniProt ID: O51317) and CTP synthase (UniProt ID: O51522), both found in 14 interactions and thirteen interaction partners were identified for ParB (UniProt ID: O51395). Analysis of the human proteins showed that the NFKB (UniProt ID: P19838) was the proteins with the highest number of interaction (23followed by Stabilin-1 (UniProt ID: Q9NY15, 11 interactions) and Complement factor H (UniProt ID: P08603, 8 interactions), CD74 (UniProt ID: P04233, 7 interactions). Six interaction partners were observed also for Trinucleotide repeat-containing gene 6A protein (UniProt ID: Q8NDV7), STAT6 (UniProt ID: P42226) and β-2-microglobulin (UniProt ID: P61769).

3.2 Selection of potential borrelial vaccine targets by compartment and localization

Interactome networks included all interacting proteins, with random function and location. The proper selection and characterization of interactome members are crucial in order to find the vaccine candidates. Therefore, we prioritized only the elements of the interactome essential for adhesion and binding. Our reduction strategy followed two main criteria. One of the borrelial strategies for the survival in the bloodstream is interaction with the complement-related proteins or mimicking of those proteins to avoid or inhibit the complement-mediated lysis of extracellular pathogens. We were focused only on the proteins related to complement system or involved in the immune response, which were used to model the subnetwork of interactome. The second criterion in the selection of the most suitable candidates for vaccine development is their topographical distribution as only the membrane proteins are accessible for the host immune cells.

3.2.2 Characterization of the *H. sapiens-B. burgdorferi* subnetwork

The human subnetwork focused on the complement-related proteins consists of 97 interactions between 41 human and 79 borrelial proteins (Table S2). CRASP-1 (UniProt ID: O50957) encoded by *cspA* gene was predicted as a top target involved in the interactions with

View Article Online
DOI: 10.1039/C8MO00130H

complement factor H (UniProt ID: P08603), complement factor H-related proteins 1 and 2 (UniProt ID: Q03591 and V9GYE7), complement components C7 (UniProt ID: P10643), C8 beta chain (UniProt ID: P07358) and C9 (UniProt ID: P02748). Employment in two interactions was observed for most of the members of Erp-proteins, flagellar protein FlgE (UniProtID: Q447679, variable large protein (G5IXI6) and HflK (UniProtID: O51221). Proteins with a high degree of connection, however, located in the cytoplasm, were excluded from the further study. High numbers of interactions among human proteins were observed for NFKB (UniProt ID: P19838), Stab1 (UniProt ID: Q9NY15) and complement factor H (UniProt ID: P08603) (Table S1A) with the values of 23, 11 and 8 interactions, respectively (Table S2). As a strong candidate for vaccine development, members of the OspE/F-related protein family (Erp) seem to be good targets: Erp are borrelial lipoproteins. They were found to interact with the complement factor H and complement factor H-related proteins of various mammalian species (5), however, the regulation and exact function of these proteins remain unclear (37).

3.3 Distribution of candidate vaccine proteins according to GO function

Functional analysis of interacting proteins was performed by BLAST-GO annotation. From the seventy-nine spirochetal proteins in the subnetwork, only 48 have been annotated with biological function. The larger cluster contains proteins that participate in response to stimulus (GO: 0050896; 81% of annotated proteins), followed by a cluster of single-organism process proteins (GO:0044699; 79%), biological regulation (GO:0065007; 75%), metabolic processes (GO:0008152; 68%), immune system processes (GO:0002376; 68%), cellular processes (GO:0009987; 66%) and signaling (GO:0023052; 50%). In addition, proteins involved in other essential biological processes, such as developmental process, cellular component organization or biogenesis, biological adhesion, locomotion, and cell killing were detected (Figure 2, Table S3). This analysis suggests proteins involved in metabolic processes as the main group to consider as vaccine candidates given that blocking of those proteins will lead to the dysfunction of the enzymatic machinery, resulting in the inability of the bacteria to infect and replicate in the host. Human interacting partners were distributed mainly in the categories GO:0005488 (binding; 59% of annotated proteins), GO:0003824 (catalytic activity;

View Article Online
DOI: 10.1039/C8MO00130H

16%), GO:0071704 (receptor activity; 8%), and others, including protein binding transcription factor activity, receptor regulator activity and transporter activity (Supp. Figure 1).

3.4 Vaccine target prioritization

The prediction of antigenic epitopes is crucial for vaccine development; therefore, we focused only on proteins which are recognized by B-cells. The work of Drouin *et al*. showed that T-cells can be activated by borrelial proteins via cross-reactive epitopes during the infectious phase of Lyme disease and can be continuously stimulated by a homologous self-antigen during the post-infectious period. Furthermore, the cross-reactive T-cell can recognize both borrelial antigen and own peptides, what can be an investigate step in the development of autoimmune disease (38). The software we used to predict the B-cell epitopes relies mainly on the identification of the linear epitope sequence from the primary amino acids of the proteins in the context of their properties, such as hydrophilicity and solvent accessibility. Among all the protein candidates localized on the cytoplasmic or outer membrane of *Borrelia*, we selected 27 borrelial proteins based on their capability of interaction with host proteins (Table 2). Using the software for the prediction of linear epitopes, each of the proteins showed the presence of putative epitopes. For further analysis, only the epitopes that were found (also as overlapping sequences) in at least two of four epitopes tools were further analyzed (Table S4, shown in bold).

To identify the most antigenic proteins, we analyzed the whole sequence of each candidate with VaxiJen using a threshold cut-off of 0.5. From the 27 proteins, fifteen appear below the threshold line (Table 2). Three of them were assessed by the program as putative antigenic molecules (threshold above 0.4):, the “binding-protein-dependent transport systems inner membrane component domain protein” (UniProt ID: O51585), the L-lactate permease (UniProt ID: O51549) and in the Pts system, the fructose-specific iiabc component (UniProt ID: O51369), however, with only a low antigenic score. In addition, the BlastP screening of each epitope against the human sequences revealed any significant similarity (Table S5).

We found that the top three candidates with the best antigenicity score and connectivity levels (Table 3) belong to the Erp proteins encoded by *erp* genes located on extrachromosomal elements. These proteins are expressed in the early phase of LD and bind

View Article Online
DOI: 10.1039/C8MO00130H

mammalian factor H, plasmin/plasminogen, laminin, as well as other components of the extracellular matrix (5, 39). Binding to the extracellular matrix is a crucial event for *Borrelia*. For the dissemination in the host body, bacterium moves in the vascular system. However, here occur strong protective mechanisms that can lead to the death of the bacterium. Therefore, *Borrelia* is seeing for evasion from vascular tissue to invade the tissue with a low immune response. The group of *erp* genes is very diverse, resulting in the interaction with different host components and inhibition of complement-mediated lysis caused by the recruiting of the complement-regulatory proteins (40). *Borrelia* express Erp proteins in very small quantities at a low temperature of 23°C (the ambient temperature of the unfed tick). Switching the temperature when the tick starts feeding in warm-blood animal (23 to 34°C) significantly up-regulates the expression of Erp proteins (41). Temperature-dependent expression is probably one of the critical events in the natural infectious cycle of *Borrelia*. Variability of the surface antigens such as Erp and CRASP is one of the strategies to avoid the host protective mechanism and facilitate the development of the LD. The binding of the Erp protein to complement regulatory proteins depends on the amino acid sequence and glycolysation pattern of host proteins, which differ among the species. Activating of the immune defence depends on the several factors. One of them is complement resistance which depends on the binding affinity of borrelial proteins to factor H (42, 43). It was shown, that ErpA/I/N, ErpC and ErpP and ErpX bind the factor H in several groups of animals and humans. However, ErpL binds the factor H exclusively in cow, which is resistant to *Borrelia* infection (5). Although several studies in mammals showed the high immunogenic capacity of Erp proteins, they had never been investigated in detail for technical reasons (reviewed in (44).

The best candidate among the chosen proteins in our analysis is ErpX, a surface-exposed lipoprotein. ErpX binds to the complement factor H and factor H-like proteins as well as laminin, the component of the extracellular matrix (5, 39). The binding affinity of each of the Erp proteins to mammalian Factor H varies among species. The human, cat, dog and horse factor H have great affinity to the ErpX; while the rat, mouse and rabbit only exhibit weak affinity and, finally, cattle Factor H shows no binding affinity. The primary sequence is only approximately 20% identical to the primary sequence of other members of Erp proteins; however, they share the same sequence motifs and probably the function (5).

View Article Online
DOI: 10.1039/C8MO00130H

The second top candidate, lipoprotein ErpL, is a member of the OspF-related family, which binds heparin sulfate and plasminogen, as well as complement regulatory factor H. The heparin sulfate-mediated borrelial attachment leads to the adhesion of the glial, synovial and respiratory epithelial cells and promotes tissue colonization (45). Interestingly, it is the only Erp protein able to bind Factor H of cattle (5).

ErpY protein, encoded in the cp32 prophage family of plasmid-like elements, had been also shown as a binding partner to factor H from rodents, dogs and cats (5). The *erpY* gene is consistently expressed in the infected tissue even after a long time, but in lower levels compared to *erpL* (46).

A high antigenicity score was observed also for the Variable large protein (Vlp), the crucial player in evasion mechanisms of host immune response. It occurs typically during the mammalian host infection, interestingly, in alternation with the Variable small protein (Vsp). Both lipoproteins Vlp and Vsp are encoded by the same expression plasmid, but they are expressed in different cycles of infection. When the host immune system recognizes the antigen and starts to produce Vlp antibody, *Borrelia* switch the expression and instead of the Vlp, produce the Vsp protein. The antigenic variation mechanism is the best reported in the species caused the relapsing fever (47).

The infection-associated membrane lipoproteins OspA and OspC were the most popular targets for vaccine development (587 research articles for OspA, 158 for OspC and 116 research articles for OspA/OspC related to vaccine development until February 2018, PubMed). However, they do not belong to the top-three candidates, as they are interacting only with one interacting partner among the immune-related host proteins. However, both of those proteins are strong antigens according to the results of the antigenicity validation. To get a complete view of the selected candidates, we analyzed them in order to search for their homologues in the host under BlastP analysis. We found that only two of them have to be restricted from the list of potential candidates: OppD (UniProt ID: H7C7Q3) and ABC transporter, ATP-binding protein (UniProt ID: O51695), as both possess orthologues in human proteome.

View Article Online
DOI: 10.1039/C8MO00130H

Concerning overall results obtained in this study, we suggest additionally to the Erp proteins (ErpX, ErpL, ErpY) also the proteins encoded by *bb_F0041* (Variable large protein, UniProt ID: G5IXI6) as a novel target for vaccine development against *Borrelia*. These Erp proteins and Vlp were validated by the interactome and ortholog reconstruction, together with their compartment localization and high B-cell antigenicity and high antigenicity score make them good vaccine candidates.

4 Conclusion

The present study suggests four proteins of *B. burgdorferi* that can be promising candidates for new vaccine research. Development of vaccines against *Borrelia* for human is challenging and a long-term process, which at present did not lead to satisfactory protection against LD. We performed in this study extensive bioinformatics studies to get broadly new candidates that can serve in vaccine developments including building the protein-protein interaction network based on the orthologues of experimentally verified PPI, elucidation of the proteins involved in the immune response (e.g. interaction with complement-regulatory proteins), selection of topologically-interesting membrane proteins and their prioritization based on the antigenicity. To avoid harmful effects on the host, homologous proteins were removed from the list of suggestions. Extracellular pathogens use several strategies to evade the complement attack. *Borrelia* developed several strategies, mainly addressed to the binding to CRPs, to survive in the bloodstream. Our filtering method resulted in identification of 15 proteins interacting with immune proteins. Among these proteins, four are suggested for further studies to test and probe their potential as novel vaccine candidates being lipoproteins and pathogenicity factors, highly exposed, in the right compartment and with high antigenicity.

Declarations of interest: none.

Acknowledgements: This research was funded by Project Grants Deutsche Forschungsgemeinschaft [grant number TR124/B2]; Agentúra na podporu výskumu a vývoja [grant number APVV-14-0218]; and Vedecká grantová agentúra Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied [grant number VEGA 1/0258/15]. We thank Daniel Villalobos for grammar corrections.

View Article Online
DOI: 10.1039/C8MO00130H

321

View Article Online
DOI: 10.1039/C8MO00130H

References

1. Takayama K, Rothenberg RJ, Barbour AG. Absence of lipopolysaccharide in the Lyme disease spirochete, Borrelia burgdorferi. Infection and immunity. 1987;55(9):2311-3.
2. Pulzova L, Bhide M. Outer surface proteins of Borrelia: peerless immune evasion tools. Current protein & peptide science. 2014;15(1):75-88.
3. Bhide MR, Escudero R, Camafeita E, Gil H, Jado I, Anda P. Complement factor H binding by different Lyme disease and relapsing fever Borrelia in animals and human. BMC research notes. 2009;2:134.
4. Madar M, Bencurova E, Mlynarcik P, Almeida AM, Soares R, Bhide K, et al. Exploitation of complement regulatory proteins by Borrelia and Francisella. Molecular bioSystems. 2015;11(6):1684-95.
5. Stevenson B, El-Hage N, Hines MA, Miller JC, Babb K. Differential binding of host complement inhibitor factor H by Borrelia burgdorferi Erp surface proteins: a possible mechanism underlying the expansive host range of Lyme disease spirochetes. Infection and immunity. 2002;70(2):491-7.
6. Hovis KM, Tran E, Sundy CM, Buckles E, McDowell JV, Marconi RT. Selective binding of Borrelia burgdorferi OspE paralogs to factor H and serum proteins from diverse animals: possible expansion of the role of OspE in Lyme disease pathogenesis. Infection and immunity. 2006;74(3):1967-72.
7. Hovis KM, McDowell JV, Griffin L, Marconi RT. Identification and characterization of a linear-plasmid-encoded factor H-binding protein (FhbA) of the relapsing fever spirochete Borrelia hermsii. Journal of bacteriology. 2004;186(9):2612-8.
8. Rogovskyy AS, Bankhead T. Variable VlsE is critical for host reinfection by the Lyme disease spirochete. PloS one. 2013;8(4):e61226.
9. Schnarr S, Franz JK, Krause A, Zeidler H. Lyme borreliosis. Best Practice & Research Clinical Rheumatology. 2006;20(6):1099-118.
10. Kaaijk P, Luytjes W. Vaccination against Lyme disease: Are we ready for it? Human vaccines & immunotherapeutics. 2016;12(3):757-62.
11. Nigrovic LE, Thompson KM. The Lyme vaccine: a cautionary tale. Epidemiol Infect. 2007;135(1):1-8.
12. Chawley P, Samal HB, Prava J, Suar M, Mahapatra RK. Comparative genomics study for identification of drug and vaccine targets in Vibrio cholerae: MurA ligase as a case study. Genomics. 2014;103(1):83-93.
13. Khan AM, Hu Y, Miotto O, Thevasagayam NM, Sukumaran R, Abd Raman HS, et al. Analysis of viral diversity for vaccine target discovery. BMC Med Genomics. 2017;10(Suppl 4):78.
14. Birhanu BT, Lee SJ, Park NH, Song JB, Park SC. In silico analysis of putative drug and vaccine targets of the metabolic pathways of Actinobacillus pleuropneumoniae using a subtractive/comparative genomics approach. J Vet Sci. 2018;19(2):188-99.
15. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome research. 2004;14(6):1107-18.
16. Zhou H, Gao S, Nguyen NN, Fan M, Jin J, Liu B, et al. Stringent homology-based prediction of H. sapiens-M. tuberculosis H37Rv protein-protein interactions. Biology direct. 2014;9:5.
17. UniProt C. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43(Database issue):D204-12.
18. Durmus Tekir S, Cakir T, Ardic E, Sayilirbas AS, Konuk G, Konuk M, et al. PHISTO: pathogen-host interaction search tool. Bioinformatics. 2013;29(10):1357-8.

View Article Online
DOI: 10.1039/C8MO00130H

19. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic acids research. 2014;42(Database issue):D581-91.
20. Kumar R, Nanduri B. HPIDB--a unified resource for host-pathogen interactions. BMC bioinformatics. 2010;11 Suppl 6:S16.
21. Sonnhammer EL, Ostlund G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic acids research. 2015;43(Database issue):D234-9.
22. Yellaboina S, Tasneem A, Zaykin DV, Raghavachari B, Jothi R. DOMINE: a comprehensive collection of known and predicted domain-domain interactions. Nucleic acids research. 2011;39(Database issue):D730-5.
23. Luo Q, Pagel P, Vilne B, Frishman D. DIMA 3.0: Domain Interaction Map. Nucleic acids research. 2011;39(Database issue):D724-9.
24. Kim Y, Min B, Yi GS. IDDI: integrated domain-domain interaction and protein interaction analysis system. Proteome science. 2012;10 Suppl 1:S9.
25. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1):D279-85.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.
27. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.
28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.
29. El-Manzalawy Y, Dobbs D, Honavar V. Predicting flexible length linear B-cell epitopes. Comput Syst Bioinformatics Conf. 2008;7:121-32.
30. Singh H, Ansari HR, Raghava GP. Improved method for linear B-cell epitope prediction using antigen's primary sequence. PloS one. 2013;8(5):e62216.
31. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2:2.
32. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65(1):40-8.
33. Gupta SK, Srivastava M, Akhoon BA, Smita S, Schmitz U, Wolkenhauer O, et al. Identification of immunogenic consensus T-cell epitopes in globally distributed influenza-A H1N1 neuraminidase. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2011;11(2):308-19.
34. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research. 2011;39(Database issue):D561-8.
35. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research. 2002;30(1):303-5.
36. Goll J, Rajagopala SV, Shiau SC, Wu H, Lamb BT, Uetz P. MPIDB: the microbial protein interaction database. Bioinformatics. 2008;24(15):1743-4.
37. Embers ME. The Pathogenic Spirochetes: strategies for evasion of host immunity and persistence: Springer Science & Business Media; 2012.
38. Drouin EE, Glickstein L, Kwok WW, Nepom GT, Steere AC. Human homologues of a Borrelia T cell epitope associated with antibiotic-refractory Lyme arthritis. Mol Immunol. 2008;45(1):180-9.
39. Brissette CA, Verma A, Bowman A, Cooley AE, Stevenson B. The Borrelia burgdorferi outer-surface protein ErpX binds mammalian laminin. Microbiology. 2009;155(Pt 3):863-72.

View Article Online
DOI: 10.1039/C8MO00130H

40. Kraiczy P, Hellwage J, Skerka C, Kirschfink M, Brade V, Zipfel PF, et al. Immune evasion of Borrelia burgdorferi: mapping of a complement-inhibitor factor H-binding site of BbCRASP-3, a novel member of the Erp protein family. European journal of immunology. 2003;33(3):697-707.
41. Miller JC, von Lackum K, Babb K, McAlister JD, Stevenson B. Temporal analysis of Borrelia burgdorferi Erp protein expression throughout the mammal-tick infectious cycle. Infect Immun. 2003;71(12):6943-52.
42. Bhide MR, Travnicek M, Levkutova M, Curlik J, Revajova V, Levkut M. Sensitivity of Borrelia genospecies to serum complement from different animals and human: a host-pathogen relationship. FEMS Immunol Med Microbiol. 2005;43(2):165-72.
43. Kisova-Vargova L, Mucha R, Cernanska D, Bhide M. Host-dependent differential expression of factor H binding proteins, their affinity to factor H and complement evasion by Lyme and relapsing fever borreliae. Vet Microbiol. 2011;148(2-4):341-7.
44. Skotarczak B. Why is There Still no Human Vaccine Against Lyme Borreliosis? Folia Biol-Krakow. 2015;63(3):159-65.
45. Lin YP, Bhowmick R, Coburn J, Leong JM. Host cell heparan sulfate glycosaminoglycans are ligands for OspF-related proteins of the Lyme disease spirochete. Cell Microbiol. 2015;17(10):1464-76.
46. Miller JC, Narayan K, Stevenson B, Pachner AR. Expression of Borrelia burgdorferi erp genes during infection of non-human primates. Microbial pathogenesis. 2005;39(1-2):27-33.
47. Norris SJ. Antigenic variation with a twist--the Borrelia story. Molecular microbiology. 2006;60(6):1319-22.
48. Garcia BL, Zhi H, Wager B, Hook M, Skare JT. Borrelia burgdorferi BBK32 Inhibits the Classical Pathway by Blocking Activation of the C1 Complement Complex. PLoS pathogens. 2016;12(1):e1005404.
49. Meri T, Cutler SJ, Blom AM, Meri S, Jokiranta TS. Relapsing fever spirochetes Borrelia recurrentis and B. duttonii acquire complement regulators C4b-binding protein and factor H. Infection and immunity. 2006;74(7):4157-63.
50. Hammerschmidt C, Klevenhaus Y, Koenigs A, Hallstrom T, Fingerle V, Skerka C, et al. BGA66 and BGA71 facilitate complement resistance of Borrelia bavariensis by inhibiting assembly of the membrane attack complex. Molecular microbiology. 2016;99(2):407-24.
51. Siegel C, Hallstrom T, Skerka C, Eberhardt H, Uzonyi B, Beckhaus T, et al. Complement factor H-related proteins CFHR2 and CFHR5 represent novel ligands for the infection-associated CRASP proteins of Borrelia burgdorferi. PloS one. 2010;5(10):e13519.
52. Schott M, Grosskinsky S, Brenner C, Kraiczy P, Wallich R. Molecular characterization of the interaction of Borrelia parkeri and Borrelia turicatae with human complement regulators. Infection and immunity. 2010;78(5):2199-208.
53. Rossmann E, Kraiczy P, Herzberger P, Skerka C, Kirschfink M, Simon MM, et al. Dual binding specificity of a Borrelia hermsii-associated complement regulator-acquiring surface protein for factor H and plasminogen discloses a putative virulence factor of relapsing fever spirochetes. Journal of immunology. 2007;178(11):7292-301.
54. Hovis KM, Jones JP, Sadlon T, Raval G, Gordon DL, Marconi RT. Molecular analyses of the interaction of Borrelia hermsii FhbA with the complement regulatory proteins factor H and factor H-like protein 1. Infection and immunity. 2006;74(4):2007-14.
55. Kraiczy P, Skerka C, Brade V, Zipfel PF. Further characterization of complement regulator-acquiring surface proteins of Borrelia burgdorferi. Infection and immunity. 2001;69(12):7800-9.
56. Dyer A, Brown G, Stejskal L, Laity PR, Bingham RJ. The Borrelia afzelii outer membrane protein BAPKO_0422 binds human factor-H and is predicted to form a membrane-spanning beta-barrel. Bioscience reports. 2015;35(4).

View Article Online
DOI: 10.1039/C8MO00130H

465

466

View Article Online
DOI: 10.1039/C8MO00130H

Table 1: The protein-protein interaction network between complement-related proteins and *Borrelia burgdorferi* sensu lato[1]

Protein of *Borrelia*	Pathogen	Host protein	Host	Reference
BBK32	*B. burgdorferi*	C1, C1r	human	(48)
App. 40 kDa, app. 30 kDa: *e.g.* gi|426202330 gi|106534293 gi|1507705 gi|75344820 gi|1507653 gi|106534293	*B. valaisiana, B. lusitaniae, B. bissetti, B. japonica, B. afzelii, B. burgdorferi, B. bavariensis, B. garinii, B. hermsi*	C4Bp	human	(4)
App. 140 kDa	*B. corriaceae*	C4Bp	human	(4)
gi|239835984	*B. burgdorferi*	C4Bp	sheep	(4)
N/A	*B. recurrentis, B. duttoni,B. burgdorferi*	C4Bp	human	(49)
BGA66, BGA71	*B. bavariensis*	complement components C7, C8 and C9	human	(50)
CRASP-3 CRASP-5	*B. burgdorferi*	Factor H – related protein (CFHR)1, CFHR2, CFHR5	human	(51)
BpcA	*B. parkeri*	Factor H CFHR 1	human	(52)
BhCRASP-1	*B. hermsii*	Factor H CFHR 1	human	(53)
OspE	*B. burgdorferi*	Factor H CFHR 1	human	(6)
FhbA	*B. hermsii*	Factor H	human	(7)
App. 17-26 kDa *e.g.* gi|488741558 gi|218964081 gi|254952778 gi|108796607	*B. afzelii, B. burgdorferi, B. garinii, B. valaisiana, B. andersoni, B. japonica, B. hermsi, B. parkeri, B. anserina*	Factor H	human	(4)
gi|254952778 gi|51707666	*B. burgdorferi*	Factor H	sheep	(4)

View Article Online
DOI: 10.1039/C8MO00130H

	B. bissetti			
gi|195942064	*B. coriaceae*	Factor H	cattle	(4)
FhbA	*B. hermsii*,	Factor H CFHR 1	human	(54)
N/A	*B. recurrentis, B. duttoni,B. burgdorferi*	Factor H	human	(49)
OspE-related lipoprotein	*B. japonica*	Factor H	human, mouse	(3)
CRASP-1 outer surface protein	*B. afzelii,*	Factor H	human, mouse	(3)
Complement regulator-acquiring surface protein 1	*B. burgdorferi*	Factor H	human	(3)
N/A	*B. garinii, B. valaisiana, B. andersoni, B. bissetti, B. hermsi, B. parkeri, B. anserina*	Factor H	human	(3)
Membrane protein (WP_011703765)	*B. afzelii*	Factor H	mouse	(3)
N/A	*B. bissetti, B. japonica, B. coriaceae, B. burgdorferi, B. andersoni, B. hermsi*	Factor H	mouse	(3)
15-58 kDa proteins	*B. hermsi, B. coriaceae*	Factor H	rat	(3)
20 kDa protein	*B. hermsi*	Factor H	guinea pig	(3)
40 and 58 kDa proteins	*B. coriaceae*	Factor H	cow	(3)
15-20 kDa proteins	*B. afzelii, B. valaisiana, B. andersoni, B. japonica*	Factor H	dog	(3)
15 and 22 kDa protens	*B. afzelii, B.*	Factor H	cat	(3)

View Article Online
DOI: 10.1039/C8MO00130H

	japonica			
ErpA/I/N, ErpC, ErpP, ErpX	*B. burgdorferi*	Factor H	human, horse	(5)
ErpA/I/N, ErpC, ErpP, ErpX, ErpY	*B. burgdorferi*	Factor H	rat, mouse, rabbit	(5)
ErpA/I/N, ErpC, EpcG, ErpL, ErpP	*B. burgdorferi*	Factor H	cattle	(5)
ErpA/I/N, ErpC, ErpG, ErpP, ErpX, ErpY	*B. burgdorferi*	Factor H	dog, cat	(5)
BaCRASP-4 BaCRASP-5	*B. afzelii*	Factor H	human	(55)
BbCRASP-3 BbCRASP-4 BbCRASP-5	*B. burgdorferi*	Factor H	human	(55)
BAPKO_0422	*B. afzelii*	Factor H	human	(56)
BaCRASP-1 BbCRASP-2 BbCRASP1	*B. burgdorferi*	CFHR1, reconectin	human	(55)
30, 55 and 130 kDa proteins	*B. garinii* *B. afzelii* *B. valaisiana* *B. andersoni* *B. anserina*	Vitronectin	human	(4)

[1]after bioinformatics analysis and detailed manual curation

N/A – the interacting proteins were not characterized

View Article Online
DOI: 10.1039/C8MO00130H

Table 2: The protein interaction network between complement-related proteins and *Borrelia burgdorferi sensu lato* [1]

UniProt Acces. Nr.	Protein (*Borrelia*)	Degree	Localization	Interacting partners (human)	Homology with human	DEG[2]	VaxiJen score	Probable antigen
O51369	Pts system, fructose-specific iiabc component	8	Cytoplasmic Membrane	E1BAK9, E1BDD9, F1MYZ3, F1N0R5, P01888, Q0VCJ2, Q3SZA6, Q862A9	N	3	0.4046	yes
Q44767	FlgE	8	Extracellular	A0JN51, E1BDD9, F1MYZ3, F1N0R5, P01888, Q0VCJ2, Q3SZA6, Q862A9	N	0	0.5288	yes
O51039	UPF0118 membrane protein BB_0006	7	Cytoplasmic Membrane	A1A4L7, E1BDD9, F1MYZ3, F1N0R5, P01888, Q3SZA6, Q862A9	N	2	0.7313	yes
P70859	HAP1	5	Extracellular	Q2KI50, E1BAL5, FIMB08, FIMKW9, F1N1T8	N	0	0.3499	no
O50957	CRASP-1	4	Extracellular	F1N102, Q28085, O50957, Q3MHN2	N	0	0.5800	yes
O51550	Serine-type D-Ala-D-Ala carboxypeptidase	4	Cytoplasmic Membrane	E1BEJ5, F1MVJ8, P26201, Q9TU03	N	2	0.3068	no
O51310	OppC	3	Cytoplasmic Membrane	F1MAZ0, F1MDU9, P01888	N	11	0.3648	no
O51272	FtsK	3	Cytoplasmic Membrane	G3N026, Q56JW7, Q9XS49	N	23	0.5404	yes
O51278	Signal peptidase I	3	Membrane	A6QLZ0, E1BNJ1, G5E5X0	N	1	0.3314	no
O51161	Penicillin-binding protein	3	Membrane	A6H767, E1BM52, Q95KV1	N	47	0.3744	no
G5IXI6	VlsE1	2	Membrane	Q28065, Q3ZBS7	N	0	0.6848	yes
O50835	BBK32	2	Membrane	A5D9E9, F1MJ12	N	0	O50835	yes
O51549	L-lactate permease	2	Cytoplasmic Membrane	E1BGB0, P01888	N	1	0.4754	yes
O51754	Sodium/pantothenate symporter	2	Cytoplasmic Membrane	F1MKW9, P01966	N	1	0.7552	yes

View Article Online
DOI: 10.1039/C8MO00130H

H7C7Q3	OppD	2	Cytoplasmic Membrane	F1MKW9, Q3MHI2	Y	100	0.3997	no
Q9S0D9	ErpY	1	Extracellular	Q28085	N	0	0.8473	yes
Q9S036	ErpP	1	Extracellular	Q28085	N	0	0.6611	yes
O50889	Antigen, P35	1	Extracellular	Q28085	N	0	0.6404	yes
H7C7N5	ErpA8	1	Extracellular	Q28085	N	0	0.6246	yes
H7C7M3	ErpL	1	Extracellular	Q28085	N	0	0.8833	yes
H7C7L6	ErpX	1	Extracellular	Q28085	N	0	0.9108	yes
Q07337	OspC	1	Extracellular	Q28065	N	0	0.6518	yes
P0CL66	OspA	1	Extracellular	Q28065	N	0	0.8151	yes
O51585	Binding-protein-dependent transport systems inner membrane component domain protein	1	Cytoplasmic Membrane	E1BNJ1	N	5	0.4953	yes
O51735	Outer membrane protein	1	Extracellular	F1MKW9	N	21	0.3901	no
O51695	ABC transporter, ATP-binding protein	1	Cytoplasmic Membrane	F1MKW9	Y	100	0.1688	no

[1]after detailed bioinformatics and computational analysis

[2]DEG – Number of homologs found in DEG, N – no homology, Y – homologous proteins found

View Article Online
DOI: 10.1039/C8MO00130H

Table 3. Summary of selected target proteins.

Gene	Protein	UniProt ID	MW	Function	Compartment	Expression in infection cycle	Score B-cell /VaxiJen*	MHC I - class epitopes**
erpX	ErpX protein	H7C7L6	39.8 kDa	Contributes in invasion and long-term colonization of host tissues, laminin binding	Outer cell membrane	Only in the mammalian host, unexpressed in the tick host	1/0.9108	12
erpL	ErpL protein	H7C7M3	26.1 kDa	Tissue colonization	Outer cell membrane	Strongly in mammalian host in the early phase of LD	1/0.8833	3
erpY	ErpY protein	Q9S0D9	25.4 kDa	Unknown	Outer cell membrane	Stronger expressed by low temperature, expressed also in mammalian host	1/0.8473	5
BB_F0041	Vlp protein (syn. VlsE1, BB_F0041)	G5IXI6	36.2 kDa	Involved in pathogenesis (antigenic variation and tissue tropism)	Outer cell membrane	Expressed in each mammalian and tick host, strongly expressed in the early phase of LD	1/0.6848	7
erpP	ErpP protein (syn.	Q9S036	20,6 kDa	Binding multiple serum and	Outer cell membrane	Only in the mammalian host, unexpressed in	1/0.6611	2

View Article Online
DOI: 10.1039/C8MO00130H

	CRASP-3)			tissue components		the tick host		

MW: molecular weight

*Score B-cell is the score of the most antigenic epitope of the protein generated by FBCpres.

** MHC class I alleles distribution for antigenic epitopes was generated according (33) (Supp. Figure S2)

View Article Online
DOI: 10.1039/C8MO00130H

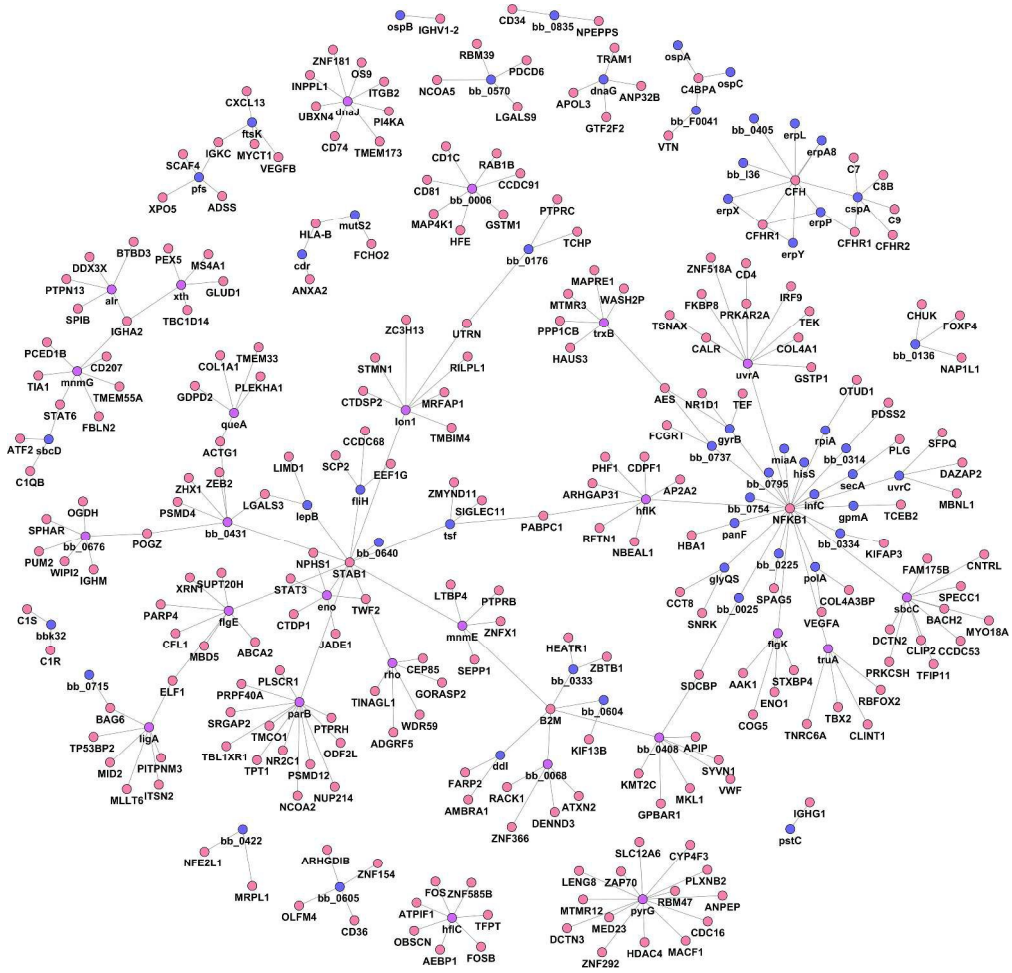
Figure legends

Figure 1: A map of the host-pathogen protein-protein interaction (PPI) network between human and *B. burgdorferi*. The map is based on interologs of eukaryotic-bacteria PPIs and literature information. The color of the nodes indicates different information; Color code: blue = *B. burgdorferi* proteins, purple = *B. burgdorferi* proteins with >=5 host interacting proteins, pink = human immune-related proteins.

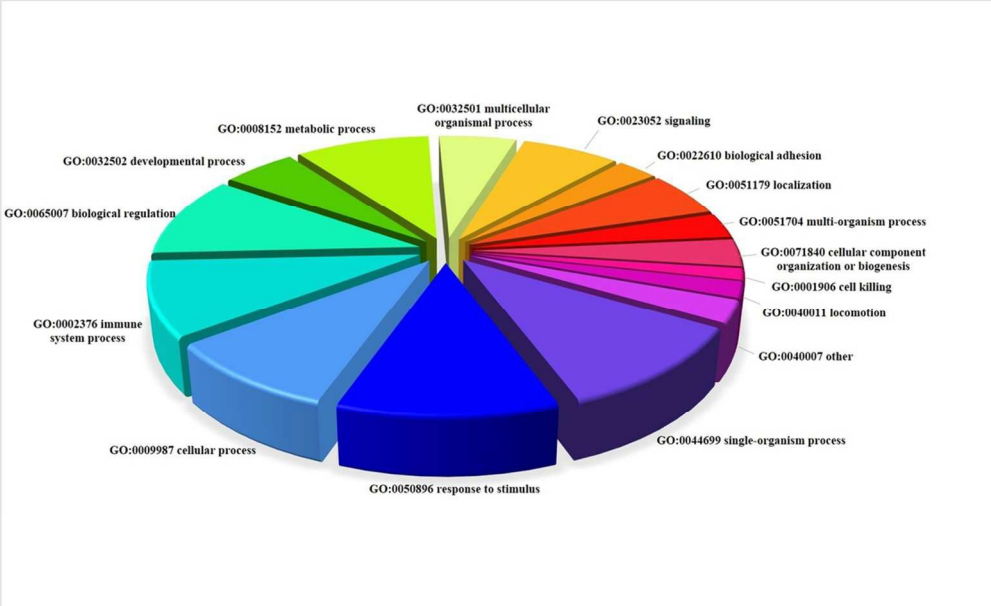
Figure 2: Distribution of interacting borrelial proteins according to GO annotation. Pie chart representations of the distribution of identified 48 borrelial proteins interacting with human proteins according to their molecular functions.

Figure 3: Comparision of immune-related proteins subnetworks. Same proteins from pathogen and the orthologs of human proteins are oriented at the same place in the networks. Color code: blue = *B. burgdorferi* proteins, purple = *B. burgdorferi* proteins with >=5 host interacting proteins, pink = human immune-related proteins. The proteins higlighed in black are the proteins suggested as the novel protein targets based on their antigenicity, molecular characteristics and interactome data.

View Article Online
DOI: 10.1039/C8MO00130H

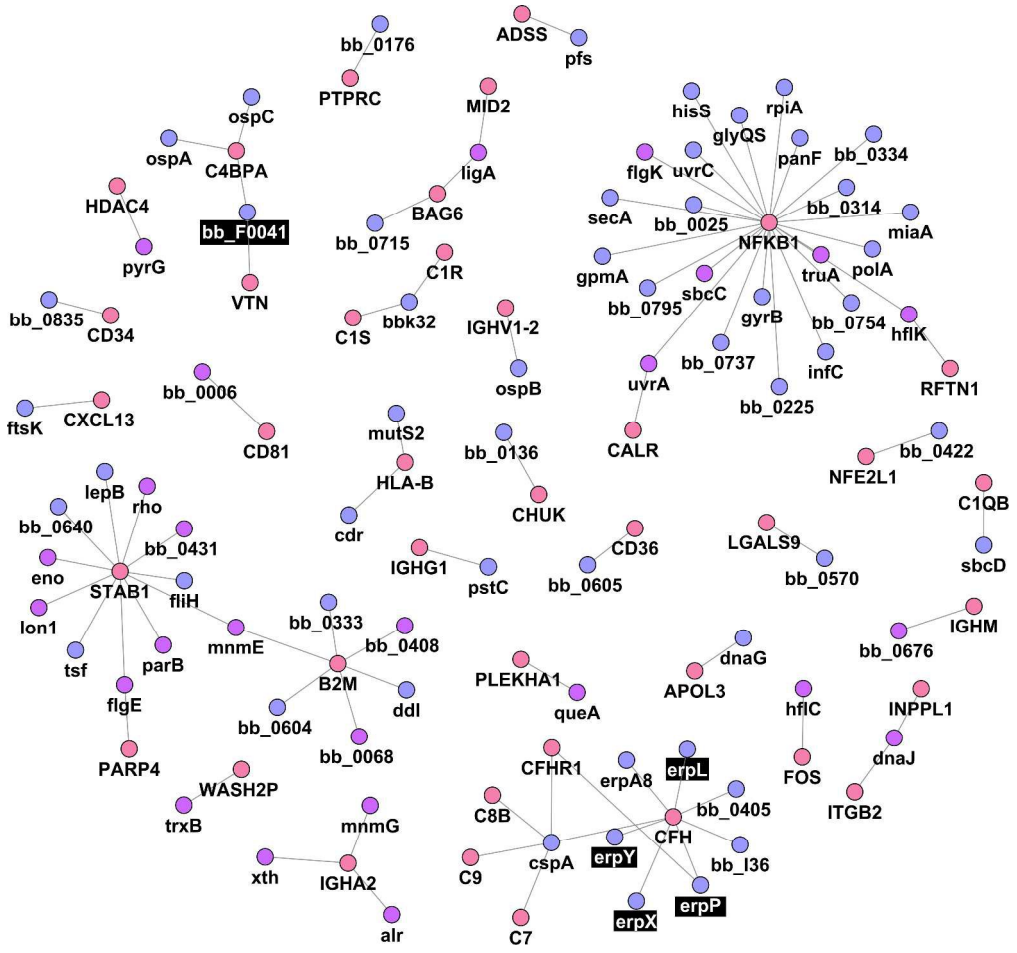


View Article Online
DOI: 10.1039/C8MO00130H



104x64mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C8MO00130H



1466x1375mm (96 x 96 DPI)