BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis
Volume 8(19)

Cluster analysis identifies aminoacid compositional features that indicate *Toxoplasma gondii* adhesin proteins

Ailan F Arenas[1]*, Gladys E Salcedo[2], Diego M Moncada[1], Diego A Erazo[1], Juan F Osorio[1] & Jorge E Gomez-Marin[1]

[1]Grupo de Parasitología Molecular (GEPAMOL), Centro de Investigaciones Biomédicas, Universidad del Quindío, Armenia, Colombia; [2]Grupo de Investigación y Asesoría en Estadística, Universidad del Quindío, Armenia, Colombia; Ailan F Arenas – Email: aylanfarid@yahoo.com; *Corresponding author

Received August 30, 2012; Accepted September 03, 2012; Published October 01, 2012

Abstract:
Toxoplasma gondii invade host cells using a multi-step process that depends on the regulated secretion of adhesions. To identify key primary sequence features of adhesins in this parasite, we analyze the relative frequency of individual amino acids, their dipeptide frequencies, and the polarity, polarizability and Van der Waals volume of the individual amino acids by using cluster analysis. This method identified cysteine as a key amino acid in the *Toxoplasma* adhesin group. The best vector algorithm of non-concatenated features was for 2 attributes: the single amino acid relative frequency and the dipeptide frequency. Polarity, polarizability and Van der Waals volume were not good classificatory attributes. Single amino acid attributes clustered unambiguously 67 apicomplexan hypothetical adhesins. This algorithm was also useful for clustering hypothetical *Toxoplasma* target host receptors. All of the cluster performances had over 70% sensitivity and 80% specificity. Compositional aminoacid data can be useful for improving machine learning-based prediction software when homology and structural data are not sufficient.

Keywords: Cluster analysis, adhesin, Toxoplasma

Background:

Adhesins in microbes are cell surface proteins that confer the ability of attachment to cells and tissue surfaces **[1].** Adhesins are the first line of a pathogen's strategy of host cell invasion and are an indispensable determinant of its virulence. Investigations into this primary event of host–pathogen interaction have revealed a wide array of proteins with adhesin function in a variety of pathogenic microbes **[2]**. *Toxoplasma gondii* is an apicomplexan parasite that is capable of infecting a broad host range, including humans **[3]**. The most important human health consequences of toxoplasmosis are the congenital transmission and the reactivation in immune suppressed patients, which are an important public health problem in some countries **[4]**.The emergence of parasites that are resistant to commonly used drugs and the lack of availability of vaccines aggravate the problem. One of the preventive approaches targets the adhesion of parasites to host cells and tissues. The abrogation of adhesion using the adhesins could be a focus for the development of new drugsor vaccine targets **[1]**.

The *Toxoplasma* tachyzoite lytic cycle begins with an active invasion of host cells that involves the release of adhesive proteins from apical secretory organelles called micronemes. Many microneme proteins (MICs) contain well-conserved functional domains, which are associated with adhesive activity **[4].** Such protein regions are the thrombospondin type 1 (TSP-1), von Wille brand Factor A (VWA) and plasminogen apple nematode (PAN) domains, which were originally defined based on their role in mediating protein-protein and cell-cell interactions in mammalian cells **[5]**. They are thought to interact with the extracellular matrix to mediate motility, attachment and / or invasion into host cells **[6, 7]**.

Experimental methods used for characterizing adhesin-like proteins are time-consuming and demand large resources. Computational methods such as homology searching can aid in identification, but this procedure suffers from limitations when the homologues are not well characterized. Sequence analysis based on the compositional properties provides relief for this problem **[8]**. The amino acid composition is a fundamental attribute of a protein and has a significant correlation with the protein's location, function, folding type, shape and in vivo stability. In recent years, compositional properties have been applied to problems as diverse as the prediction of functional roles **[9]**. One of the statistical methods to analyze these properties is the cluster analysis of proteins according to shared annotation, which can reveal related subsets that warrant further investigation **[10]**. In this method, a successful hierarchical clustering is defined as the point in the hierarchy at which one of the clusters contains no false positive annotations **[11]**. The results based on the metrical distance of protein families are very useful for classifying according to the distinct biological context without relying on another type of information such as domains or phylogenetic profiles. The advantage of this methodology relies on the fact that, without complex information, good classification power can be obtained that complements the traditional classification methods. Accordingly, we wonder whether a cluster statistical method would identify the primary structural level features that exclusively characterize *Toxoplasma* adhesin proteins, providing novel amino acid features that surely will indicate a protein sequence to be an adhesin.

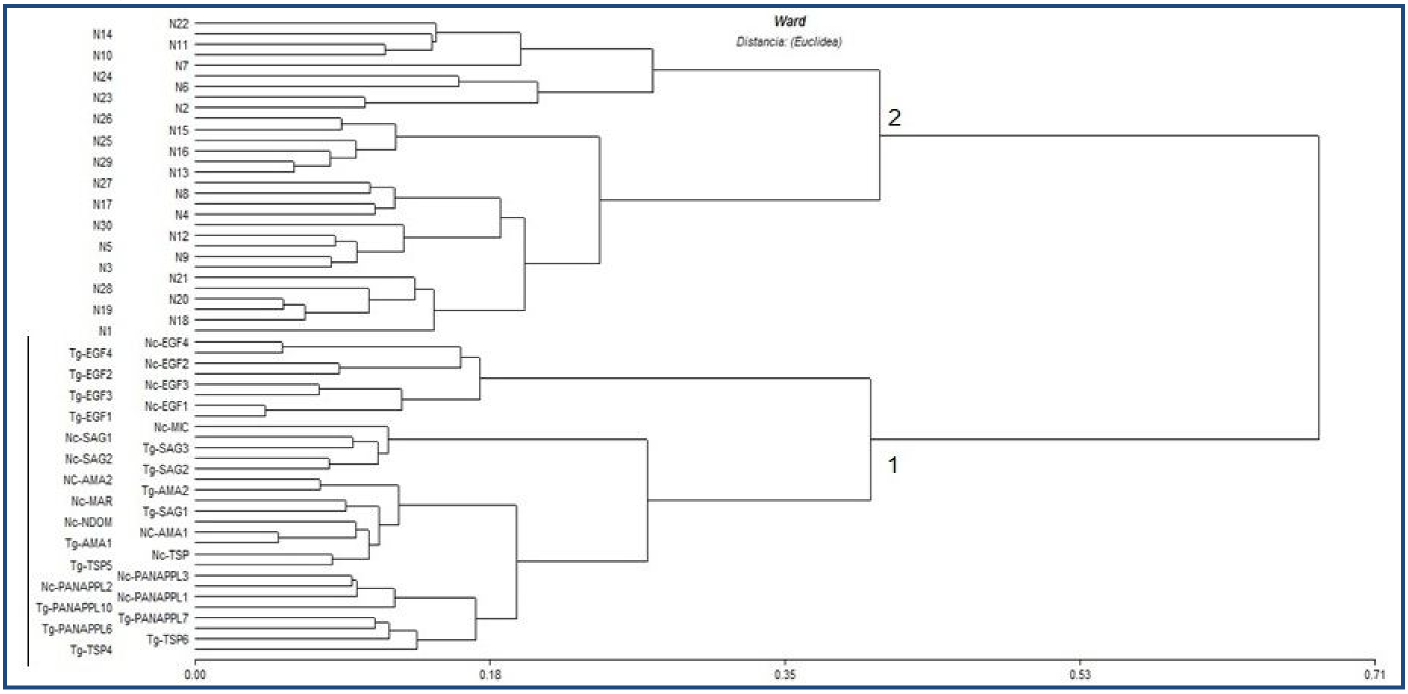


Figure 1: Cluster analysis by Euclidean distance using the Ward method from the relative frequency of each amino acid, taken up from a set of 30 *Toxoplasma* and *Neospora* proteins (representing its adhesive domains) and 30 proteins with no adhesin function. Tg and Nc means *Toxoplasma gondii* and *Neospora caninum,* respectively. Here, EGF means epidermal grow factor, PAN / APPLE meansthe pan or apple domain, TSP means trombospondin, SAG means the surface antigen group, and AMA means the apical membrane antigen. The numbers at the end of the parasite domain signs mean the numbers of repeat domains, and N means the non-adhesins. The dendrogram shows that this simple attribute can separate the two groups.

Methodology:

Dataset

Toxoplasma adhesin-like proteins were downloaded from the recent release (Version 7.0, 21 July2011) of the predicted proteome of the*Toxoplasma gondii* ME49 strain database (www.toxodb.org). The sequences were filtered, searching the experimental data (we considered only sequences with a proven adhesion function).To obtain a better sequence representation, the searches for adhesin domains such as EFG (epidermal growth factor),TSP-1, VWA, PAN and functional motifs were performed by using Smart and the Prosite domain and motif databases **[12]**. We found 20 well-characterized *Toxoplasma* proteins with an adhesion function that was experimentally tested. To increase the adhesin data set, we also searched the orthologous adhesins in the *Neospora caninum* genome because the *Neospora* and *Toxoplasma* genomes are closely related species, and we obtained, in total, an adhesin set with 30 *Toxoplasma Neospora*.

For the negative dataset, we included ribosomal, metabolic, and other intracellular and membrane-associated protein sequences. In total, a negative dataset of 600 proteins was assembled, which was grouped into 12 sets with 50 non-adhesin proteins in each set. All of these sequences were filtered with a 60% sequence identity using the program CD-HIT.

Compositional properties as numerical features

Each protein sequence is represented by a set of five attribute feature vectors: **(i)** ***Amino acid frequencies***: amino acid frequency *fi* = (counts of the i-th amino acid in the sequence) / 1, where *i* = 1, . . . , 20 and 1 is the length of the protein; **(ii)** ***Dipeptide frequencies***: the frequency of a dipeptide (i, j) *fij* = (counts of theij-th dipeptide) / (total dipeptide counts), where i, j are from 1 to 20.There are 20*20 = 400 possible dipeptides; **(iii)** ***Multiplet frequencies***: Multiplets are defined as homopolymeric stretches (X)n, where X is an amino acid and n (an integer) > 2. After identifying all of the multiplets, the frequencies of the amino acids in the multiplets were computed as follows: (a) *fi*(m) = (counts of the i-th amino acid that occurs as a multiplet) / 1; (b) where 1 is the length of the sequence. There are 20 possible values for *fi* (m) for the 20 amino acids; **(iv)** ***Hydrophobic composition***: The amino acids were classified into five groups, based on their hydrophobicity scores: 1 (−8 for K, E, D and R), 2 (−4 for S, T, N and Q), 3 (−2 for P and H), 4 (+1for A, G, Y, C and W) and 5 (+2 for L, V, I, F and M) **[13]**. The inputs for each group are as follows: (a) *fi* = (counts of the i-th group) / (total counts in the protein), where i = 1,2,…,5; **(v)** ***Polarity, polarizability and Van der Waals volume:*** used as concatenated attributes. For each attribute, twenty amino acids were divided into three groups **(supplementary material S1)**, and for each protein sequence, every amino acid was replaced by the index 1, 2, or 3, depending on its group. Polarizability, polarity and Van der Waals volume composition was calculated for each group based on the simple formula: (a) *fi* = (counts of the i-th group) / (total counts in the protein), where i = 1, 2, 3.

Therefore, we had 442 frequencies that were used as numerical feature inputs for each sequence. Thus, each protein was represented by 442 numerical features obtained from its amino acid sequence. We implemented 5 algorithms for each attribute using a MATLAB 2009Ra platform. The algorithms were implemented to read FASTA sequences files. Once we had all of the frequencies from each attribute within a matrix, the Euclidean distances were calculated for adhesins as well as for non-adhesin groups through cluster analysis. These analyses were conducted using the STATGRAPHICS plus package.

External cluster evaluation

Clustering results were evaluated based on knowing the class labels, which were, in our case, proteins with or without adhesin function. We calculated the Rand index RI **[14]**, the Fowlkes.0-Mallows index FM **[15]** and the Matthews correlation coefficient MCC **[16]** **(For equations see supplementary material)**.

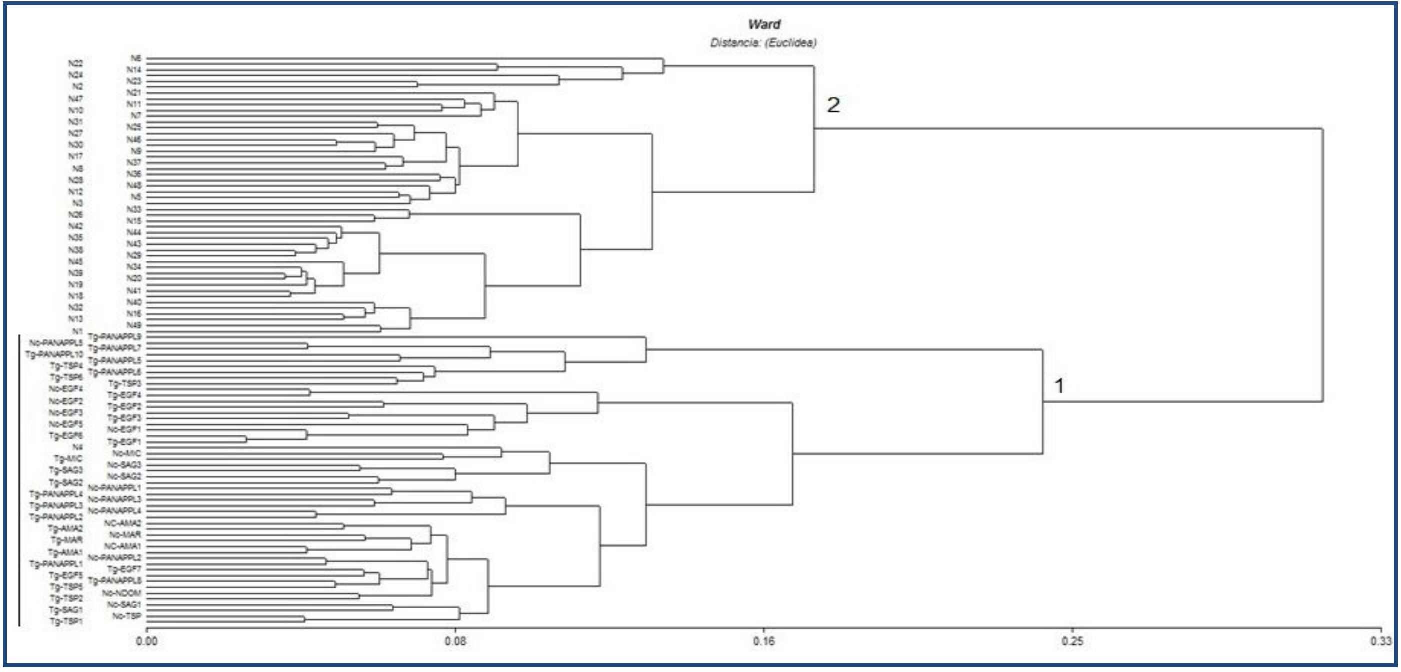


Figure 2: cluster analysis by Euclidean distance using the Ward method, from the relative frequency of 400 dipeptide combinations calculated from a set of 50 *Toxoplasma* and *Neospora* and 50 non-adhesins. Tg and Nc means *Toxoplasma gondii* and *Neospora caninum,* respectively, EGF means the epidermal growth factor, PAN / APPLE means the pan or apple domain, TSP means trombospondin, SAG means the surface antigen group, and AMA means the apical membrane antigen. The numbers at the end of the parasite domain signs mean the numbers of repeat domains, where N means non-adhesins. This attribute grouped most of the toxoplasma adhesins away from toxoplasma non-adhesins.

Results:

We found that, when we used each single amino acid frequency as an attribute into a set of 30 *Toxoplasma* and *Neospora,*the adhesin proteins could branch off from the non-adhesin set (proteins with no adhesive domains); the analysis pictured two large cluster groups that separated the *Toxoplasma* and *Neospora* adhesin domains' subcluster 1 from the negative subcluster 2 **(Figure 1)**.

We wanted to know what is the relative frequency for each amino acid that grouped subcluster 2 (non-adhesins) away from subcluster 1 (adhesins) in **(Figure 1)**. We found that cysteine "C" had the highest relative frequency difference between the two sets. Likewise, leucine "L", isoleucine "I", arginine "R", and threonine "T" also showed measurable differences **(supplementary material S1)**. We applied a hypothesis test for the difference between two means for each amino acid betweenthe parasites' adhesin and non-adhesin sets. We found that a t-Test forthe mean differences of the five amino acids mentioned above had a high significance level P<0.01 **Table 1 (see supplementary material).**

When we used the dipeptide frequency as a classifier of the adhesin feature, we also observed that this property in the *Toxoplasma* and *Neospora* adhesins set made this group cluster together **(Figure 2)**. Among the 400 possible combinations of dipeptides, those with a large relative frequency in the Toxoplasma adhesin set were the following: AC, DC, EC, GC, KC, PC RC, SC, and TC **(supplementary material S2)**. It can be noted that all of these combinations include cysteine.

Afterward, we merged the two attributes of the individual amino acid and the dipeptide occurrences into a characteristics vector, to strengthen the classification of the adhesins; we included 15 *Cryptosporidium* and *P. falciparum* adhesins. We found than *Toxoplasma* and *Neospora* adhesins merge into only one subcluster; however, 7 non-adhesins (FP) clustered within it, but those separated from the *P. falciparum* and Cryptosporidium branch had none mixed with the non-adhesins **(Figure 3, sub cluster 1a)**.

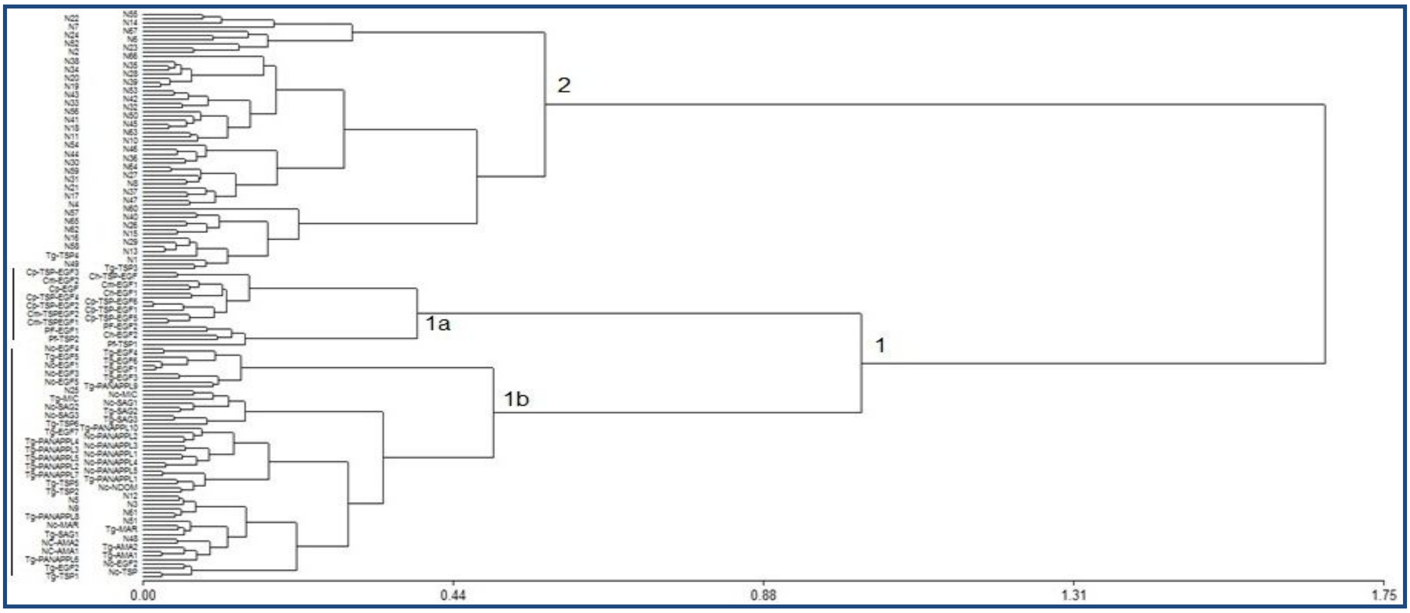


Figure 3: cluster analysis by Euclidean distance using the Ward method, from the relative frequency of the individual amino acids; 400 dipeptides concatenate into only one feature vector, which was calculated from 50 hypothetical *Toxoplasma*, *Neospora* and 17 *Plasmodium falciparum*, *Cryptosporidium* adhesins. Pf and Cp mean *P.falciparum* and *Crypstosporidium,* respectively, EGF means the epidermal growth factor, PAN / APPLE means the pan or apple domain, TSP means trombospondin, SAG means a surface antigen, and AMA means the apical membrane antigen group. The numbers at the end of the parasite domain signs mean the numbers of repeat domains, and N means the non-adhesins. (TSP1-EGF is a special multi-domain architecture in the *Crypstosporidium* adhesins).

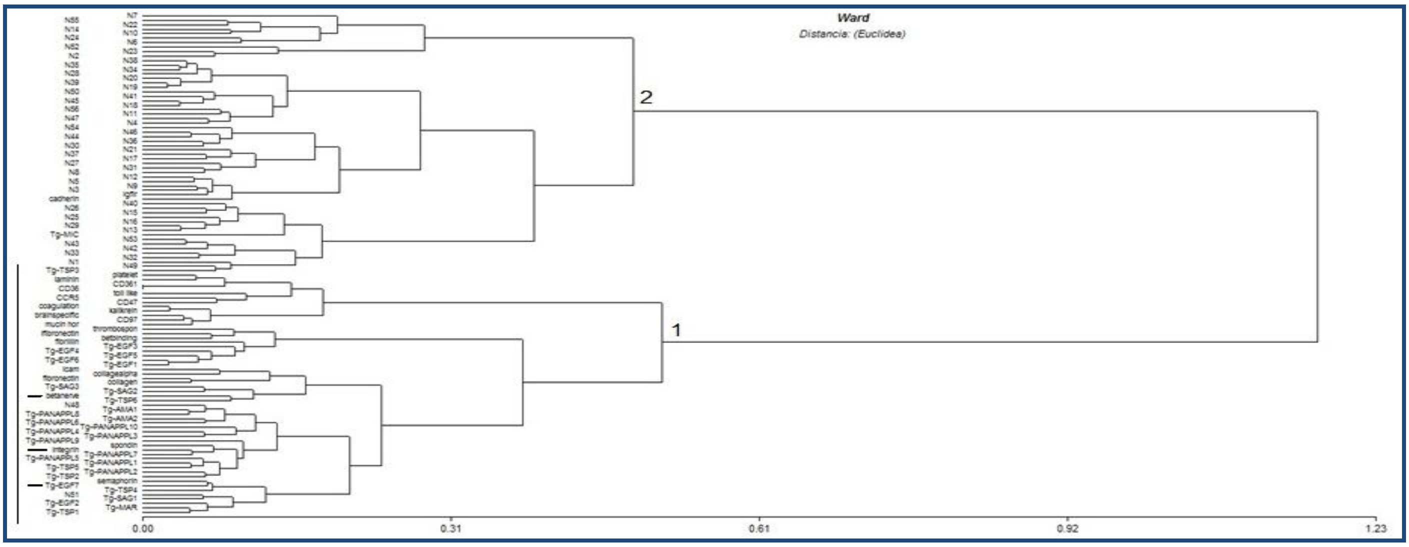


Figure 4: Cluster analysis by the Euclidean distance using the Ward method, from the relative frequency of the individual amino acids; 400 dipeptides concatenate into only one feature vector, which is calculated from 30 hypothetical *Toxoplasma* adhesins and 26 human receptors extracted from the literature, which is suspected to interact with *Toxoplasma* adhesins. Tg means *Toxoplasma gondii,* EGF means epidermal growth factor, PAN / APPLE means the pan or apple domain, TSP means trombospondin, SAG means

the surface antigen group, and AMA means the apical membrane antigen. The numbers at the end of the parasite domain signs mean the numbers of repeat domains, and N means non-adhesins. (Human extra-cell receptors were included in the dendrogram with their respective names).

We calculated the frequency of multiplets (a repetition of more than 2 of the same amino acid) in addition to physical and chemical characteristics such as hydrophobicity, polarity, polarizability and Van der Waals volume, but these properties do not work as good classifier attributes, at least in the adhesin family proteins. We observed that hydrophobic amino acids are less frequently in adhesins compared with non-adhesins and those there are no large differences in the frequencies that are observed between the two groups **(observations in supplementary material S1)**

We also wanted to know whether these two attributes can group the human extra-cellular domain; according to other reports, there is possible interaction with toxoplasma adhesive motifs. We extracted single amino acids and dipeptide frequencies from 26 human extra-cell receptors and applied a cluster analysis along with 30 *Toxoplasma* adhesins; we found that human extra-cell proteins clustered with *Toxoplasma* proteins but not with non-adhesin proteins **(Figure 4)**. According to the clustering, it was observed that some human proteins are closely related to *Toxoplasma* adhesins, such as integrin and spondin 1 with Tg_PAN/APPLE 7, beta-nerve growth factor with Tg_TSP6 and Tg_EGF7 with semaphoring 5 **(Figure 4)**.

We evaluated all of the cluster results based on knowing the class labels (in our case, adhesin or non-adhesin from parasites). We performed 3 indexes, the Rand index, the Fowlkes-Mallows index and the Matthews correlation coefficient for each cluster. We calculated the indexes from 36 dendrograms with 4 different negative sets for 3 attributes into 3 species adhesin groups. We found that the sensitivity and specificity as well as the 3 indexes are over 90% for *Toxoplasma* and *Neospora* clusters using the frequency of each amino acid and the dipeptide-amino acid merge **Table 2 (see supplementary material)**. These algorithms also classified adhesin in other apicomplexa with a sensitivity of over 80%, even though it could classify Human receptors as adhesins with a sensitivity of over 70% **(Table 2)**. These results demonstrated that the information at the primary structure level of the proteins has a high sensitivity and specificity for the classification of proteins that are involved in the same processes.

Discussion:

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters) in such a way that the data in each subset (ideally) share common traits that implicate more proximity according to a defined distance measure **[11]**. In our case, the amino acid composition feature, which was based on normalized counts of single or pairs of amino acids, identified clusters of proteins that were close to each other from a biological perspective.

It is well known that cysteine is a key amino acid because of its capacity to form disulfide bonds and to contribute to the folding and bioactivity of some adhesive domains such as apple and epidermal growth factor (EGF) **[17]**. Although cysteine is not the most frequent amino acid in Toxoplasma adhesin proteins, we have observed that this amino acid is one of the less frequent in the negative set **(Table 1)**, which are cytosolic proteins, and we demonstrated that cysteine frequency is a valuable clue to classifying a protein family according to its function and location.

Cysteine is unique among the coded amino acids because it contains a reactive sulph-hydryl group. Therefore, two cysteine residues can form a cysteine (disulfide link) between various parts of the same protein or between two separate polypeptide chains. It is known that one or more disulfide links are frequently found in excreted or plasma membrane proteins. In contrast, cytosolic proteins often lack disulfide links **[8]**. The known scarcity of disulfide bridges in cytosolic proteins may or may not translate into lower protein cysteine content for this reason **[18]**.

In accordance with the implication to infection, previous analyses have shown that conserved cysteine-rich domains play important roles at critical times during the invasion process in the life cycle of apicomplexan parasites. For example, Duffy-binding–like (DBL) domains, which are expressed as a part of the erythrocyte-binding proteins (DBLEBP), are essential cysteine-rich ligands that recognize specific host cell surface receptors. DBL domains also mediate cytoadherence as a part of the variant erythrocytic membrane protein-1 (PfEMP-1) on the surface of *P. falciparum*-infected erythrocytes **[19]**.

Hydrophobic amino acids in proteins are a crucial attribute for the proteins' function and location. Hydrophobic amino acids that segment in the protein could be important because of possible interaction with plasmatic membranes **[20, 21]**. According to our analysis, the hydrophobic amino acid composition was not a useful compositional attribute for separating most of the adhesin from the non-adhesin proteins **(supplementary material S2)**. Most of the cytosolic proteins interact with cytoplasmic organelle membranes, which make hydrophobic composition an unimportant feature when classifying proteins those are located extra-cellularly.

Apicomplexa parasites such as *Toxoplasma* and *Plasmodium* might invade different organisms and even different types of cells; they share some domains that are evolutionarily conserved among them, such as Apple, EGF, PAN and TSP1, which are crucial for invading host cells **[4, 22]**. Even they are also conservative in animal cytoadhesion, which make it possible for the parasite and host to exploit similar mechanisms for cytoadherence. We found that features at the amino acid level allow us to gather information that is common among them. It was observed that certain attributes make the best classification when only applied to a single species or regarding a species protein set (for example, *Toxoplasma* and *Neospora* adhesins vs. a negative set); as a result, we avoid evolutionary confusion in conserved and polymorphic residues in other species that are from different adaptations by the parasites to the respective host environment **(Table 2)**. These data could be useful for classifying and predicting new sequences with regard to the adhesin function into the apicomplexa group, and the data help to predict the possible human receptor interaction. This information also suggests the possibility that certain properties of proteins are not fully captured in algorithms that search only for protein domains.

It is possible that evolution also works at the amino acid level because the frequencies of certain amino acids are maintained when evolution attempts to retain conserved structures; there can be increases in the occurrence of a more reactive amino acid such as cysteine, with more cysteine developing in new adaptations in protein families such as the proteins that are involved in the adhesion process.

In conclusion, cluster statistical analysis of aminoacid compositional attributes of *Toxoplasma* adhesin proteins revealed that single amino acids and dipeptides that included cysteine are common characteristics for this group of proteins. An exhaustive analysis of primary sequence level attributes based on amino acid compositional features could improve the classification and prediction power. Our method provided essential attributes that will be included in the algorithm for learning machine techniques to look at predicting the functional roles of amino acid sequences that are not yet experimentally validated.

Acknowledgement:

This work was funded by Colciencias (Grant 111351928976), a Colombian scientific governmental agency.

References:

[1] Wizemann TM *et al. Emerg Infect Dis*. 1999 **5**: 395 [PMID: 10341176]
[2] Finlay BB & Falkow S, *Microbiol Mol Biol Rev*. 1997 **61**: 136 [PMID: 9184008]
[3] Gómez-Marin JE, *Manual Moderno*. 2010
[4] Tomley FM & Soldati DS, *Trends Parasitol*. 2001 **17**: 81 [PMID: 11228014]
[5] Lawler J & Hynes RO, *J Cell Biol*. 1986 **103**: 1635 [PMID: 2430973]
[6] Naitza S *et al. Parasitol Today*. 1998 **14**: 479 [PMID: 17040860]
[7] Tordai H *et al. FEBS Lett*. 1999 **63**: 67
[8] Nakashima K *et al. J Mol Biol*. 1994 **54**: 61
[9] Ramana J & Gupta D, *PLoS One*. 2010 **15**: e9695 [PMID: 20300572]
[10] Wolting C *et al. BMC Bioinformatics*. 2006 **7**: 338 [PMID: 16836750]
[11] Kunin V & Ouzounis CA, *BMC Bioinformatics*. 2005 **6**: 24 [PMID: 15703069]
[12] Hobohm U & Sander C, *J Mol Biol*. 1995 **251**: 390 [PMID: 7650738]
[13] Brendel V *et al. Proc Natl Acad Sci U S A*. 1992 **89**: 2002
[14] Rand WM, *Journal of the American Statistical Association*. 1971 **66**: 846
[15] Fowlkes EB & Mallows CL, *Journal of the American Statistical Association*. 1983 **78**: 553
[16] Baldi P *et al. Bioinformatics*. 2000 **16**: 412 [PMID: 10871264]
[17] Meissner M *et al. J Cell Sci*. 2002 **115**: 563 [PMID: 11861763]
[18] Miseta A & Csutora P, *Mol Biol Evol*. 2000 **17**: 1232 [PMID: 10908643]
[19] Michon P *et al. Mol Biol Evol*. 2002 **19:** 1128 [PMID: 12082132]
[20] Takahashi N *et al. ProcNatlAcadSci U S A*. 1986 **82**: 1906
[21] Zhang YP *et al. Biophys J*. 1995 **68**: 847 [PMID: 7756552]
[22] Chen ZQ *et al. PLos One*. 2008 **3**: e3611

Edited by P Kangueane
Citation: Arenas *et al*. Bioinformation 8(19): 916-923 (2012)

Supplementary material:

Methodology:

External cluster evaluation

Clustering results were evaluated based on knowing the class labels, which were, in our case, proteins with or without adhesin function. We calculated the Rand index RI **[14]**, the Fowlkes.0-Mallows index FM **[15]** and the Matthews correlation coefficient MCC **[16]**.Each index was calculated as follows:

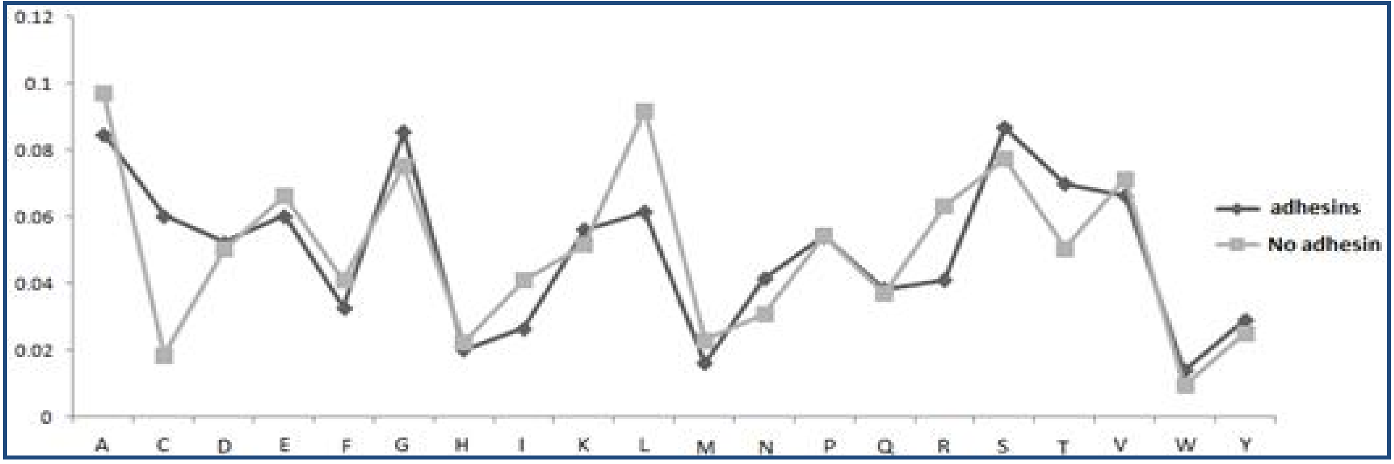
$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

$$FM = \sqrt{\frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}$$

$$MCC = \frac{TP \text{ x } TN - FP \text{ x } FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Supplementary material S1: The relative frequency of each of the 20 amino acids in 50 *Toxoplasma* - *Neospora* and 17 *P.falciparum* - Cryptosporidium adhesins and 250 proteins with no adhesin function. We observed that cysteine and leucine were involved in most of the differences between the two groups.



Supplementary material S2: The relative frequency of 400 dipeptide combinations in 50 *Toxoplasma* - *Neospora* and 17 *P.falciparum* - Cryptosporidium adhesins and 250 proteins with no adhesin function. We observed that each amino acid combined with cysteine (AC, DC, EC, GC, KC, PC RC, SC, and TC) were involved in most of the differences between the two groups.

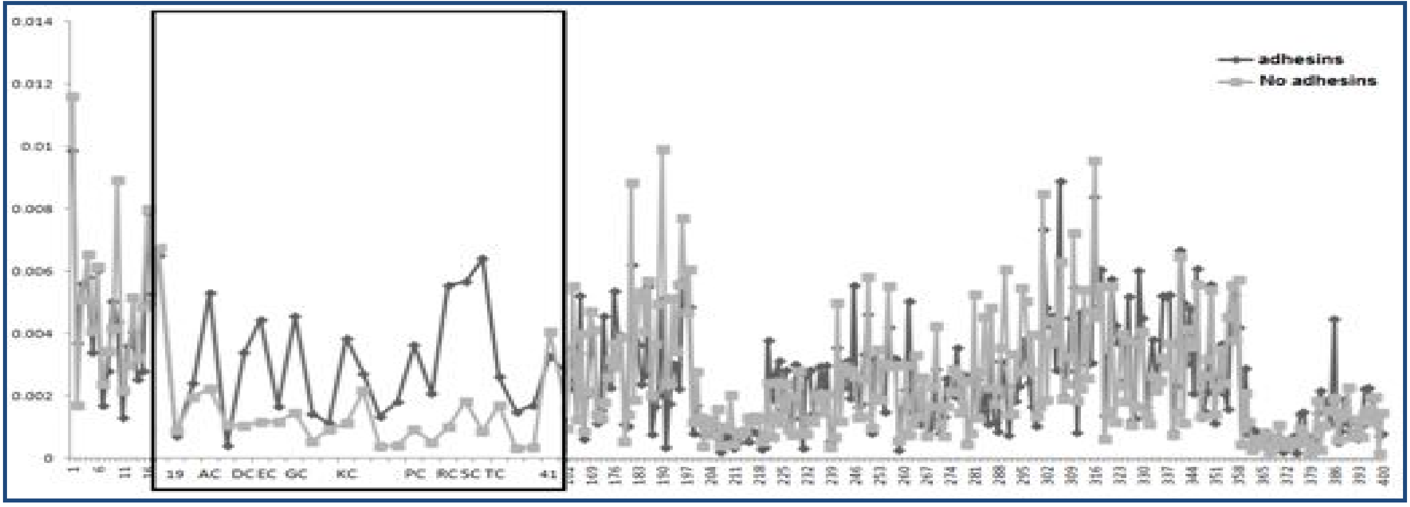


Table 1: Percentage of each of the 20 amino acids in 50 *Toxoplasma* - *Neospora* and 17 *P.falciparum* - *Cryptosporidium* adhesins and 250 proteins with no adhesin function. We observed that cysteine had the most differences between the two groups. * Means: t-Test for the mean differences, with a significance level of (1%) P<0.01.

Amino acid	% adhesin	% no-adhesin
A	8.497224902*	9.900973419
C	6.045344621*	1.661277004
D	5.25053692	6.105769973
E	6.042241469*	9.034318467
F	3.293497963	3.52775801
G	8.550577767	7.378893222
H	2.009254617	2.193342321
I	2.665180434	3.527213641
K	5.609956941	5.631834518
L	6.168125373*	8.526700865
M	1.622759623*	2.164307488
N	4.155321046*	2.729724402
P	5.426616579	4.884755541
Q	3.857550409	4.122343618
R	4.11982123*	6.67064215
S	8.699353022	7.504551763
T	6.989995246*	4.825976449
V	6.654400461	6.531048511
W	1.400208485	1.045167191
Y	2.914045253	1.983979969

Table 2: cluster validation for 3 species of adhesin group proteins (*Toxoplasma gondii – Neospora caninum*), (*Toxoplasma gondii – Neospora caninum – Plasmodium falciparum – Criptosporidium* sp) and (*Toxoplasma gondii* – Human extra-cell receptor), using 3 attributes (single amino acid, dipeptides and amino acids-dipeptides merged). Each species' adhesin group was mixed with 4 different negative sets; we obtained 36 cluster evaluations. A total of 3 indexes were calculated, namely the Rand index, the Fowlkes-Mallows index and the Matthews correlation coefficient, for each cluster.

Toxoplasma - *Neospora* adhesins												
Atribbute	Amino acids-Dipeptides merge				Amino acid				Dipeptide			
FP	1	3	5	1	4	1	5	1	1	3	1	5
TP	45	47	47	46	47	46	47	46	49	47	43	47
FN	4	2	2	3	2	3	2	3	0	2	6	2
TN	48	46	44	48	45	48	44	48	48	46	48	44
Sensitivity	0.9184	0.9592	0.9592	0.9388	0.9592	0.9388	0.9592	0.9388	1	0.9592	0.8776	0.9592
Specificity	0.9796	0.9388	0.898	0.9796	0.9184	0.9796	0.898	0.9796	0.9796	0.9388	0.9796	0.898
MCC	0.8996	0.8991	0.8588	0.9191	0.8783	0.9191	0.8588	0.9191	0.9798	0.8981	0.8616	0.8588
RI	0.949	0.949	0.9286	0.9592	0.9388	0.9592	0.9286	0.9592	0.9898	0.949	0.9286	0.9286
FM	0.9478	0.9495	0.9311	0.9585	0.9402	0.9585	0.9311	0.9585	0.9899	0.9495	0.9261	0.9311
Toxoplasma – Neospora – P.Falciparum - Cryptosporidium adhesins												
FP	0	3	0	3	7	0	2	3	10	4	0	8
TP	54	57	64	57	66	54	64	57	67	65	65	67
FN	13	10	3	10	1	13	3	10	10	2	2	0
TN	67	64	67	64	60	67	65	64	57	63	67	59
Sensitivity	0.806	0.8507	0.9552	0.8507	0.9851	0.806	0.9552	0.8507	0.8701	0.9701	0.9701	1
Specificity	1	0.9552	1	0.9552	0.8955	1	0.9701	0.9552	0.8507	0.9403	1	0.8806
MCC	0.8216	0.8104	0.9562	0.8104	0.8841	0.8216	0.9255	0.8104	0.7209	0.9109	0.9706	0.8869
RI	0.903	0.903	0.9776	0.903	0.9403	0.903	0.9627	0.903	0.8611	0.9552	0.9851	0.9403
FM	0.8978	0.899	0.9774	0.899	0.9437	0.8978	0.9627	0.899	0.8701	0.956	0.985	0.9452
Toxoplasma adhesins - Human extra-cell receptors												
FP	15	11	3	4	2	0	4	3	13	7	4	1
TP	55	48	48	52	52	41	50	40	49	41	43	44
FN	1	8	8	4	4	15	6	16	7	15	13	12
TN	41	45	53	52	54	56	52	53	43	49	52	55
Sensitivity	0.9821	0.8571	0.8571	0.9286	0.9286	0.7321	0.8929	0.7143	0.875	0.7321	0.7679	0.7857
Specificity	0.7321	0.8036	0.9464	0.9286	0.9643	1	0.9286	0.9464	0.7679	0.875	0.9286	0.9821
MCC	0.7377	0.6617	0.8068	0.8571	0.8934	0.7599	0.822	0.6793	0.6466	0.6134	0.7056	0.7831
RI	0.8571	0.8304	0.9018	0.9286	0.9464	0.8661	0.9107	0.8304	0.8214	0.8036	0.8482	0.8839
FM	0.8785	0.8351	0.8982	0.9286	0.9456	0.8557	0.9092	0.8151	0.8316	0.7908	0.8382	0.8765