

## Prediction of Virulence Factors Using Bioinformatics Approaches

Rupanjali Chaudhuri and Srinivasan Ramachandran

### Abstract

Virulence factors produced by a pathogen are essential for causing disease in the host. They enable the pathogen to establish itself within the host thus enhancing its potential to cause disease and in some instances underlie evasion of host defense mechanisms. Identification of these molecules, especially those of immunological interest and their use in vaccine development are attractive and are among the initial steps of reverse vaccinology. Surface localized virulence factors such as adhesins serve as excellent immunogenic candidates in this regard. In this chapter we have described the bioinformatics approaches for adhesin prediction, which include specific adhesin prediction algorithms.

**Key words** Virulence factors, Host, Pathogen, Immunogenic, Adhesins, Vaccine

---

### 1 Introduction

Despite advances in technologies to combat infections, infectious diseases continue to challenge humans. This may be attributed to the rise in drug-resistant strains of pathogens such as *Mycobacterium tuberculosis* and new emerging infectious pathogens such as SARS coronavirus and influenza virus. A key step in the establishment of infectious disease is microbial virulence, which has been described as an emergent property of host–microbe interaction [1]. At the molecular level, entities like proteins, carbohydrates, or lipids enable the pathogens to establish themselves in a susceptible host. These molecules form inherent part of the pathogen cellular system and are collectively termed “virulence factors” [2, 3]. Virulence factors in various pathogens play diverse roles in the establishment of disease. These include colonization of the host, evasion of host defense mechanisms, immunosuppression, acquisition of nutrients from host cell, mediation of entry and exit into host cell in intracellular pathogens, and sensing change of environment [4, 5]. These factors enable colonization of host niche and eventually cause damage to host tissues [2, 4, 6].

It was therefore realized that targeting these microbial molecules by identifying their immunogenicity and use in vaccine formulations could serve as efficient anti-infective strategy. Vaccinologists are therefore preparing vaccine formulations with these molecules for priming the immune system in order to neutralize their activity in the event of a host–pathogen contact [5, 7].

A diverse array of molecules is involved during host–pathogen interaction and the prominent players vary between the pathogens. These include adhesins, toxins, enzymes, and capsules (polysaccharides or polypeptides).

Adhesins have attracted interest from immunological perspective because they are located on the cell surface and are likely to be accessible to the molecules of the immune system [8]. In the subsequent sections we provide an overview of these molecules and describe their prediction using Bioinformatics.

### 1.1 Adhesins

Adhesins enable adherence of the pathogen to host cells and constitutes the initial major step in the process of infection. This role of adhesins qualifies them for vaccine candidates as targeting adhesins could arrest infection at the initial stage [8]. Even though adhesins exhibit sequence polymorphisms, the conserved regions may serve for potential vaccine especially those containing receptor binding domain [9]. Recently, a potent combination of adhesins of *Plasmodium falciparum* has been identified, which could transcend strain variations [10].

Examples include FimH adhesin of uropathogenic *Escherichia coli*. Vaccination with this protein proved effective against urinary tract infection caused by *E. coli* in both mice and in nonhuman primates [11]. Filamentous hemagglutinin (FHA) and pertactin adhesins of gram-negative bacteria *Bordetella pertussis* elicits long-lasting cell mediated respiratory immune response [12]. These adhesins are components of the approved acellular pertussis licensed vaccine [13]. Another adhesin *Neisseria meningitidis* adhesin A (NadA) is part of a multicomponent meningococcal serogroup B vaccine named Bexero, which is capable of eliciting a robust immune response. This vaccine has cleared all clinical trials and awaiting license for use [14].

---

## 2 Materials and Methods

### 2.1 Bioinformatics Approaches of Adhesin Characterization

The advent of genomics technologies has revolutionized biological research. The complete genome sequence of a pathogen provides an abundance of opportunities to identify putative virulence factors through sequence analysis. These investigations are being aided by the development of new computational algorithms in this area.

In the sections below, we discuss and outline the methods used in several investigations:

1. Sequence Similarity Search: Sequence similarity search is very popular and is among the first to be applied in sequence analysis. The goal here is to obtain orthologous sequences corresponding to a given query. This approach has been used to identify orthologues of known adhesins characterized in other pathogens (*see Note 1*). The best known algorithm is the Basic Local Alignment Search Tool (BLAST) algorithm [15]. Examples include application of BLAST algorithms in screening for potential adhesins in *Mycoplasma agalactiae*, *Escherichia coli*, *Mycoplasma conjunctivae*, *Mycoplasma pneumonia*, *Rickettsial species* [16–21]. In addition BLAST can be used to identify orthologues of enzymes from pathogens involved in virulence: Hyaluronidase, Neuraminidase, Phospholipases, Proteases, Collagenase, Kinase, Coagulase, Leukocidins, Hemolysins.

2. Sequence Motif search: Sequence motif refers to a particular arrangement or pattern of amino acids within a protein sequence, or nucleotides within a DNA sequence, which is characteristic of a specific biochemical function [22]. In particular, majority of protein sequence motifs, provide unique detectable sequence features for a set of protein sequences and thus act as signatures of protein families. Such motifs indicate similar functional roles.

For example, in fungi, many Glycosylphosphatidylinositol-modified (GPI) proteins linked to plasma membrane via pre-formed GPI anchor play role in adhesion and virulence [23, 24]. These proteins have C-terminal GPI-motif described as follows: “[GNSDAC]-[GASVIETKDLF]-[GASV]-X(4,19)-[FILMVAGPSTCYWN](10)>” in Prosite format, where “>” indicates the C-terminal end of the protein [26]. Algorithm based on identifying sequences having a C-terminal, fungus-specific, consensus sequence for GPI modification (GPI-motif) helps screen a set of potential fungal adhesins [25]. Table 1 lists the motifs identified in several adhesins.

3. Signal Peptide: Signal Peptide (SP) is a short stretch of sequence present in the N-terminus of the protein directing it to the secretory pathway [31]. Adhesins being membrane attached proteins usually possess N-terminal signal peptide for translocation across the membrane of the endoplasmic reticulum [32, 33]. Therefore, algorithms using this information to screen for proteins having N-terminal signal peptide may help identifying potential adhesins (*see Note 2*). However, there are adhesins called “anchorless adhesins,” which do not have Signal peptide or Transmembrane domain. These “anchorless adhesins” cannot be identified through these approaches.

**Table 1**  
**Motifs in adhesins and other virulence factors**

Motif	Description	Reference
Beta-helix motif	These are right-handed parallel beta-helix supersecondary structural motifs in primary amino acid sequences. Present in toxins, virulence factors, adhesins, and surface proteins of <i>Chlamydia</i> , <i>Helicobacteria</i> , <i>Bordetella</i> , <i>Leishmania</i> , <i>Borrelia</i> , <i>Rickettsia</i> , <i>Neisseria</i> , and <i>Bacillus anthracis</i>	[26]
FxxN, GGA(I,L,V)	These are tetrapeptide motifs FxxN and GGA(I, L, V) present in polymorphic membrane protein family (Pmp) of <i>Chlamydia pneumonia</i> . They are required as duplicate copies for adhesion to host cells	[27]
RGD, SGxG	These are arginine-glycine-aspartic acid (RGD) and glycosaminoglycan binding site (SGxG) motifs present in autotransporter family proteins of <i>Bordetella pertussis</i> —pertactin (Prn), <i>Bordetella</i> resistance to killing (BrkA) and <i>Bordetella</i> autotransporter protein-C (BapC). The arrangement of motifs confer BapC adhesive property to binding sites on the macrophages and epithelial cells	[28]
PARF motif (A/T/E) xYLxx(LYF)N	This is a (A/T/E)XYLXXLN amino acid sequence motif referred to as PARF (peptide associated with rheumatic fever). It is located in the N-terminal hypervariable region of the collagen binding M protein type 3 of <i>Streptococcus pyogenes</i> and <i>Streptococcus dysgalactiae</i> ssp. <i>equisimilis</i> (SDSE)	[29]
HExxH containing metalloprotease adhesins	This is a zinc binding sequence motif His-Glu-Xaa-Xaa-His. It is present in certain adhesins like <i>Treponema pallidum</i> extracellular matrix binding adhesin Tp0751	[30]

4. Transmembrane domain: Transmembrane domains are the regions of membrane proteins which traverse in and out, looping through the membrane. They are characteristics of integral membrane proteins. Since adhesins are mostly membrane proteins, the prediction of proteins having transmembrane domain would contribute to the set of putative adhesins (*see Note 3*). However, this approach would lead to large number of false positives as not all proteins possessing transmembrane domains are adhesins.
5. Domain Search: Domains are conserved autonomously folding functional unit of a protein [34]. The domains of a protein together define the function of the protein. The domain information of an unannotated protein sequence can be used to predict its function (*see Note 4*).

Some adhesin domains are known. Examples include GLEYA adhesin domain, PA14 domain, ALS\_N domain in fungal species,

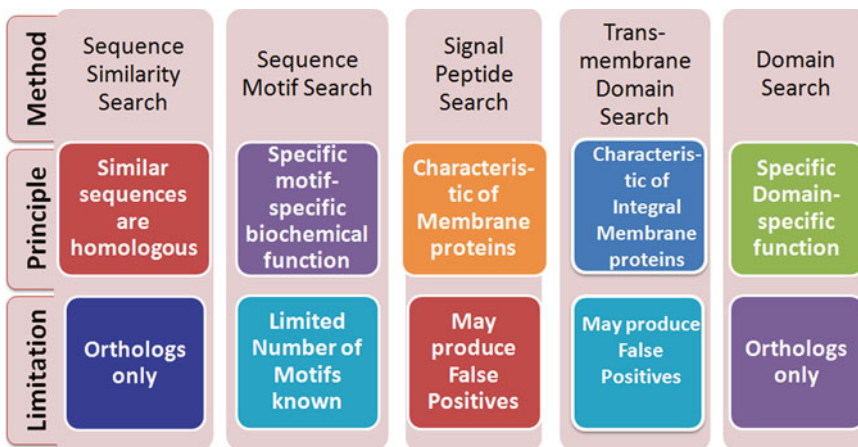
YadA adhesin protein domain, fibrinogen-binding domain, Gingipain adhesin domains forming part of cleaved adhesin domain in bacterial species [35–40]. Sequence analysis to study the presence of such adhesin related domains in the query protein sequence may help predicting potential adhesins.

## 2.2 Challenges in Bioinformatics Characterization of Adhesins

Although the computational methods described in preceding section permit identifying potential adhesins they are limited in their scope. Unlike many families of proteins, adhesins lack a well defined common sequence pattern or signatures, rendering their identification using the general signature sequence search or unique motif search difficult. This is mainly because adhesins include diverse proteins. Even adhesins belonging to same species include diverse molecular types and lack a common specific pattern in sequence. For example, the adhesins- M proteins in *Streptococcus pyogenes*, Gal/GalNAc lectin in *Entamoeba histolytica*, Fimbrial adhesins in *Escherichia coli*, Blood group antigen binding adhesin (BabA) in *Helicobacter pylori*, YadA collagen binding adhesin in *Yersinia enterocolitica* [41–45] lack significant similarity among each other.

However, in certain cases like in fungal species where many adhesins possess fungal specific GPI-motif, sequence motif search algorithm can be used to screen for potential fungal adhesins. However, identification methods solely based on motif searches such as GPI-anchor searches could return several false positives because all GPI-anchored proteins are not adhesins. Similar concerns apply to other identification methods such as Signal peptide search. The basic principles and limitations of various bioinformatics approaches used to characterize adhesins are summarized in Fig. 1.

These limitations formed the foundation for developing non-homology group of algorithms, which use a large number of compositional properties.



**Fig. 1** Advantages and limitations of different sequence and motif based approaches for prediction of potential virulence factors

## 2.3 Specialized Algorithms for Adhesin Prediction

### 2.3.1 SPAAN: A Software Program for Prediction of Adhesins and Adhesin-Like Proteins

SPAAN is an adhesin prediction tool developed using artificial neural network trained on compositional properties of known adhesins and non-adhesins. The algorithm is trained to predict adhesins and adhesin-like proteins solely from the sequence data. It is a non-homology method. SPAAN was trained using 105 compositional properties including 20 amino acid frequencies, 20 selected dipeptide frequencies, 20 multiplet frequency, 20 charge compositions, and 25 hydrophobic compositions. It showed an optimal sensitivity of 89 % and specificity of 100 % on a defined test set and could identify 97.4 % of known adhesins at high Pad value from a wide range of bacteria. Though SPAAN was trained on datasets dominated by bacterial adhesins, it can be used for general purpose to identify adhesins from a wide spectrum of species belonging to diverse phyla. Many novel adhesins in diverse species have been characterized using SPAAN [46]. It is one of the most widely used adhesin prediction tool available. The standalone software package of SPAAN can be downloaded from <http://sourceforge.net/projects/adhesin/files/>.

System Requirement: Red Hat Linux version 7.3 or above.

Other requirements: C compiler

Instruction for usage

1. SPAAN is provided as a tar-gzipped file. Post download, it should be unzipped and untarred by the command “tar xvzf SPAAN.tar.gz.”
2. The query sequences should be in FASTA format. Multiple sequences can be present in the input file.
3. The input file should be named as “query.dat.”
4. The command to run the software SPAAN is “./askquery.”
5. The output data is stored in “query.out.”
6. If the existing binary files are not compatible to the system, the source C codes provided need to be recompiled using the following example command—“gcc -lm standard.c -o standard.o.”

List of C source codes to be compiled—standard.c, filter.c, annotate.c, and finalp1.c in the main SPAAN directory; recognize.c, AAcompo.c, hdr.c, multiplets.c, querydipep.c, and charge.c in their respective directories: AAcompo, hdr, multiplets, dipep, and charge: recognize.c needs to be compiled individually in each of the five mentioned directories.

Figure 2 describes an example of a run of SPAAN output result file “query.out.”

### 2.3.2 MAAP: Malarial Adhesin and Adhesin-Like Proteins Predictor

MAAP was developed using Support Vector Machine (SVM) trained through compositional properties for classifying malarial adhesins and adhesin-like proteins [47]. The SVM<sup>light</sup> package [48] of Support Vector Machine was used for this purpose. A total of 420 compositional properties including amino acid frequencies of

1 SN	Pad-value	Protein name (Annotation)
2 1	0.429278	>ADK22398 A/Acre/15093/2010 2010/03/03 HA
3 2	0.431378	>ADK22399 A/Acre/26954/2010 2010/04/04 HA
4 3	0.370007	>CAA24269 A/Aichi/2/1968 1968// HA
5 4	0.364150	>AAA43239 A/Aichi/2/1968 1968// HA
6 5	0.398248	>BAF37221 A/Aichi/2/1968 1968// HA
7 6	0.378111	>BAF48361 A/Aichi/2/1968 1968// HA
8 7	0.413332	>AFH00065 A/Akita/4/1993 1993// HA
9 8	0.387092	>AGF44915 A/Alabama/3101/2012 2012/12/11 HA
10 9	0.437835	>ADY05347 A/Alabama/AF2692/2010 2010/12/16 HA
11 10	0.516410	>ABV30415 A/Alabama/UR06-0545/2007 2007/03/07 HA
12 11	0.426751	>AGF68882 A/Alaska/03/2010 2010/08/06 HA
13 12	0.438203	>ACJ73748 A/Alaska/04/2008 2008/08/09 HA
14 13	0.507199	>ACA33696 A/Alaska/05/2007 2007/03/20 HA
15 14	0.464186	>AGF44917 A/Alaska/3103/2012 2012/11/15 HA
16 15	0.439399	>ADL39357 A/Alaska/WRAIR1145P/2009 2009/03/ HA
17 16	0.468747	>ABQ01366 A/Albany/1/1976 1976// HA
18 17	0.374754	>ABN51099 A/Albany/11/1968 1968// HA
19 18	0.411987	>AB033069 A/Albany/14/1978 1978// HA
20 19	0.469509	>ABP49492 A/Albany/15/1976 1976// HA
21 20	0.377399	>AB052357 A/Albany/17/1968 1968// HA
22 21	0.398604	>AB044068 A/Albany/18/1968 1968// HA
23 22	0.371164	>ABN51121 A/Albany/19/1968 1968// HA
24 23	0.408893	>AB052335 A/Albany/20/1974 1974// HA
25 24	0.378263	>ABN51132 A/Albany/3/1969 1969// HA
26 25	0.385848	>ABS49910 A/Albany/3/1970 1970// HA

**Fig. 2** An example of a run of SPAAN output result file “query.out.” The results are output under three column heads, Serial No. (SN), Probability of adhesin (Pad-value), Protein name (Annotation)

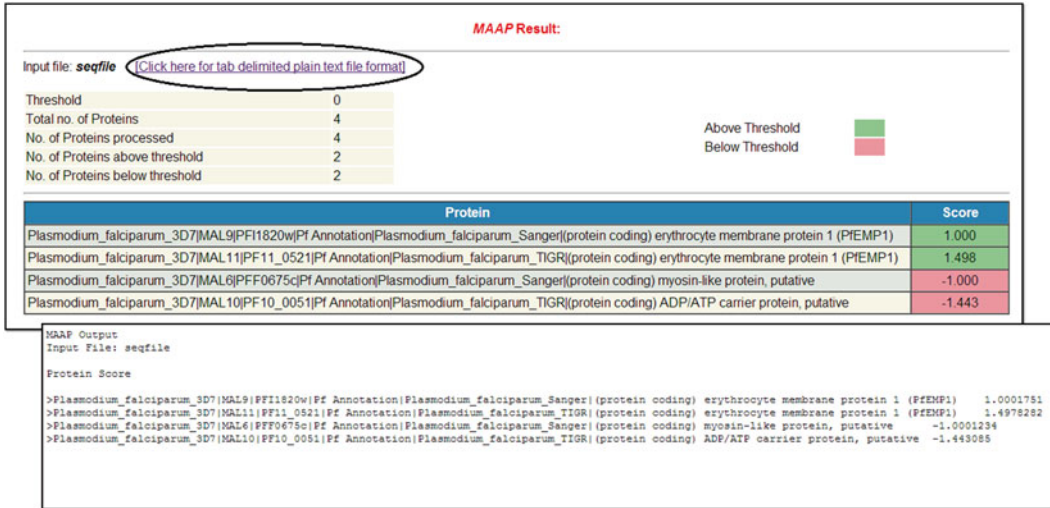
20 and 400 dipeptide frequencies were used to characterize the sequences of known adhesins and nonadhesins of *Plasmodium* species. MAAP runs on complete proteomes of *Plasmodium* species revealed that in *Plasmodium falciparum* at *Pmaap* scores above 0.0, a sensitivity of 100 % was observed with two false positives. In *P. vivax* and *P. yoelii* an optimal threshold *Pmaap* score of 0.7 was found optimal with very few false positives (upto 5). The MAAP Web server provides users with an interface where they can paste or upload their query sequences and predict whether the protein sequence is an adhesin (*see Note 5*). Users have the facility to set their own desired threshold cutoff value. The result can be exported as tab delimited text file by the users. The standalone version can be downloaded from the “Download” tab of MAAP Web server or <http://sourceforge.net/projects/adhesin/files/>.

Figure 3 describes the output result obtained using MAAP Web server.

### 2.3.3 FungalRV adhesin Predictor

In pathogenic fungi, adhesins play major roles as virulence factors mediating the interaction of the pathogens to variety of host cell types. In addition, adhesins in fungi aid in biofilm formation contributing to increased drug resistance and persistence of infections [49]. It has been established that differences in adhesion are responsible for greater virulence of one strain compared to other in fungi [50]. The fungal pathogens represent a diverse group of species.





**Fig. 3** Screenshot of output result obtained using MAAP Web server. The protein sequences scoring above threshold are highlighted in *green* color, whereas those scoring below the threshold are highlighted in *red* color. The result can be saved in a tab delimited plain text file format by clicking on the *purple* colored link (*encircled*)

FungalRV adhesin predictor was developed using Support Vector Machine (SVM) trained through compositional properties for classifying human pathogenic fungal adhesins and adhesin like proteins [51]. This tool was developed using SVM<sup>light</sup> package of Support Vector Machine trained through 3,945 compositional properties including amino acid frequencies of 20 from amino acids, 247 selected dipeptide frequencies, 3,653 selected tripeptide frequencies, 20 amino acid multiplets frequencies, frequency of the hydrophobic amino acids and four moments of hydrophobic amino acid distribution of order 2–5. This is a non-homology based prediction tool. We obtained an overall MCC value of 0.8702 considering all 8 pathogens, namely, *Candida albicans*, *Candida glabrata*, *Aspergillus fumigatus*, *Coccidioides immitis*, *Coccidioides posadasii*, *Histoplasma capsulatum*, *Blastomyces dermatitidis*, and *Paracoccidioides brasiliensis* thus showing high sensitivity and specificity at a threshold of 0.511. In case of *P. brasiliensis* the algorithm achieved a sensitivity of 66.67 %. This tool was made into FungalRV Web server available at <http://fungalrv.igib.res.in>. The “Adhesin Predictor” tab of the FungalRV Web server provides users with an interface where they can paste or upload their query sequences and predict whether the protein sequence is a fungal adhesin (*see Note 6*). Users have been provided the facility to set their own desired threshold cutoff value. This facility has been provided to allow users to optimize the threshold for other fungi for which “FungalRV adhesin predictor” was not trained. The result can be exported as tab delimited text file by the users. The facility to search for fungal specific GPI



**FungalRV**  
Adhesin Prediction and Immunoinformatics portal for human fungal pathogens

IGRB  
INSTITUTE OF GENOMICS & BIOTECHNOLOGY

Home Adhesin Predictor Immunoinformatics Data Known Vaccines Download Help Contact Us

**FungalRV adhesin predictor Result:**

Input file: [seqfile](#) [\(Click here for tab delimited plain text file format\)](#)

Threshold	0	
Total no. of Proteins	6	
No. of Proteins processed	6	Above Threshold <span style="color: green;">■</span>
No. of Proteins above threshold (Adhesin)	3	Below Threshold <span style="color: red;">■</span>
No. of Proteins below threshold (Non Adhesin)	3	

Protein	Score	Blast with Href Proteins	Search for GPI fungal pattern
Afu3g09690   extracellular thaumatin domain protein, putative   <i>Aspergillus fumigatus</i>   chr_3   AAHF01000002   93	1.0010	Blast	Pattern
gi 71001330 ref XP_755346.1  adhesin [ <i>Aspergillus fumigatus</i> Af293]	1.0005	Blast	Pattern
gi 1022896 gb AAA91036.1  WI-1 adhesin	1.0002	Blast	Pattern
gi 74591477 sp Q5AGC1.1 FIP1_CANAL RecName: Full=Pre-mRNA polyadenylation factor FIP1	-0.7367	Blast	Pattern
gi 239606518 gb EEQ83505.1  RNA polymerase II transcriptional coactivator [ <i>Ajellomyces dermatitidis</i> ER-3]	-0.9991	Blast	Pattern

```

FungalRV adhesin predictor Output
Input File: seqfile

Protein Score
>Afu3g09690 | extracellular thaumatin domain protein, putative | Aspergillus fumigatus | chr_3 | AAHF01000002 | 93 | 1.0009641
>gi|71001330|ref|XP_755346.1| adhesin [Aspergillus fumigatus Af293] | 1.0004667
>gi|1022896|gb|AAA91036.1| WI-1 adhesin | 1.0001888

>gi|74591477|sp|Q5AGC1.1|FIP1_CANAL RecName: Full=Pre-mRNA polyadenylation factor FIP1 | -0.73674514
>gi|239606518|gb|EEQ83505.1| RNA polymerase II transcriptional coactivator [Ajellomyces dermatitidis ER-3] | -0.99906948
>gi|239595576|gb|EEQ78157.1| RNA polymerase II transcriptional coactivator [Ajellomyces dermatitidis SLH14081] | -1.0359635

```

**Fig. 4** Screenshot of output result obtained using FungalRV Web server. The protein sequences scoring above threshold are highlighted in *green* color, whereas those scoring below the threshold are highlighted in *red* color. The result can be saved in a tab delimited plain text file format by clicking on the *purple* colored link (*encircled*). Additional data on BLAST with Href proteins and GPI patterns are also displayed

pattern in the predicted adhesins and adhesin like proteins using fuzzpro program of EMBOSS has been provided. Users also have been provided the facility to conduct BLAST search with human reference proteins (*see Note 7*). The standalone version can be downloaded from the “Download” tab of FungalRV Web server or <http://sourceforge.net/projects/adhesin/files/>. Figure 4 describes the output adhesin prediction results obtained using FungalRV Web server.

### 2.3.4 Faapred

In addition to FungalRV, another Support Vector Machine (SVM) based algorithm named Faapred for prediction of fungal adhesins and adhesin-like proteins is available [52]. The SVM models for Faapred development were trained with compositional features- amino acid, dipeptide, multiplet fractions, charge and hydrophobic compositions, as well as PSI-BLAST derived PSSM matrices. The best classifiers were screened based on high MCC and accuracy. The amino acid composition model (ACHM), PSSM-a, and PSSM-b came out as the best classifiers with ACHM providing the highest MCC value of 0.610. Thus the prediction of Faapred uses classifiers based on compositional properties as

well as PSSM. Faapred provides overall accuracy of 86 %. The prediction method is freely available as a World Wide Web based server at <http://bioinfo.icgeb.res.in/faap>.

---

### 3 Notes

1. BLAST algorithm is widely used to fetch orthologues. Reciprocal Best Hits (RBH) method has shown good efficiency in identifying orthologues. RBH is based on the principle that two genes from different genomes are orthologous if they find each other as the best hit in BLAST search in the other genome. Here BLASTP is usually carried out at a maximum E-value threshold of  $1 \times 10^{-6}$ , including Smith–Waterman algorithm and Soft-filtering.
2. Various bioinformatics algorithms are available, which aid identifying signal peptides. SignalP algorithm available at <http://www.cbs.dtu.dk/services/SignalP/> is widely used. The query sequences input in FASTA format can be submitted to predict presence of signal peptides.
3. Transmembrane prediction algorithms for example TMHMM available at <http://www.cbs.dtu.dk/services/TMHMM/> is generally used to predict presence of transmembrane regions.
4. Conserved Domains can be predicted using domain prediction algorithms for example CDD search available at <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>. The presence of known adhesin related domains in the query sequences can be predicted.
5. The query proteins in FASTA format can be uploaded in the MAAP Web server. The server can be used to analyze the whole genome in one run.
6. Query protein sequences in FASTA format can be uploaded in FungalRV Web server. This Web server can be used to analyze the whole genome.
7. An adhesin vaccine should ideally not have similarity to human reference proteins to avoid cross-reactivity. The facility to conduct BLAST search with human reference proteins has therefore been provided in the FungalRV Web server. The cutoff E-value used here is 0.01, which borders on the limits of threshold similarity.

---

### Acknowledgements

RC thanks The Indian Council of Medical Research for fellowship. This work was funded through grants “GENESIS” BSC0121 to SR from CSIR.

## References

- Casadevall A, Fang FC, Pirofski LA (2011) Microbial virulence as an emergent property: consequences and opportunities. *PLoS Pathog* 7:e1002136
- Weiss RA (2002) Virulence and pathogenesis. *Trends Microbiol* 10:314–317
- Poulin R, Combes C (1999) The concept of virulence: interpretations and implications. *Parasitol Today* 15:474–475
- Cross AS (2008) What is a virulence factor? *Crit Care* 12:196
- Casadevall A, Pirofski LA (2009) Virulence factors and their mechanisms of action: the view from a damage-response framework. *J Water Health* 7(Suppl 1):S2–S18
- Rothy A, James L, Ed. (1988) Virulence mechanisms of bacterial pathogens. American Society for Microbiology, ISBN 0-914826-99-9
- Casadevall A, Pirofski LA (1999) Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infect Immun* 67:3703–3713
- Wizemann TM, Adamou JE, Langermann S (1999) Adhesins as targets for vaccine development. *Emerg Infect Dis* 5:395–403
- Ofek I, Hasty DL, Sharon N (2003) Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol Med Microbiol* 38:181–191
- Pandey AK, Reddy KS, Sahar T, Gupta S, Singh H, Reddy EJ, Asad M, Siddiqui FA, Gupta P, Singh B, More KR, Mohammed A, Chitnis CE, Chauhan VS, Gaur D (2013) Identification of a potent combination of key *Plasmodium falciparum* merozoite antigens that elicit strain-transcending parasite-neutralizing antibodies. *Infect Immun* 81:441–451
- Langermann S, Mollby R, Burlein JE, Palaszynski SR, Auguste CG, DeFusco A, Strouse R, Schenerman MA, Hultgren SJ, Pinkner JS et al (2000) Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic *Escherichia coli*. *J Infect Dis* 181:774–778
- Ausiello CM, Lande R, Urbani F, la Sala A, Stefanelli P, Salmaso S, Mastrantonio P, Cassone A (1999) Cell-mediated immune responses in four-year-old children after primary immunization with acellular pertussis vaccines. *Infect Immun* 67:4064–4071
- Halperin SA, Scheifele D, Mills E, Guasparini R, Humphreys G, Barreto L, Smith B (2003) Nature, evolution, and appraisal of adverse events and antibody response associated with the fifth consecutive dose of a five-component acellular pertussis-based combination vaccine. *Vaccine* 21:2298–2306
- Gorringe AR, Pajón R (2012) Bexsero: a multicomponent vaccine for prevention of meningococcal disease. *Hum Vaccin Immunother* 8:174–183
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Pichel M, Binsztein N, Viboud G (2000) CS22, a novel human enterotoxigenic *Escherichia coli* adhesin, is related to CS15. *Infect Immun* 68:3280–3285
- Fleury B, Bergonier D, Berthelot X, Peterhans E, Frey J, Vilei EM (2002) Characterization of P40, a cytoadhesin of *Mycoplasma agalactiae*. *Infect Immun* 70:5612–5621
- Belloy L, Vilei EM, Giacometti M, Frey J (2003) Characterization of LppS, an adhesin of *Mycoplasma conjunctivae*. *Microbiology* 149:185–193
- Nakane D, Adan-Kubo J, Kenri T, Miyata M (2011) Isolation and characterization of P1 adhesin, a leg protein of the gliding bacterium *Mycoplasma pneumoniae*. *J Bacteriol* 193:715–722
- Renesto P, Samson L, Ogata H, Azza S, Fourquet P, Gorvel JP, Heinzen RA, Raoult D (2006) Identification of two putative rickettsial adhesins by proteomic analysis. *Res Microbiol* 157:605–612
- Palaniappan RU, Chang YF, Jusuf SS, Artiushin S, Timoney JF, McDonough SP, Barr SC, Divers TJ, Simpson KW, McDonough PL, Mohammed HO (2002) Cloning and molecular characterization of an immunogenic LigA protein of *Leptospira interrogans*. *Infect Immun* 70:5924–5930
- D'haeseleer P (2006) What are DNA sequence motifs? *Nat Biotechnol* 24:423–425
- De Groot PW, Hellingwerf KJ, Klis FM (2003) Genome-wide identification of fungal GPI proteins. *Yeast* 20:781–796
- Richard ML, Plaine A (2007) Comprehensive analysis of glycosylphosphatidylinositol-anchored proteins in *Candida albicans*. *Eukaryot Cell* 6:119–133
- Weig M, Jansch L, Gross U, De Koster CG, Klis FM, De Groot PW (2004) Systematic identification in silico of covalently bound cell wall proteins and analysis of protein-polysaccharide linkages of the human pathogen *Candida glabrata*. *Microbiology* 150:3129–3144
- Bradley P, Cowen L, Menke M, King J, Berger B (2001) BETAWRAP: successful prediction of parallel beta-helices from primary sequence reveals an association with many microbial pathogens. *Proc Natl Acad Sci U S A* 98:14819–14824

27. Mölleken K, Schmidt E, Hegemann JH (2010) Members of the Pmp protein family of *Chlamydia pneumoniae* mediate adhesion to human cells via short repetitive peptide motifs. *Mol Microbiol* 78:1004–1017
28. Bokhari H, Bilal I, Zafar S (2012) BapC auto-transporter protein of *Bordetella pertussis* is an adhesion factor. *J Basic Microbiol* 52:390–396
29. Reissmann S, Gillen CM, Fulde M, Bergmann R, Nerlich A, Rajkumari R, Brahmadathan KN, Chhatwal GS, Nitsche-Schmitz DP (2012) Region specific and worldwide distribution of collagen-binding M proteins with PARF motifs among human pathogenic streptococcal isolates. *PLoS One* 7:e30122
30. Houston S, Hof R, Francescutti T, Hawkes A, Boulanger MJ, Cameron CE (2011) Bifunctional role of the *Treponema pallidum* extracellular matrix binding adhesin Tp0751. *Infect Immun* 79:1386–1398
31. Blobel G, Dobberstein B (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* 67:835–851
32. Lodish H, Berk A, Zipursky SL et al. (2000) *Molecular cell biology*, 4th edn. W. H. Freeman, New York, NY. Section 17.4, translocation of secretory proteins across the ER membrane. <http://www.ncbi.nlm.nih.gov/books/NBK21532/>
33. Lee VT, Schneewind O (2001) Protein secretion and the pathogenesis of bacterial infections. *Genes Dev* 15:1725–1752
34. Phillips DC (1966) The three-dimensional structure of an enzyme molecule. *Sci Am* 215:78–90
35. Rigden DJ, Mello LV, Galperin MY (2004) The PA14 domain, a conserved all-beta domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends Biochem Sci* 29:335–339
36. Linder T, Gustafsson CM (2008) Molecular phylogenetics of ascomycotal adhesins—a novel family of putative cell-surface adhesive proteins in fission yeasts. *Fungal Genet Biol* 45:485–497
37. Phan QT, Myer CL, Fu Y, Sheppard DC, Yeaman MR, Welch WH, Ibrahim AS, Edwards JE Jr, Filler SG (2007) Als3 is a *Candida albicans* invasin that binds to cadherins and induces endocytosis by host cells. *PLoS Biol* 5:e64
38. Nummelin H, Merckel MC, Leo JC, Lankinen H, Skurnik M, Goldman A (2004) The *Yersinia adhesin* YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J* 23:701–711
39. Li N, Yun P, Nadkarni MA, Ghadikolaei NB, Nguyen KA, Lee M, Hunter N, Collyer CA (2010) Structure determination and analysis of a haemolytic gingipain adhesin domain from *Porphyromonas gingivalis*. *Mol Microbiol* 76:861–873
40. Konkel ME, Christensen JE, Keech AM, Monteville MR, Klena JD, Garvis SG (2005) Identification of a fibronectin-binding domain within the *Campylobacter jejuni* CadF protein. *Mol Microbiol* 57:1022–1035
41. Ellen RP, Gibbons RJ (1972) M protein-associated adherence of *Streptococcus pyogenes* to epithelial surfaces: prerequisite for virulence. *Infect Immun* 5:826–830
42. Schembri MA, Kjaergaard K, Sokurenko EV, Klemm P (2001) Molecular characterization of the *Escherichia coli* FimH adhesin. *J Infect Dis* 183(Suppl 1):S28–S31
43. Mann BJ (2002) Structure and function of the *Entamoeba histolytica* Gal/GalNAc lectin. *Int Rev Cytol* 216:59–80
44. Wen S, Moss SF (2009) *Helicobacter pylori* virulence factors in gastric carcinogenesis. *Cancer Lett* 282:1–8
45. Nummelin H, Merckel MC, Leo JC, Lankinen H, Skurnik M, Goldman A (2004) The *Yersinia adhesin* YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J* 23:701–711
46. Sachdeva G, Kumar K, Jain P, Ramachandran S (2005) SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics* 21:483–491
47. Ansari FA, Kumar N, Bala SM, Gnanamani M, Ramachandran S (2008) MAAP: malarial adhesins and adhesin-like proteins predictor. *Proteins* 70:659–666
48. Joachims T (1999) Making large-scale SVM learning practical. In: Scholkopf B, Burges C, Smola A (eds) *Advances in Kernel methods—support vector learning*. MIT, Cambridge, MA, pp 169–185
49. Hawser SP, Douglas LJ (1994) Biofilm formation by *Candida* species on the surface of catheter materials in vitro. *Infect Immun* 62:915–921
50. Verstrepen KJ, Klis FM (2006) Flocculation, adhesion and biofilm formation in yeasts. *Mol Microbiol* 60:5–15
51. Chaudhuri R, Ansari FA, Raghunandan MV, Ramachandran S (2011) FungalRV: adhesin prediction and immunoinformatics portal for human fungal pathogens. *BMC Genomics* 12:192
52. Ramana J, Gupta D (2010) FaaPred: a SVM-based prediction method for fungal adhesins and adhesin-like proteins. *PLoS One* 5:e9695