PLoS NEGLECTED TROPICAL DISEASES

High-Throughput Analysis of Synthetic Peptides for the Immunodiagnosis of Canine Visceral Leishmaniasis

Angélica R. Faria[1], Míriam M. Costa[2], Mário S. Giusta[2], Gabriel Grimaldi Jr.[3], Marcus L. O. Penido[2], Ricardo T. Gazzinelli[2,4,5], Hélida M. Andrade[1]*

1 Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, **2** Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, **3** Instituto Oswaldo Cruz – Fundação Oswaldo Cruz, Rio de Janeiro, Brasil, **4** Centro de Pesquisas René Rachou – Fundação Oswaldo Cruz, Belo Horizonte, Brasil, **5** Division of Infectious Diseases and Immunology, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America

Abstract

Background: Visceral leishmaniasis is the most severe form of leishmaniasis. Approximately 20% of zoonotic human visceral leishmaniasis worldwide is caused by *Leishmania infantum*, which is also known as *Leishmania chagasi* in Latin America, and disease incidence is increasing in urban and peri-urban areas of the tropics. In this form of disease, dogs are the main reservoirs. Diagnostic methods used to identify *Leishmania* infected animals are not able to detect all of the infected ones, which can compromise the effectiveness of disease control. Therefore, to contribute to the improvement of diagnostic methods for canine visceral leishmaniasis (CVL), we aimed to identify and test novel antigens using high-throughput analysis.

Methodology/Principal Findings: Immunodominant proteins from *L. infantum* were mapped in silico to predict B cell epitopes, and the 360 predicted peptides were synthesized on cellulose membranes. Immunoassays were used to select the most reactive peptides, which were then investigated with canine sera. Next, the 10 most reactive peptides were synthesized using solid phase peptide synthesis protocol and tested using ELISA. The sensitivity and specificity of these peptides were also compared to the EIE-LVC Bio-Manguinhos kit, which is recommended by the Brazilian Ministry of Health for use in leishmaniasis control programs. The sensitivity and specificity of the selected synthesized peptides was as high as 88.70% and 95.00%, respectively, whereas the EIE-LVC kit had a sensitivity of 13.08% and 100.00% of specificity. Although the tests based on synthetic peptides were able to diagnose up to 94.80% of asymptomatic dogs with leishmaniasis, the EIE-LVC kit failed to detect the disease in any of the infected asymptomatic dogs.

Conclusions/Significance: Our study shows that ELISA using synthetic peptides is a technique with great potential for diagnosing CVL; furthermore, the use of these peptides in other diagnostic methodologies, such as immunochromatographic tests, could be beneficial to CVL control programs.

Citation: Faria AR, Costa MM, Giusta MS, Grimaldi G Jr, Penido MLO, et al. (2011) High-Throughput Analysis of Synthetic Peptides for the Immunodiagnosis of Canine Visceral Leishmaniasis. PLoS Negl Trop Dis 5(9): e1310. doi:10.1371/journal.pntd.0001310

Editor: Ana Rodriguez, New York University School of Medicine, United States of America

Received May 6, 2011; **Accepted** July 26, 2011; **Published** September 13, 2011

Funding: This work was supported by Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa de Minas Gerais (FAPEMIG) and Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: helida@icb.ufmg.br

Introduction

Leishmaniasis, which is one of the major parasitic diseases recognized by the World Health Organization, affects approximately 1–2 million individuals annually. Dogs are considered the main domestic reservoir of *Leishmania infantum* (also known as *L. chagasi*) [1], which is the causative agent of zoonotic visceral leishmaniasis (VL) in both the Old and New Worlds [2]. In endemic areas, up to 85% of infected dogs may be asymptomatic [3], and they serve as reservoir for vector transmission to susceptible animals and humans [4].

The epidemiological control of VL in Brazil involves the elimination of infected dogs, widespread insecticide use and the systematic treatment of human cases [5]. Reliable diagnostic tests for *L. infantum* detection are essential to prevent disease transmission and the unnecessary culling of dogs. Given the frequency of asymptomatic infections in dogs and the difficulty of direct parasite detection, the development of rapid and accurate indirect diagnostic methods for canine infection is essential for VL surveillance programs. The principal serodiagnostic tests include the immunofluorescent antibody test (IFAT) and the enzyme-linked immunosorbent assay (ELISA). These conventional tests employ crude antigen preparations of either whole promastigotes or their soluble extracts, which limits assay standardization and result reproducibility [6].

An alternative method for the production of antigens for immunoassays is the synthesis of peptides. These peptides are relatively simple to synthesize and are cheaper to produce compared to the production of whole proteins [7]. In general, the use of synthetic peptides increases the specificity of immunoassays compared to crude antigens [8].

Author Summary

Globally, the number of new human cases of visceral leishmaniasis (VL) is estimated to be approximately 500,000 per year. This is the most severe of all forms of leishmaniasis, and the zoonotic form of VL, caused by *Leishmania infantum (*also known as *Leishmania chagasi*), represents 20% of human visceral leishmaniasis worldwide; additionally, its prevalence is increasing in urban and peri-urban areas of the tropics. In Brazil, the identification and elimination of infected dogs, which act as a reservoir for *Leishmania* parasites, is a control measure employed in addition to the use of insecticides against the vectors and the identification and treatment of infected humans. Currently, the diagnostic methods employed to identify infected animals are not able to detect all of these dogs, which compromises the effectiveness of control measures. Moreover, one of the most important issues in controlling VL is the difficulty of diagnosing asymptomatic dogs, which act as parasite reservoirs. Therefore, to contribute to the improvement of the diagnostic methods for CVL, we aimed to identify and characterize new antigens that were more sensitive and specific and could be applied in epidemiologic surveys.

In previous studies, we identified almost 50 immunodominant proteins of *L. infantum*, mapped their B cell epitopes and submitted 180 peptides to Spot synthesis and immunoassay. A total of 25 peptides showed promising characteristics for serodiagnosis of visceral leishmaniasis [9]. Here, we increased the B cell epitopes mapping, performed a high-throughput analysis of 360 peptides and selected the top 10 peptides for ELISA testing. When assessed, the specificity and sensitivity of the selected peptides was as high as 88.7% and 95.0%, respectively. These new antigens represent solid candidate peptides for the diagnosis of VL with great accuracy, especially in asymptomatic animals.

Methods

Ethics Statement

Experiments with dogs were performed in compliance with the guidelines of the Institutional Animal Care and Committee on Ethics of Animal Experimentation (Comitê de Ética em Experimentação Animal – CETEA, national guidelines Lei 11.794, de 8 de outubro de 2008) from Universidade Federal de Minas Gerais (UFMG); protocol 211/07 was approved on 03/12/2008.

Canine Sera

For the initial screening of *Leishmania* antigens on cellulose membranes, we used a pool of sera from ten animals per experimental group, *i.e.*, chronically infected dogs and uninfected control dogs. The chronically infected dogs were naturally infected with *Leishmania*, and they were found in the metropolitan region of Belo Horizonte, Minas Gerais state, Brazil, rescued and maintained in our facility for laboratory and clinical evaluations. VL in chronically infected dogs was certified by the presence of clinical symptoms and parasitological tests that were conducted on bone marrow cells examined by optical microscopy. The uninfected dogs were negative based on parasitological as well as serological tests for VL; these animals served as negative controls in our study. Blood from all the dogs was withdrawn and maintained at room temperature for 3 h to obtain serum. For each animal in a group, 100 µL of serum was deposited in a single tube to obtain a pool of sera that was representative of chronically infected or uninfected dogs; each group was comprised of 10 dogs.

For the ELISA tests, we used the serum described above and 62 serum samples from 23 symptomatic and 39 asymptomatic dogs from Pancas, Espírito Santo state, Brazil. Dogs were scored for 6 typical signs of canine visceral leishmaniasis: alopecia, dermatitis, chancres, conjunctivitis, onychogryphosis and lymphadenopathy. Each sign was scored on a semi-quantitative scale from 0 (absent) to 3 (severe), and these scores were added together to give an overall clinical score. Dogs with a total score of 0-2 were arbitrarily classed as asymptomatic, 3-6 as oligosymptomatic and 7-18 as symptomatic [10]. Additionally, several serum samples from dogs that were experimentally infected with other pathologies were also tested; samples from dogs that were seropositive for *Trypanosoma cruzi* by RIFI (n = 15) and positive for *Leishmania braziliensis* based on parasitological examination and molecular identification (n = 20) were kindly provided by Prof Dr. Ricardo Toshio Fujiwara of UFMG and Prof Dr. Alexandre Barbosa Reis of UFOP, respectively.

Epitope Identification and Spot Synthesis

Previously, we identified almost 50 immunodominant proteins from *L. infantum*, performed the mapping of their B cell epitopes using BepiPred program that is based on propensity scale methods (http://www.cbs.dtu.dk/services/BepiPred/[11]), and a total of 180 peptides were submitted for Spot synthesis and immunoassays, and 25 peptides were shown to be of interest for use in VL serodiagnosis [9]. Here, we completed the mapping of the same proteins [9] using two different programs: ABCPred, based on machine learning methods that apply a recurrent neural network (http://www.imtech.res.in/raghava/abcpred [12]) and BCPreds, which is also based on machine learning methods but involve those that apply a support vector machine (http://ailab.cs.iastate.edu/bcpreds/[13]). Epitopes that were predicted by the two programs simultaneously (excluding those previously identified [9]) were synthesized using the Spot synthesis method [14] on derivatized cellulose membranes with an Ala-Ala linker; peptide size ranged from 9 to 14 amino acids [15].

Immunoassays with cellulose-bound peptides

Initially, the selection of the most immunoreactive peptides was performed using immunoassays of pooled canine serum (previously described in [9]) and alkaline phosphatase-conjugated goat anti-dog immunoglobulin G. We tested all of the peptides mapped in both studies. The relative intensity of the spots representing each peptide was determined by overlapping positive and negative membranes with ImageMaster™ Platinum program. Peptides with relative intensity values of 2 or greater (RI $\geq$ 2) were considered potential candidate antigens [16]. Next, new membranes were synthesized with only the selected peptides, and they were tested with individual canine sera. To evaluate these assays, cut off values were calculated for each peptide using the mean color intensity + 2 standard deviations (SD) from 5 known negative individual canine sera. All assays were performed in duplicate.

Synthesis of soluble peptides

Based on the results of immunoassays conducted using cellulose-bound peptides that were probed with canine sera, 10 peptides that exhibited reactivity with the largest number of serum samples from infected animals were chemically synthesized using 9-florenyl-methoxy-carbonyl (Fmoc) chemistry [17] in an automated synthesizer, model PSSM8 (Shimadzu, Kyoto, Japan). Peptide purity was assessed with reverse-phase high performance liquid chromatography (HPLC) and mass spectrometry (MALDI-TOF-TOFAutoFlexIII™, BrukerDaltonics, Billerica,

Massachussets, USA). The synthetic peptides were diluted in PBS and used as antigens in ELISA assays.

ELISA

All ELISA procedures were optimized in terms of antigen concentrations and the dilutions of serum and conjugated immunoglobulins to develop a reproducible and robust assay. The optimal antigen concentration was 20μg/mL. A clear separation between sera from *L. infantum*-infected and uninfected animals was possible using 1:100 dilutions for sera and 1:5,000 dilutions for the conjugated immunoglobulins.

Falcon flexible microtitration plates purchased from Becton Dickinson Labware Europe (Becton Dickinson, France S.A.) were coated for 16 h approximately with 100 μL/well of synthetic peptides (20 μg/mL) in PBS. Wells were then blocked with 5% powdered skim milk in PBS/T (PBS containing 0.05% Tween20) at 37°C for 1 hour. Serum samples, diluted 1:100 in PBS/T containing 0.5% powdered skim milk, were added and incubated at 37°C for 1 hour. Plates were washed three times with PBS/T and then incubated with peroxidase-conjugated anti-dog immunoglobulin G (Sigma-Aldrich, St. Louis, MO) diluted 1:5,000 in PBS/T containing 0.5% powdered skim milk at 37°C for 1 h. After washing three times with PBS/T, reactions were developed with Fast-OPD^{TM} (Sigma-Aldrich, St. Louis, MO). Plates were incubated for 30 min in the dark. The reactions were stopped with 2 M H_2SO_4, and the plates were read at 492 nm in a Multiskan Plate Reader (MCC/340).

The results of the ELISA using synthetic peptides as antigens (EP) were compared with those obtained with the EIE-LVC Bio-Manguinhos kit, which is based on immunoenzymatic detection of canine visceral leishmaniasis. The EIE-LVC Bio-Manguinhos kit uses crude antigens and is currently recommended by the Brazilian Ministry of Health for the screening of seroreactive animals [18]. To do this comparison, the same serum samples were tested using both assays; the EIE-LVC kits were used according to the manufacturer's instructions and also calculating a different cut off, employing ROC curve and the same control serum samples already described.

Statistical analysis

A cut off point for optimal sensitivity and specificity was determined using ROC analysis [19], and the area under the curve (AUC) was calculated to assess the performance of the tests. All of the statistical analyses were performed using GraphPad Prism^{TM} (version 5.0) and MedCalc^{TM} (version 7.3).

Agreement beyond chance between the tests was assessed using the Cohen Kappa (κ) coefficient [20] and interpreted according to the following scale: 0.00–0.20, negligible; 0.21–0.40, weak; 0.41–0.60, moderate; 0.61–0.80, good and 0.81–1.00, excellent [21]. The accuracy of each test was evaluated according to the AUC referent to the ROC curve and the Youden index J [22].

Results

Epitope identification, Spot synthesis and immunoassays with cellulose-bound peptides

ABCPred and BCPreds programs simultaneously predicted 191 distinct peptides. However, 11 peptides predicted by both programs were previously predicted by BepiPred [9]. All of these peptides (n = 360) were synthesized in cellulose membranes and submitted to immunoassays. Among these 360 peptides, there were 48 with RI ≥ 2, which are presented in Table S1.

All of the 48 selected peptides were synthesized onto new cellulose membranes and subjected to immunoassays with individual canine serum samples obtained from 5 uninfected dogs and 20 *L. infantum*-infected dogs. The pattern of recognition of the various serum samples was similar against the same peptide; furthermore, uninfected serum samples always showed lower reactivity compared to serum samples from infected animals. In this step, peptides that were reactive with pooled sera from *T. cruzi*-infected animals were excluded (data not shown).

Ten peptides were reactive against multiple individual canine serum samples (from 35% to 75% of samples) and also did not cross-react with pooled sera from *T. cruzi*-infected animals (presented in Table S2). Coincidently, these 10 peptides resulted from the simultaneous prediction by two programs (ABCPred and BCPreds). These peptides were then synthesized in a soluble form using solid phase technique to be used in ELISA tests.

ELISA

The antigens (peptides) selected for ELISA testing were named as follows: PSLc1, PSLc2, PSLc3, PSLc4, PSLc5, PSLc6, PSLc7, PSLc8, PSLc9 and PSLc10. All of the peptides were mixed into a single solution (Mix10), which was used as another antigen in ELISA testing.

Peptide sensitivity and specificity were calculated using parasitological results as a gold standard. Most of the peptides were able to detect a large percentage of symptomatic and asymptomatic infected dogs, which was not observed with EIE-LVC kit. Diagnostic performances of the EP and EIE-LVC kit for canine serum samples are shown in Table 1. Based on the accuracy of the EP, those peptides with higher AUC and Youden index J values were selected. Thus, PSLc6, PSLc8 and PSLc10, as well as the Mix10, were tested again with a higher number of serum samples.

The next stage of testing was performed with 107 serum samples from *L. infantum*-infected dogs and the same 20 serum samples that were used as the negative controls in the previous assays. Additionally, 15 serum samples from *T. cruzi*-infected dogs and 20 from *L. braziliensis*-infected dogs were also tested. Figure 1 shows the results obtained with the three selected peptides and the Mix10 tested separately. Some cross-reactivity with *T. cruzi* and *L. braziliensis* occurred for all of the antigens, being cross-reactivity with *L. braziliensis* more frequent. For PSLc6, 40% of *T. cruzi* and 70% of *L. braziliensis*-infected serum samples were considered positive (Figure 1A). PSLc8 exhibited cross-reactivity with 26.6% of *T. cruzi* and 85% of *L. braziliensis*-infected serum samples (Figure 1B), whereas 93.3% of *T. cruzi* and 95.0% of *L. braziliensis*-infected samples reacted with PSLc10 (Figure 1C). Mix10 showed a similar pattern, with 80% of *T. cruzi*-infected serum samples and 70% of *L. braziliensis*-infected serum samples testing positive (Figure 1D).

The same serum samples were tested using the kit recommended by the Brazilian Ministry of Health (EIE-LVC kit). The cut off value obtained as recommended by manufacturer (negative control absorbance multiplied by two) was very high, and was therefore outside the detection range for many of the infected dogs (Figure 2A). Furthermore, when ROC curve was applied using the previous described control serum samples, the cut off obtained was lower, which increased sensitivity, but cross-reactivity occurred with 53.3% of *T. cruzi* infected sera and with 40.0% of *L. braziliensis* infected sera (Figure 2B).

The sensitivity, specificity, AUC and Youden index J values were calculated for the investigated serological tests (Table 2). The sensitivity of the widely used EIE-LVC kit performed according to the manufacturer was 13.08%, but when ROC curve was applied to calculate cut off value, the sensitivity was 87.85%. EP showed sensitivities that ranged between 71.03% and 84.10%, depending

Table 1. Diagnostic performance of EP and EIE-LVC kit in serum samples from symptomatic and asymptomatic dogs.

Antigen/test	Se %	Sp %	AUC	Youden index J	Sym + (%)	Asym + (%)	Total+ (%)
PSLc1	72.50	75.00	0.754	0.475	17 (73.9%)	28 (71.7%)	45 (72.5%)
PSLc2	75.80	70.00	0.769	0.458	16 (69.5%)	31 (79.4%)	47 (75.8%)
PSLc3	74.10	70.00	0.778	0.441	22 (95.6%)	24 (61.5%)	46 (74.1%)
PSLc4	75.80	70.00	0.797	0.458	11 (47.8%)	36 (92.3%)	47 (75.8%)
PSLc5	70.96	55.00	0.642	0.259	16 (69.5%)	28 (71.7%)	44 (70.9%)
PSLc6	79.00	85.00	0.904	0.64	16 (69.5%)	33 (84.6%)	49 (79.0%)
PSLc7	85.40	80.00	0.863	0.654	19 (82.6%)	34 (87.1%)	53 (85.4%)
PSLc8	82.26	95.00	0.944	0.772	17 (73.9%)	34 (87.1%)	51 (82.2%)
PSLc9	79.00	75.00	0.847	0.540	12 (52.1%)	37 (94.8%)	49 (79.0%)
PSLc10	88.70	85.00	0.947	0.737	20 (86.9%)	35 (89.7%)	55 (88.7%)
Mix10	75.81	95.00	0.916	0.708	13 (56.5%)	34 (87.1%)	47 (75.8%)
EIE-LVC kit *	8.00	100.00	NA	0.08	5 (21.7%)	0 (0%)	5 (8.0%)

Samples from symptomatic (n = 23), asymptomatic (n = 39) *L. infantum*-infected dogs and healthy dogs (n = 20) were tested. Se: sensitivity; Sp: specificity; Sym: number of symptomatic dogs diagnose as positive; Asym: number of asymptomatic dogs diagnose as positive;
*cut off obtained according to the manufacturer.
doi:10.1371/journal.pntd.0001310.t001

upon the peptide utilized. The EP Mix10 presented the highest AUC value (0.902), showing a high accuracy [23], followed by the EP PSLc10 with an AUC value of 0.888. The Youden index J values for these two tests were also the highest, with values of 0.585 and 0.641, for the EP Mix10 and EP PSLc10, respectively. Specificity rates were determined for all uninfected dogs. The specificity of the kit was 100%, as a result of the high cut off value. In contrast, the specificity of EP ranged from 55% to 80%, depending on the peptide used.

When the ROC curves obtained from all of the investigated tests with synthetic peptides were combined, it was possible to observe that EP PSLc6 had the lowest AUC value, showing the

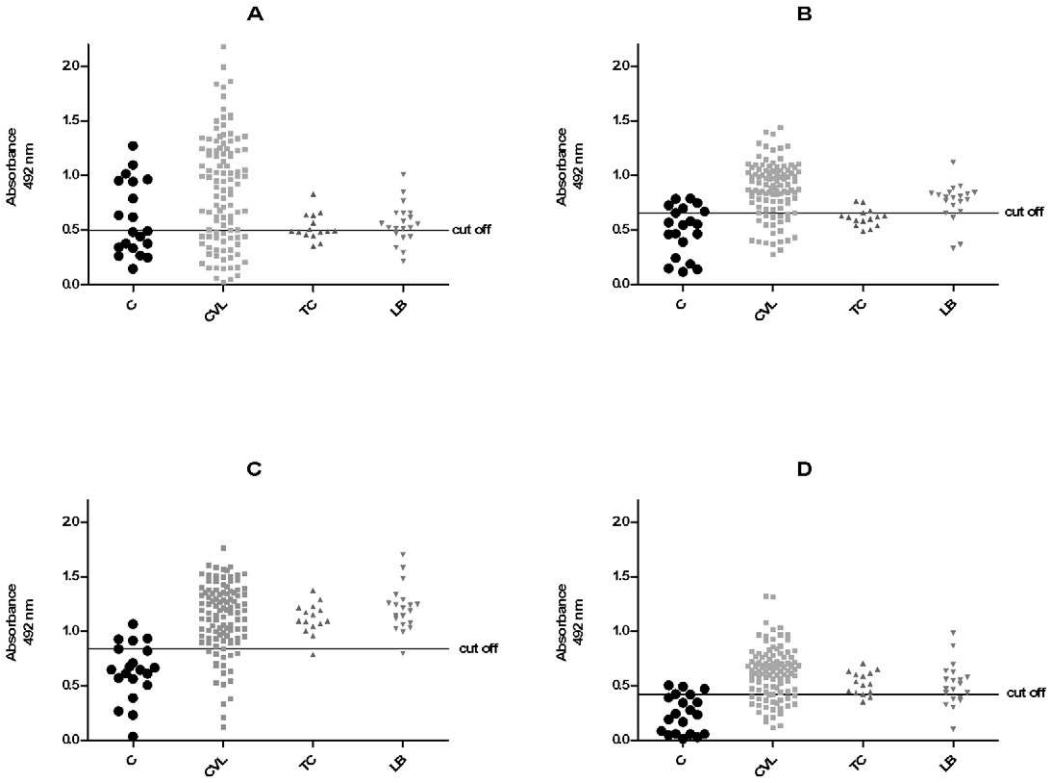


Figure 1. Comparison of ELISA reactivity of canine sera against PSLc6, PSLc8, PSLc10 and Mix10. ELISA was performed in different groups of dogs (C, control group, n = 20; CVL, *L. infantum* group, n = 107; TC, *T. cruzi* group, n = 15; LB, *L. braziliensis* group, n = 20) against PSLc6 (A), PSLc8 (B), PSLc10 (C) and Mix10 (D).
doi:10.1371/journal.pntd.0001310.g001

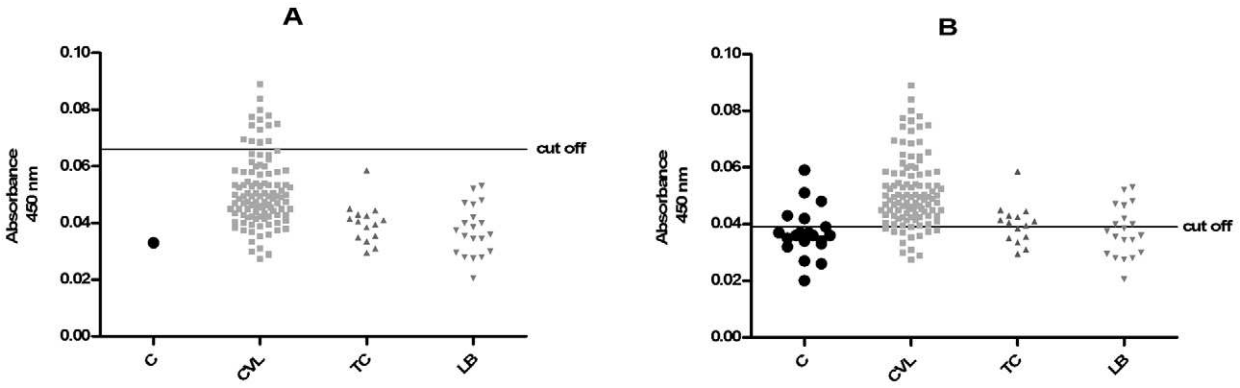


Figure 2. Comparison of ELISA reactivity of canine sera assessed by EIE-LVC kit. ELISA was performed in different groups of dogs (C, control group, n = 1, given by the kit; CVL, *L. infantum* group, n = 107; TC, *T. cruzi* group, n = 15; LB, *L. braziliensis* group, n = 20). In A, cut off was performed according to the manufacturer; in B, it was performed using ROC curve with serum samples from control group.
doi:10.1371/journal.pntd.0001310.g002

most ineffective performance of all of the tests. In contrast, the other tests presented similar performances, showing overlapping ROC curves (Figure 3).

A good agreement beyond chance (κ index) ranging from 0.402 to 0.751 was obtained when the results from each peptide (PSLc6, PSLc8, PSLc10) and Mix10 were cross-tabulated (Table 3). Among these results, the κ index value for PSLc6 and Mix10 was the lowest (0.402), showing weak agreement. The correlation between PSLc8 and Mix10 had the highest κ index value (0.751), showing a good agreement. When the EIE-LVC kit performed according to the manufacturer was cross-tabulated with the synthetic peptides, the agreement was very poor. The index ranged between 0.031 (compared to Mix10) to 0.064 (compared to PSLc8). All of these correlations indicate a negligible agreement between these tests. When EIE-LVC kit performed using ROC curve and 20 control serum samples was cross-tabulated with the synthetic peptides, the agreement was weak. It ranged from 0.213 (compared to PSLc6) to 0.406 (compared to PSLc10).

Discussion

Despite the efforts to search for new diagnostic methods, a method with satisfactory CVL diagnosis efficiency is not yet available. The use of accurate methods that are easy use in the field and are cheap is crucial for diagnosis and consequently, for disease control. Therefore, the identification of new antigens is an important research area for VL disease control. In this work, new antigens for CVL serological diagnosis were investigated using bioinformatic tools and successive screenings with immunoassays.

Using bioinformatics, 360 epitopes were predicted from 47 immunogenic proteins that were identified by bidimensional electrophoresis and Western blot [9]. B cell prediction showed great efficiency because, in the immunoassays, the majority of the cellulose-bound peptides were immunogenic. Of the 48 peptides that could be used to differentiate infected and uninfected dogs, we selected the ten most specific ones.

The ten best peptides resulted from the prediction of two programs simultaneously (BCPreds and ABCPred), and none resulted from using BepiPred considering 2 as a minimum score. However, we observed that 60% of the ten best peptides were also predicted as epitopes based on BepiPred's default score (0.35). These results suggest that using the default score of prediction programs associated with the overlapping predictions by more than one program can be better than using a single type of prediction. Several authors have already shown that combined T cell epitopes generated using consensus predictions are believed to be more accurate [24,25]. It is important to mention that ABCPred and BCPreds approaches are similar (machine learning methods), whereas the BepiPred approach is based on propensity scale methods. This difference could potentially explain the differences between the epitope predictions made by BepiPred alone compared to those made by the ABCPreds and BCPreds programs together.

After the immunoassays in cellulose membranes, the ten peptides that reacted with multiple tested serum samples were

Table 2. Diagnostic performance of EP and EIE-LVC kit in a larger panel of canine serum samples.

Serological test	Sensitivity in CVL group % (95% CI)	Specificity in control group % (95% CI)	AUC	Youden index J
EP PSLc6	71.03 (61.4–79.3)	55.0 (31.5–76.9)	0.661	0.260
EP PSLc8	81.3 (72.6–88.1)	70.0 (45.7–88.1)	0.876	0.513
EP PSLc10	84.1 (75.7–90.4)	80.0 (56.3–94.2)	0.888	0.641
EP Mix10	78.5 (69.5–85.8)	80.0 (56.3–94.2)	0.902	0.585
EIE-LVC kit *	13.08 (7.3–20.9)	100.0 (2.5–100.0)	NA	0.130
EIE-LVC kit #	87.85 (80.12– 93.3)	75.0 (50.9–91.3)	0.848	0.620

Samples from *L. infantum*-infected dogs (n = 107) and healthy dogs (n = 20) were tested. CI: confidence interval; AUC: area under the curve; NA: not applicable;
*cut off obtained according to the manufacturer;
#cut off obtained by ROC curve using 20 control serum samples.
doi:10.1371/journal.pntd.0001310.t002

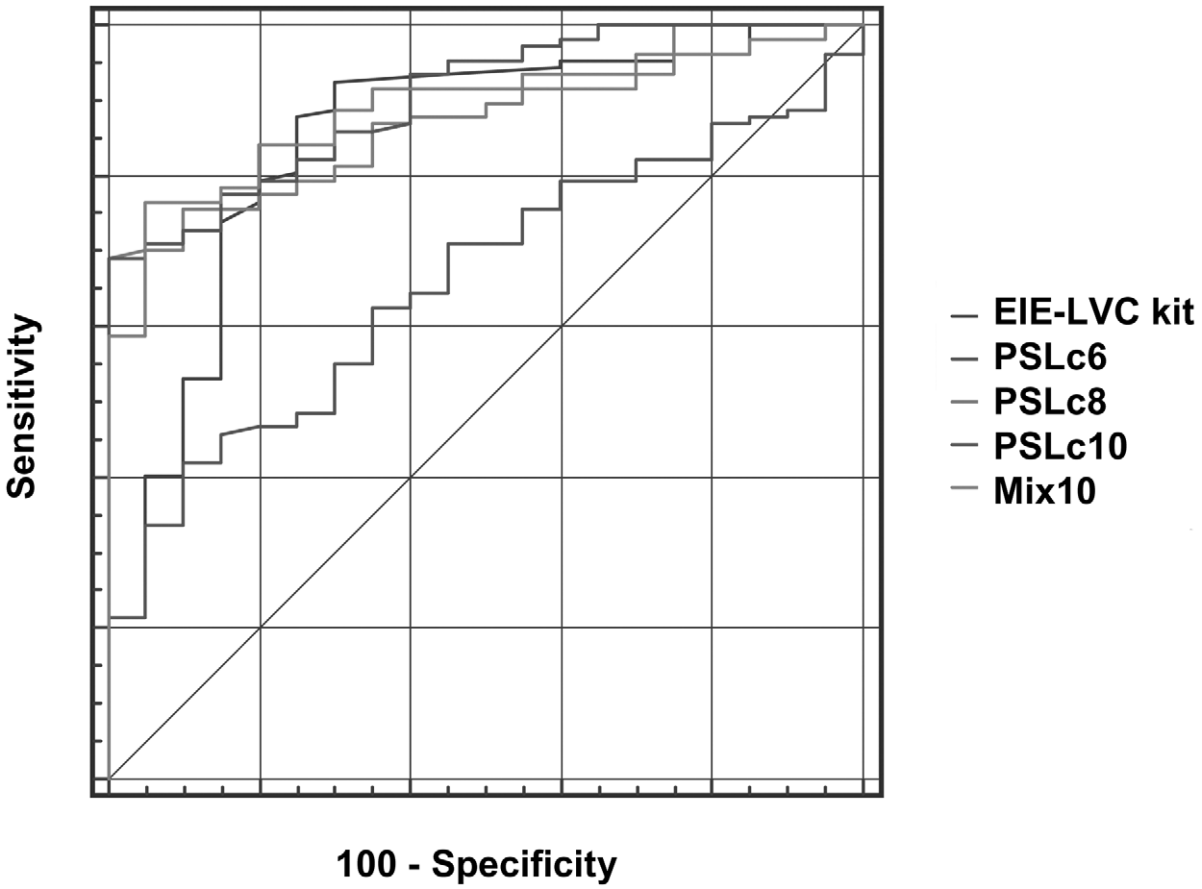


Figure 3. Comparison of ROC curves obtained from all the tests. The curves were used to determine ELISA cut off, sensitivity, specificity and AUC.
doi:10.1371/journal.pntd.0001310.g003

synthesized using Fmoc technique. These peptides showed good reactivity as antigens and were able to recognize specific antibodies in ELISA. The performance observed for the new antigens presented in this work (maximum of sensitivity and specificity of approximately 84.1% and 80.0% respectively) is compatible with the performance of many antigens, mainly recombinant proteins, that have been developed in recent years for CVL diagnosis [26-28] with the advantage of being cheaper and faster to produce.

However, there are other ELISA methods described for CVL that have greater sensitivity or specificity compared to EP. A study conducted in Brazil investigated sera from negative controls (n = 30) and from *L. infantum*-infected dogs (n = 60) in a fucose-mannose ligand based ELISA. Using ROC curve, the sensitivity was 90% and specificity was 93.3% in oligosymptomatic dogs [29]. Another Brazilian study analyzed 209 sera samples from *L. infantum*-infected dogs in a recombinant cysteine proteinase based ELISA. The cut off value was obtained by adding two standard deviation values to the mean absorbance of 22 sera samples from healthy dogs. It provided values of 98% and 96% from sensitivity and specificity respectively [30]. In Europe, some researchers investigated soluble antigens derived from promastigote or amastigote-like stages of *L. infantum* in ELISA. The cut off value was the arithmetic mean plus 3 standard deviations of 48 negative controls. When a group of 47 *L. infantum*-infected dogs was tested, sensitivity varied from 94.1 to 100%, and specificity varied from 96 to 100% [31]. All of these findings show that much has been done to improve CVL diagnosis, but the results are similar. Higher values of sensitivity and specificity are motivations for our antigen improvement.

We compared the results obtained in EP with those obtained with the widely used kit recommended by the Brazilian Ministry of Health, the EIE-LVC kit. Evaluation of EIE-LVC kit, used as recommended by manufacturer, revealed a sensitivity of 13.08% and a specificity of 100%. This low sensitivity could be explained by the high cut off value, which missed many infected dogs. However, following this manufacturer's instruction, test accuracy is related to the composition of only one negative control provided by the kit. Then, we used also 20 serum samples from known uninfected dogs, and the cut off values were obtained using ROC curves. This way, it was possible to obtain sensitivity of 87.8% and specificity of 75.0%. The performance of this kit, when evaluated by other authors is variable. Values of sensitivity of approximately 87.5% and specificity of approximately 100% were obtained elsewhere; 15 false negative results were reported, being 11 in asymptomatic dogs, in a group of 120 samples. Therefore, the prevalence of CVL was underestimated [32]. However, using the same kit, another study showed a sensitivity of 72% and a specificity of 87.5%. These authors indicated the use of EIE-LVC kit in parallel with another kit produced by Bio-Manguinhos, which employed indirect immunofluorescence, to minimize the number of false negatives [33].

Previous data have shown that serologic test performance in CVL depends on infection status [34,35] and an important limitation in CVL control programs is the inability to identify

Table 3. Kappa index (κ) between paired results of diagnostic tests.

Serological test		Serological test														
		EIE-LVC kit *			EP PSLc6			EP PSLc8			EP PSLc 10			EP Mix10		
		P (+)	N (−)	T	P (+)	N (−)	T	P (+)	N (−)	T	P(+)	N(−)	T	P(+)	N(−)	T
EIE-LVC kit #	**P(+)**	14	87	101	73	28	101	81	20	101	84	17	101	76	25	101
	N(-)	0	26	26	12	14	26	12	14	26	10	16	26	12	14	26
	T	14	113	127	85	42	127	93	34	127	94	33	127	88	39	127
κ index (95% CI)		0.062 (-0.049–0.172)			0.213 (0.011–0.415)			0.306 (0.097–0.514)			0.406 (0.208–0.605)			0.245 (0.041–0.450)		
EP Mix10	**P(+)**	11	77	88	70	18	88	84	4	88	81	7	88	-		
	N(-)	3	36	39	15	24	39	9	30	39	13	26	39			
	T	14	113	127	85	42	127	93	34	127	94	33	127			
κ index (95% CI)		0.031 (-0.098–0.161)			0.402 (0.227–0.578)			0.751 (0.622–0.878)			0.613 (0.458–0.769)					
EP PSLc 10	**P(+)**	13	81	94	75	19	94	85	9	94	-			-		
	N(-)	1	32	33	10	23	33	8	25	33						
	T	14	113	127	85	42	127	93	34	127						
κ index (95% CI)		0.060 (-0.061–0.182)			0.455 (0.280–0.629)			0.655 (0.503–0.808)								
EP PSLc8	**P(+)**	13	80	93	75	18	93	-			-			-		
	N(-)	1	33	34	10	24	34									
	T	14	113	127	85	42	127									
κ index (95% CI)		0.064 (-0.059–0.186)			0.477 (0.306–0.648)											
EP PSLc6	**P(+)**	11	74	85	-			-			-			-		
	N(-)	3	39	42												
	T	14	113	127												
κ index (95% CI)		0.041 (-0.094–0.175)														

Samples from *L. infantum*-infected dogs (n = 107) and healthy dogs (n = 20) were tested. P: positive; N: negative; T: total; CI: confidence interval;
*cut off obtained according to the manufacturer;
#cut off obtained by ROC curve using 20 control serum samples.
doi:10.1371/journal.pntd.0001310.t003

asymptomatic dogs because classic diagnostic tests are insufficiently sensitive [36]. Thus, a sensitive and specific antigen for the detection of asymptomatic dogs would be highly desirable because it would allow for effective control intervention in areas where CVL occurs. In our tests, the EIE-LVC kit was not able to detect any of the 39 serum samples from asymptomatic dogs, while the new antigens exhibited strong reactivity to the tested sera. For example, in symptomatic dogs, positive reactions reached up to 95.6% and in asymptomatic dogs, reached up to 94.8%, depending upon the used peptide.

Among the tested peptides, PSLc6, PSLc8, PSLc10 and Mix10 showed the highest accuracies when tested with serum samples from animals with defined clinical statuses. Thus, they were selected to be tested with more serum samples. Then, we observed that EP Mix10 showed the best performance, with an AUC of 0.902, characterizing the test as highly accurate. EP PSLc8 and EP PSLc10 had AUC values that characterized these tests as moderately accurate; only PSLc6 generated a test with low accuracy. The highest value of sensitivity (84.1%) also was obtained when PSLc10 was used as antigen. In a canine epidemiological screening, a test with high sensitivity is desirable.

In addition, all of the EP assays showed a good agreement according to the κ index, when they were cross-tabulated. However, when the EP tests were cross-tabulated with the EIE-LVC kit, the agreement was poor in both situations: as performed according to the manufacturer, it was negligible and as performed with our control serum samples, it was weak. It could be explained mainly by the different antigens employed in the tests. The EP appeared to be more sensitive, once it uses synthetic peptide while the EIE-LVC kit employs crude antigen. Thus, even using the ROC curve, the agreement still remained weak between EIE-LVC kit and EP tests.

The occurrence of cross-reactivity with *T. cruzi* and *L. braziliensis* in EP were observed with all synthetics antigens. These results corroborate other researchers' findings concerning to new antigens in CVL diagnosis [26,27,37,38]. In all of them, some cross-reactivity with these parasites occurred.

Although there has been speculation on the role of dogs in the zoonotic cycle of tegumentary leishmaniasis (TL) caused by *L. braziliensis*, only circumstantial evidence supports this hypothesis [39]. Indeed, the primary reservoirs of *L. braziliensis* are small mammals, particularly wild rodents [40], making the importance of this kind of cross-reactivity in the diagnosis of CVL controversial. Therefore, it could be considered that TL infected dogs have low antibodies levels similar to human TL infections. Some authors described the immune response in human TL infection as predominantly cellular, with low levels of circulating antibodies [41].

Importantly, canine infection with *L. braziliensis* is associated with rural areas. Our test has been developed with the intention of being used in urban surveys, as a diagnostic tool in the control of urban LV. Besides, the prevalence rates of *L braziliensis* in dogs are low (3.1%), when parasitological examination was performed in Brazil [42]. About the prevalence of *T. cruzi* in dogs, few data have been published. In a study conducted with 244 dogs in nine municipalities in Paraná state, Brazil, no dogs were found to be

infected with *T. cruzi* [43]. For this reason, it is difficult to establish the real importance of cross-reactivity with these trypanosomatids in EP tests.

The existence of false positives related to these tripanossomatids (*L. braziliensis* and *T. cruzi*) raises the suspicion of cross reaction with another parasites with higher prevalence in urban dogs, such as *Ehrlichia canis* and *Babesia canis*. Further investigation will be necessary in order to characterize the occurrence of potential cross reactions in EP.

Recent studies have evaluated multiple-epitope chimeric antigens as diagnostic markers for the serodiagnosis of CVL [26,44], which represents an interesting approach to our peptides. The development of an immunochromatographic strip would also be desirable, as it is a common approach in the diagnosis of CVL (rK39 strips) and other pathologies [45–47]. It would be interesting to combine multiple peptides to improve the accuracy, as CHEMBIO has done with the DPP^{TM} immunochromatographic test that employs recombinant antigens K39 and K26 to diagnose CVL. It is feasible to use in the same strip test an antigen with good sensitivity and another with good specificity. Taken together, to improve the diagnostic specificity preventing the unnecessary culling of dogs, some alternatives employing the synthetic peptides should be further investigated. For example, structural changes in the antigens, such as the production of conjugated peptides would be an interesting approach in order to increase both sensitivity and specificity.

In conclusion, we have designed new synthetic peptides for the improved serodiagnosis of CVL. The synthetic peptides named PSLc6, PSLc8, PSLc10 and Mix10 afforded high accuracy in detecting CVL cases and are faster and cheaper to produce. Our findings indicate that synthetic peptides will be useful for serodiagnosis and allow for the detection of asymptomatic dogs. The development of an immunochromatographic test using these peptides would be a valuable tool for the rapid diagnosis of CVL, an important issue for the control of this neglected disease in endemic areas.

Supporting Information

Table S1 Peptides with relative intensity (RI) equal or greater than 2.
(XLS)

Table S2 Reactivity of serum samples to 48 different peptides.
(XLS)

Acknowledgments

The authors want to thank Dr. Maria Norma Melo, Dr. Ricardo Toshio Fujiwara and Dr. Alexandre Barbosa Reis for providing the canine serum samples and Bio-Manguinhos/FIOCRUZ-RJ for providing the EIE-LVC kits.

Author Contributions

Conceived and designed the experiments: ARF MMC RTG HMA. Performed the experiments: ARF MMC MSG MLOP. Analyzed the data: ARF GG RTG HMA. Contributed reagents/materials/analysis tools: GG MSG MLOP RTG HMA. Wrote the paper: ARF RTG HMA.

References

1. Maurício IL, Stothard JR, Miles MA (2000) The strange case of *Leishmania infantum*. Parasitol Today 16: 188–189.
2. Deane LM (1961) Reservoirs of *Leishmania donovani* in Brazil. Rev Assoc Med Bras 7: 161–169.
3. Dantas-Torres F, Brito ME, Brandão-Filho SP (2006) Seroepidemiological survey on canine leishmaniasis among dogs from an urban area of Brazil. Vet Parasitol 140: 54–60.
4. Michalsky EM, Rocha MF, da Rocha Lima AC, França-Silva JC, Pires MQ, et al. (2007) Infectivity of seropositive dogs, showing different clinical forms of leishmaniasis, to *Lutzomia longipalpis* phlebotomine sand flies. Vet Parasitol 147: 67–76.
5. Tesh R (1995) Control of zoonotic visceral leishmaniasis: is it time to change strategies? Am J Trop Med Hyg 52: 287–292.
6. Scalone A, Luna R, Oliva G, Baldi L, Satta G, et al. (2002) Evaluation of the *Leishmania* recombinant K39 antigen as a diagnostic marker for canine leishmaniasis and validation of a standardized enzyme-linked immunosorbent assay. Vet Parasitol 104: 275–285.
7. González L, Boyle RW, Zhang M, Castillo J, Whittier S, et al. (1997) Synthetic-peptide-based enzyme-linked immunosorbent assay for screening human serum or plasma for antibodies to human immunodeficiency virus type 1 and type 2. Clin Diagn Lab Immunol 4: 598–603.
8. Ferrer E, Benitez L, Foster-Cuevas M, Bryce D, Wamae LW, et al. (2003) *Taenia saginata* derived synthetic peptides with potential for the diagnosis of bovine cysticercosis. Vet Parasitol 111: 83–94.
9. Costa MM, Andrade HM, Bartholomeu DC, Freitas L, Pires SF, et al. (2011) Analysis of *Leishmania chagasi* by 2-D Difference Gel Eletrophoresis (2-D DIGE) and Immunoproteomic: identification of novel candidate antigens for diagnostic tests and vaccine. J Proteome Res 10: 2172–84.
10. Quinnell RJ, Courtenay O, Davidson S, Garcez L, Lambson B, et al. (2001) Detection of *Leishmania infantum* by PCR, serology and cellular immune response in a cohort study of Brazilian dogs. Parasitology 122: 253–261.
11. Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2: 2–5.
12. Saha S, Raghava GP (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65: 40–48.
13. El-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recog 21: 243–255.
14. Frank R (1992) Spot-synthesis: an easy technique for the positionally addressable parallel chemical synthesis on a membrane support. Tetrahedron 48: 9217–9232.
15. Frank R, Overwin H (1996) SPOT-synthesis: epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. Methods Mol Biol 66: 149–169.
16. Soutullo A, Santi MN, Perin JC, Beltramini LM, Borel IM, et al. (2007) Systematic epitope analysis of the p26 EIAV core protein. J Mol Recognit 20: 227–237.
17. Atherton E, Bridgen J, Sheppard R (1976) A polyamide support for solid-phase protein sequencing. FEBS Lett 64: 173–175.
18. Alves WA, Bevilacqua PD (2004) Reflexões sobre a qualidade do diagnóstico da leishmaniose visceral canina em inquéritos epidemiológicos: o caso da epidemia de Belo Horizonte, Minas Gerais, Brasil, 1993-1997. Cad Saude Publica, 20: 259–265.
19. Greiner M, Pfeiffer D, Smith RD (2000) Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med 45: 23–41.
20. Cohen J (1968) Weighted kappa: nominal scale agreement with provisions for scales disagreement of partial credit. Psychol Bull 70: 213–220.
21. Landis JR, Koch GG (1977) An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 33: 363–374.
22. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3: 32–35.
23. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240: 1285–1293.
24. Yang X, Yu X (2009) An introduction to epitope methods and software. Rev Med Virol 19: 77–96.
25. Trost B, Bickis M, Kusalik A (2007) Strength in numbers: Achieving greater accuracy in MHC-I binding prediction by combining the results from multiple prediction tools. Immunome Res 24: 3–5.
26. Soto M, Requena JM, Quijada L, Alonso C (1998) Multicomponent chimeric antigen for serodiagnosis of canine visceral leishmaniasis. J Clin Microbiol 36: 58–63.
27. Porrozzi R, Costa MV, Teva A, Falqueto A, Ferreira A (2007) Comparative evaluation of enzyme-linked immunosorbent assays based on crude and recombinant leishmanial antigens for serodiagnosis of symptomatic and asymptomatic *Leishmania infantum* visceral infections in dogs. Clin Vacc Immunol 14: 544–548.
28. Maia C, Campino L (2008) Methods for diagnosis of canine leishmaniasis and immune response to infection. Vet Parasitol 158: 274–287.
29. Cândido TC, Perri SH, Gerzoschkwitz T, Luvizotto MC, Lima VM (2008) Comparative evaluation of enzyme-linked immunosorbent assay based on crude and purified antigen in the diagnosis of canine visceral leishmaniasis in symptomatic and oligosymptomatic dogs. Vet Parasitol 157: 175–181.
30. Pinheiro PH, Pinheiro AN, Ferreira JHL, Costa FA, Katz S, et al. (2009) A recombinant cysteine proteinase from *Leishmania (Leishmania) chagasi* as an antigen for delayed-type hypersensitivity assays and serodiagnosis of canine visceral leishmaniasis. Vet Parasitol 162: 32–39.

31. Mettler M, Grimm F, Capelli G, Camp H, Deplazes P (2005) Evaluation of enzyme-Linked immunosorbent assays, an immunofluorescent-antibody test, and two rapid tests (immunochromatographic-dipstick and gel tests) for serological diagnosis of symptomatic and asymptomatic *Leishmania* infections in dogs. J Clin Microbiol 43(11): 5515–5519.
32. Machado-Coelho GL, Silva MV, Souza DM, Araújo AP, Ferreira WR, et al. (2005) Reprodutibilidade do diagnóstico sorológico da Leishmaniose Visceral Canina utilizando a reação de ELISA em sangue dessecado em papel de filtro. In: XXXI Reunião Anual de Pesquisa Aplicada em Doenças de Chagas e Leishmanioses, Uberaba, Minas Gerais, Brasil.
33. Lira RA, Cavalcanti MP, Nakazawa M, Ferreira AGP, Silva ED, et al. (2006) Canine visceral leishmaniosis: A comparative analysis of the EIE-leishmaniose-visceral-canina-Bio-Manguinhos and the IFI-leishmaniose-visceral-canina-Bio-Manguinhos kits. Vet Parasitol 137: 11–16.
34. Quinnell RJ, Courtenay O, Davidson S, Garcez L, Lambson B, et al. (2001) Detection of *Leishmania infantum* by PCR, serology and immune response in a cohort study of Brazilian dogs. Parasitology 122: 253–261.
35. Reithinger R, Quinnell RJ, Alexander B, Davies CR (2002) Rapid detection of *Leishmania infantum* infection in dogs: comparative study using an immunochromatographic dipstick test, enzyme-linked immunosorbent assay, and PCR. J Clin Microbiol 40: 2352–2356.
36. Dye C, Vidor E, Dereure J (1993) Serological diagnosis of leishmaniasis: on detecting infection as well as disease. Epidemiol and Infect 103: 647–656.
37. Rosário EY, Genaro O, França-Silva JC, da Costa RT, Mayrink W, et al. (2005) Evaluation of enzyme-linked immunosorbent assay using crude *Leishmania* and recombinant antigens as a diagnostic marker for canine visceral leishmaniasis. Mem Inst Oswaldo Cruz 100: 197–203.
38. Romero HD, Silva L, Silva-Vergara M, Rodrigues V, Costa RT, et al. (2009) Comparative study of serologic tests for the diagnosis of asymptomatic visceral leishmaniasis in an endemic area. Am J Trop Med Hyg 81: 27–33.
39. Dantas-Torres F (2007) The role of dogs as reservoirs of *Leishmania* parasites, with emphasis on *Leishmania (Leishmania) infantum* and *Leishmania (Viannia) braziliensis*. Vet Parasitol 149: 139–146.
40. Brandão-Filho SP, Brito ME, Carvalho FG, Ishikawa EA, Cupolillo E, et al. (2003) Wild and synanthropic hosts of *Leishmania (Viannia) braziliensis* in the endemic cutaneous leishmaniasis locality of Amaraji, Pernambuco State, Brazil. Trans R Soc Trop Med Hyg 97: 291–296.
41. Ajdary S, Alimohammadian MH, Eslami MB, Kemp K, Kharazmi A (2000) Comparison of the immune profile of nonhealing cutaneous leishmaniasis patients with those with active lesions and those who have recovered from infection. Infect and Immun 68: 1760–1764.
42. Castro EA, Thomaz-Soccol V, Augur C, Luz E (2007) *Leishmania (Viannia) braziliensis:* epidemiology of canine cutaneous leishmaniasis in the state of Paraná (Brazil). Exp Parasitol 117: 13–21.
43. Falavigna-Guilherme AL, Santana R, Pavanelli GC, Lorosa ES, Araújo SM (2004) Triatomine infestation and vector-borne transmission of Chagas disease in northwest and central Paraná, Brazil. Cad Saude Publica 20: 1191–1200.
44. Boarino A, Scalone A, Gradoni L, Ferroglio E, Vitale F, et al. (2005) Development of recombinant chimeric antigen expressing immunodominant B epitopes of *Leishmania infantum* for serodiagnosis of visceral leishmaniasis. Clin Diag Lab Immun 12: 647–653.
45. Wang Y, Li X, Wang G, Yin H, Cai X, et al. (2011) Development of an immunochromatographic strip for the rapid detection of *Toxoplasma gondii* circulating antigens. Parasitol Int 60: 105–107.
46. Li Y, Hou L, Ye J, Liu X, Dan H, et al. (2010) Development of a convenient immunochromatographic strip for the diagnosis of infection with Japanese encephalitis virus in swine. J Virol Meth 168: 51–56.
47. Zhang J, Guo Y, Xiao Y, Li Z, Hu S, et al. (2010) A simple and rapid strip test for detection of antibodies to avian infectious bronchitis virus. J Vet Med Sci 72: 883–886.

Chagas disease: recombinant *Trypanosoma cruzi* antigens for serological diagnosis

José Franco da Silveira, Eufrosina Setsu Umezawa and Alejandro Ostermayer Luquetti

Diagnosis of individuals infected by *Trypanosoma cruzi* is performed mainly by serological tests using crude antigens, which might crossreact with other infections. In the past ten years, many recombinant *T. cruzi* proteins and synthetic peptides have been described, and some are already on the market. Managers of laboratories and blood banks need to make decisions on a cost–benefit basis whether to include these new-generation tests. Here, we indicate antigens that are likely to prove most useful.

José Franco da Silveira
Dept Micro, Imuno e Parasitologia da Escola Paulista de Medicina, UNIFESP, Rua Botucatu 862, CEP 04023-062, São Paulo, Brasil.

Eufrosina Setsu Umezawa
Instituto de Medicina Tropical de São Paulo, FMUSP, Av. Dr. Enéas de Carvalho Aguiar 470, CEP 05403-000, São Paulo, Brasil.

Alejandro Ostermayer Luquetti*
Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goias, PO Box 131, 74001-970 Goiania, Brasil.
*e-mail: luquetti@hc.ufg.br

Chagas disease affects around 12 million people in the Americas and, owing to migration, infected individuals can be found in nearly every country of the world. Although control of vector transmission has been achieved in at least three countries after the successful Southern Cone Initiative[1], and blood supplies are screened in this area, new cases are still increasing in countries without proper control. Furthermore, infected individuals need to be diagnosed and eventually treated[2] and candidates for blood donation should be adequately screened.

Aetiological diagnosis of American trypanosomiasis, which is required in various circumstances (Table 1), is based on the presence of antibodies against the protozoan parasite *Trypanosoma cruzi* in the serum of infected individuals. These antibodies are usually detected by an array of serological tests, from the outdated complement-fixation reaction to enzymatic immunoassays such as the enzyme-linked immunosorbent assay (ELISA). Of these conventional tests, the most widely used are indirect haemagglutination (IHA), indirect immunofluorescence (IIF) and ELISA, because of their simplicity, low costs and good performance in terms of both specificity and sensitivity. All are based on whole or semipurified antigenic fractions from *T. cruzi* epimastigotes (the noninfective form of the parasite). The WHO[3] recommends using at least two tests in parallel; if both tests are performed by trained technicians with good-quality kits, it is possible to define the status of more than 98% of sera. Nevertheless, in busy, routine diagnostic laboratories, hospitals and blood banks, use of the commercially available diagnostic kits might yield figures well below 98%[4].

Variation in the reproducibility and reliability of these tests has been reported and explained by poor standardization of the reagents[5]. Crossreactivity occurs with antibodies elicited by other pathogens (mainly *Leishmania*). In order to solve these problems, several purified antigens have been described, tested and used in research with good results, but they have not been included in kits for technical and economic reasons. Large-scale production and purification of parasite antigens by classical biochemistry is a very difficult and time-consuming task, and only very small amounts of antigenic components are obtained. Some of these problems have been overcome with the development of recombinant DNA technology, which has led to the construction of various bacterial and eukaryotic gene expression systems that allow the production of parasite antigens in large quantities, with a high degree of purity and standardized quality.

Isolation of *T. cruzi* recombinant antigens

Trypanosoma cruzi antigen genes have been cloned by screening genomic or cDNA expression libraries with sera from chagasic patients or *T. cruzi*-infected animals[6]. Libraries were constructed in phage vector (λgt11 or λZAP) using randomly generated fragments of genomic DNA or cDNA molecules transcribed from mRNAs of epimastigotes or trypomastigotes[6,7]. *Trypanosoma cruzi* recombinant antigens relevant for serodiagnosis have been isolated by several laboratories (Table 2). Although many antigens show sequences that are either identical or very similar to each other, they have been given different names (e.g. CRA, Ag30, JL8 and TCR27). For this reason, identical or similar genes are grouped together in Table 2. Several of the genes have tandemly repeated sequences and so the predicted lengths of the amino acid repeat units are also given.

Comparisons of genes cloned from different *T. cruzi* strains and isolates showed that the sequence of the repeat units is almost identical, indicating that the repetitive domains of these antigens are highly conserved. For instance, antigens FRA, Ag1, JL7 and H49 are built up of repeats of 68 amino acids that are very conserved between strains and isolates of *T. cruzi*[8]. The high frequency with which antigens bearing repetitive domains are isolated could be explained by a high concentration of specific antibodies against repeats in sera from infected individuals and/or by the fact that such antibodies

http://parasites.trends.com 1471-4922/01/$ – see front matter PII: S1471-4922(01)01897-9

Table 1. Serological tests for American trypanosomiasis

Use	Testing laboratory	Requirements	Observations/ pitfalls
Confirmation of etiology in a patient	Diagnostic laboratory	High specificity	Mislabelling of sample may give a wrong result. A false positive result may have serious consequences as rejection for a job and psychological fear of a severe disease
Exclusion of blood from a donor	Blood bank	High sensitivity	A false negative may transmit the infection through blood to the recipient
Epidemiological work (certify area free of infection)	Public health service network	High sensitivity	Crossreaction with leishmaniasis may give false positives
Follow up after etiological treatment	Research laboratory (comparison with antibody concentration before treatment)	High specificity and sensitivity	Need long period of observation. Stored sample of serum for comparison of titres
Check up in immunosuppressed/ AIDS/transplant	Diagnostic laboratory	High specificity	Possibility of reactivation of the infection
Suspected acute phase without detectable parasites	Diagnostic laboratory (assay with IgM conjugates searching for specific IgM)	First priority to search for parasites. If negative results, proceed with IgM search	If positive, should be specifically treated (Ref. 2)
Suspected congenital infection: (1) immediately after delivery, proceed as acute phase. (2) otherwise, recall infant after six months of age	Diagnostic laboratory	Mother's serology should be positive. Search for specific IgG after six months of age	If specific IgG is present after six months of age, the infant should be treated with trypanocidal drugs (Ref. 2)

bind with high affinity. However, this strategy also allowed the identification of nonrepetitive *T. cruzi* antigens such as the ribosomal P protein JL5, the 24 kDa flagellar Ca^{2+}-binding proteins (FCaBP, 1F8, Tc-24, Tc-28), A13, Tc40, heat-shock proteins, flagellum-associated membrane proteins (FL-160, CEA, CRP), and ubiquitin.

Evaluation of the diagnostic potential of *T. cruzi* recombinant antigens

Three multicentre studies have been performed using 17 recombinant antigens[21–23]. The first study[21], undertaken by the WHO, used sera from 50 chagasic and non-chagasic individuals from endemic areas of Central Brazil. Ten out of 17 antigens tested had a kappa index (KI) ≥0.80 and their specificities and sensitivities were 0.86–1.00 and 0.95–1.00, respectively. (The KI measures the agreement of results obtained by a given antigen with reference samples[21]; that is, it can compare results obtained in different laboratories with the same samples. Identical results to the reference are given a KI of 1.0 and a KI >0.80 indicates excellent agreement.) The CRA antigen was found to be the best diagnostic antigen (KI = 1.00), followed by antigens B13 (KI = 0.96) and H49 (KI = 0.92). It is noteworthy that antigens carrying common amino acid repeats presented different KI values. For instance, CRA, Ag30 and JL8 (Table 2) share identical or similar 14-amino acid repeats but had KI values of 1.00, 0.84 and 0.80, respectively[21]. These differences could reflect different protocols used to isolate them and the type of the immunoassay used (radioimmunoassay, ELISA, phage dot blot immunoassay).

In a second study[22], co-ordinated by the Project of Biotechnology of the Science and Technology for Development organization (CYTED), ten recombinant antigens (Ag2, Ag13, SAPA, H49, A13, JL5, JL7, JL8, JL9 and RAI) were examined in a reference laboratory using the phage dot blot immunoassay. Sera from 215 individuals were included in this study: 148 chagasic patients from different endemic areas for American trypanosomiasis in Argentina, Brazil and Venezuela, and 67 non-chagasic subjects. Antigens JL7, H49, Ag2 and A13 were the best diagnostic reagents in this study, with KI values of 0.82–0.93. However, none of these antigens could detect specific antibodies in sera from four chronic chagasic patients out of 148 (2.7%), indicating that single antigens would not be suitable for serological diagnosis.

The third study[23] examined the diagnostic efficiency of six antigens (H49, JL7, JL8, A13, B13 and 1F8) using ELISA on a panel of 541 serum samples (from 304 infected and 237 non-infected individuals) from nine countries in South and Central America (Argentina, Brazil, Bolivia, Chile, Colombia, Venezuela, El Salvador, Guatemala and Honduras). Four antigens (1F8, H49, JL7 and B13) showed high sensitivity (93.4–99.0%). Many individuals living in endemic areas produced specific antibodies against repetitive amino acid antigens (H49, JL7 and B13). Interestingly, the sensitivity (99%) and specificity (99.6%) of the 1F8 antigen were comparable to those of the repetitive antigens, indicating that chronic chagasic patients also display antibodies against non-repetitive antigens. Serum samples from infected individuals reacted with at least one recombinant antigen, suggesting that a mixture of recombinant antigens could detect anti-*T. cruzi* antibodies in all serum samples used in this study. The positivity of a hypothetical antigenic mixture comprising the

Table 2. *Trypanosoma cruzi* recombinant proteins or synthetic peptides with potential clinical and epidemiological use

Antigen[a]	Repeat length (amino acids)	Native protein[b] (kDa)	Remarks	Diagnosis/use	Refs
CRA	14	225	Cytoplasmic antigen	Chronic infections	6
Ag30		180–225			6
JL8		>170			6
TCR27		150–200			6
FRA	68	>300	Cytoskeleton-associated protein	Chronic infections	6
Ag1		205			6
JL7		>170			6
H49		>300			8
B13	12	116–140	Trypomastigote surface protein	Chronic infections	9
Ag2		85			6
TCR39		82			6
PEP-2					10
Ag36	38	85	Microtubule-associated protein	Chronic and acute infections	6
JL9		110			6
MAP					6
SAPA	12	105–205	Trans-sialidases (TS family)	Acute and congenital infections	6
TCNA					6
TS					6
Ag13	5	85		Chronic and acute infections	6
TcD		260			11
B12	20	200–230		Chronic infections	9
TcE	7	35	Ribosomal protein	Chronic infections	12
JL5	None	38	Ribosomal P protein	Cardiac clinical forms	6
A13	None	230		Chronic and acute infections	6
FCaBP	None	24	Flagellar Ca^{2+} binding protein	Chronic infections monitoring of cure	6
1F8		24			6
Tc-24		24			13
Tc-28		28			14
Tc-40	None	38–100		Chronic infections	15
cy-hsp70	None	70	Heat shock proteins	Chronic infections monitoring of cure	6,16
mt-hsp70		70			16
grp-hsp78		78			16
FL-160	None	160	Flagellum-associated surface protein (TS-like family)	Chronic infections monitoring of cure	17
CEA					18
CRP					19
SA85-1.1	None	85	Trypomastigote surface protein (TS-like family)	Chronic infections	17
Ubiquitin	None			Chronic infections	20

Abbreviations: CEA, chronic exoantigen of 160 kDa; CRA, cytoplasmic repetitive antigen; CRP, complement regulatory protein of 160 kDa; cy-hsp70, cytoplasmic heat shock protein of 70 kDa; FCaBP, flagellar Ca^{2+}-binding protein; FL-160, flagellar surface protein of 160 kDa; FRA, flagellar repetitive antigen; grp-hsp78, endoplasmic reticulum heat shock protein of 78 kDa; MAP, microtubule associated protein; mt-hsp70, mitochondrion heat shock protein of 70 kDa; SAPA, shed acute-phase antigen; SA85-1.1, surface protein of 85 kDa; TCNA, *Trypanosoma cruzi* neuraminidase; TS, trans-sialidase.

[a]Several different names were given to identical or similar peptides, and they are grouped together.

[b]Sizes of some native proteins may differ among different *T. cruzi* strains or isolates.

peptides H49 or JL7, B13 and 1F8 was calculated to be 100%. Furthermore, results indicated that one of the major advantages of recombinant ELISA for the serodiagnosis of Chagas disease was the lack of cross-reaction with other parasitic diseases, such as leishmaniasis.

Evaluation of mixtures of recombinant antigens or synthetic peptides

Recombinant antigen mixtures

Serodiagnostic tests with recombinant antigens were improved using a mixture of the antigens CRA and FRA in an ELISA[24–26] (Table 3). The performance of the CRA+FRA mixture was compared with four commercial ELISA kits, IHA and IIF tests using 524 well-defined chagasic and non-chagasic human serum samples from endemic areas of Brazil and from blood donors of the State Blood Bank of São Paulo, as well as 60 serum samples that had given discrepant results with conventional serology[26]. The CRA+FRA mixture showed 98.3% sensitivity and 100% specificity, and no crossreactivity was observed in the recombinant ELISA with 58 sera that were positive for other diseases. The use of the CRA+FRA recombinant ELISA might lead to a reduction of more than 50% in the number of discordant sera[26].

Table 3. Diagnostic performance of serodiagnostic assay using different combinations of *T. cruzi* recombinant proteins or synthetic peptides

Antigens[a]	Assay	Type of antigen	Sensitivity (%)	Specificity (%)	Refs
CRA+FRA mixture	ELISA	Fusion protein (β-galactosidase)	100.0	100.0	24,25
			98.3	100.0	26
Ag1+Ag2+Ag30+SAPA mixture	Enzyme immunoassay (immunodot)[b]	Fusion protein (gluthatione-*S*-transferase)	99.6	99.1	27
FCaBP+hsp70 mixture	ELISA	Fusion protein (His_6-tagged peptide)	97.0	92.3	16
(CRA+FRA+Tc-24+SAPA+MAP+TcD+Ag39)[c]	Line immunoassay[d]	Recombinant proteins and synthetic peptides	100.0	99.3	28
			99.4	98.1	29
TcD+PEP-2 mixture	ELISA	Synthetic peptide	99.7	99.0	34
TcD+Ag2+TcE mixture	Particle gel immunoassay[e]	Synthetic peptide	96.8	94.6	35
TcD+TcE+PEP-2 multi-epitope	ELISA	Linear synthetic peptide	99.6	99.3	12
			100.0	100.0	36
TcD+TcE+PEP-2+TcLo1.2 multi-epitope	ELISA	Branched synthetic tetrapeptide	100.0	ND	12
			100.0	93.3	37
TcD+TcE+PEP-2+TcLo1.2 multi-epitope	ELISA	Linear synthetic peptide	100.0	ND	12
TcD+TcE+PEP-2+TcLo1.2 multi-epitope	ELISA	Linear fusion protein (His_6-tagged peptide)	100.0	ND	12
			100.0	96.6	37

[a]Abbreviations: CRA, cytoplasmic repetitive antigen; FCaBP, flagellar Ca^{2+}-binding protein; FRA, flagellar repetitive antigen; hsp70, heat shock protein of 70 kDa; SAPA, shed acute-phase antigen; ND, not determined.
[b]Enzyme immunoassay (EIA), Dia KitTM Bio-Chagas assay®, Gador S.A., Argentina.
[c]Recombinant proteins or synthetic peptides are fixed individually in a single strip.
[d]Line immuno assay (LIA) Chagas antibody, Innogenetics, Belgium.
[e]Particle gel immunoassay (PaGIA), DiaMed AG, Switzerland.

A new immunodot assay (Dia Kit Bio-Chagas assay®, Gador, Buenos Aires, Argentina) has been developed that uses a mixture of five recombinant antigens (Ag1, Ag2, Ag13, Ag30 and SAPA) coated in a single line onto a reinforced nitrocellulose membrane, together with a second human immunoglobulin G (IgG) control line to monitor the conjugate- and colour-development steps[27]. This study used 995 chagasic and non-chagasic serum samples from Argentina, Brazil and Chile. The test displayed a sensitivity of 99.6% and a specificity of 99.1% (Table 3) but four false positive reactions were observed among 16 sera from patients with visceral leishmaniasis. The combination of recombinant antigens hsp78 (grp78) and FCaBP in ELISA (Table 3) has also been tested with a panel of 176 serum samples from Brazilian chagasic and non-infected individuals[16]. The mixture improved sensitivity in comparison with hsp78 and FCaBP alone (from 90% to 97%), but it also increased the cross-reactivity with serum from patients with cutaneous leishmaniasis (from 3% to 8%).

The loss of specificity found with different antigenic mixtures could be due to the packing of antigen molecules in a limited physical space, which might interfere with the binding of the antibodies. To overcome this problem, a line immunoassay (INNO-LIA® Chagas Ab, Innogenetics, Ghent, Belgium) combining recombinant antigens and synthetic peptides was developed and evaluated with a panel of 1062 serum samples from patients and healthy individuals from four Brazilian regions endemic for American trypanosomiasis[28]. Seven recombinant antigens (CRA, FRA, Ag39, TcD, Tc24, SAPA and MAP) were coated as discrete and independent lines onto a nylon membrane with plastic backing, and the assay gave 100% sensitivity and 99.3% specificity (Table 3). No crossreactivity was found with a set of 40 sera from patients with leishmaniasis. A second study using 1604 serum samples from the State Blood Bank of São Paulo, Brazil, suggested that the INNO-LIA Chagas could be used as a confirmatory test in serological diagnosis[29].

Synthetic peptide mixtures

Synthetic peptides derived from the amino acid sequences of repeated domains of *T. cruzi* antigens (Ag1, Ag2, Ag13, Ag30, Ag36 and SAPA) were used to develop ELISA and immunoradiometric assays[30–32]. Peptides derived from Ag2 and Ag36 reacted with 93% and 65%, respectively, of 60 serum samples from Chilean chronic chagasic patients[32]. By contrast, peptides based on the amino acid repeats of antigens Ag1, Ag13 and Ag30 gave poor results (44–65% sensitivity)[30,32] when compared with those obtained with the fusion proteins (>90% for all antigens). These results suggested that some synthetic peptides cannot mimic the immunodominant epitopes of native antigens. The correct identification of epitopes present in the repetitive domains of *T. cruzi* antigens will help the design of synthetic peptides. The synthetic peptide Ag30 (Refs 30,32) carries only one copy of the main immunodominant B-cell epitope[33]. This could explain why this peptide was only recognized by 60% of chagasic patients, whereas the recombinant protein

carrying many repeats reacts with 99% of sera[9,11,21–25,28,29].

The synthetic peptides PEP-2 and TcD have been tested individually and combined in an ELISA format (Table 3) with 378 serum samples from chagasic and non-chagasic individuals (260 living in an endemic area for Chagas disease in Brazil and 118 healthy individuals and patients with different infectious diseases)[34]. Individually, TcD and PEP-2 peptides gave sensitivities of 93% and 91%, respectively. Combination of the TcD and PEP-2 in a single ELISA significantly improved the sensitivity (99.7%) and specificity (99%) of the test. However, in a separate study[30], the peptide TcD (Table 2) showed a specificity of only 44%, suggesting that the result depends on the format of the test used.

The synthetic peptides Ag2, TcD and TcE were used to develop a particle gel immunoassay (ID-PaGIA®, Diamed, Cressier sur Morat, Switzerland)[35]. When coloured gel particles sensitized with the three peptides are mixed with the specific serum, they agglutinate and can be visualized after centrifugation. The ability of ID-PaGIA to discriminate between negative and positive sera was tested using 111 negative and 119 positive sera collected in four different Brazilian institutions. The sensitivity and specificity of this assay were 96.8% and 94.6%, respectively[35] (Table 3). The assay has the advantages of simplicity of operation and a rapid reaction time (20 min).

Multiepitope antigens

A multiepitope synthetic peptide or recombinant protein carrying *T. cruzi* repeating B-cell epitopes has recently been constructed and evaluated with sera from infected and non-infected individuals from Brazilian regions endemic for Chagas disease and from Ecuador[12,36,37] (Table 3). Consensus positive sera for *T. cruzi* infection from the Centers for Disease Control and Prevention (Atlanta, GA, USA) were also used[12,37]. Initially, the repeat units of antigens TcD, TcE and PEP-2 were combined into a single linear multiepitope synthetic peptide (Table 3). An ELISA based on this multiepitope peptide gave a sensitivity of 99.6% and a specificity of 99.3%.

A fourth repeating epitope, TcLo1.2, was isolated by expression cloning, using a serum sample from a chagasic patient who was negative for reactivity with the linear tripeptide TcD–TcE–PEP-2. TcLo1.2 was further combined into a branched tetrapeptide with TcD–TcE–PEP-2 on one arm and TcLo1.2 in the other. A linear multiepitope recombinant protein (TcLo1.2–TcD–TcE–PEP-2) was also expressed in *Escherichia coli*. An ELISA based on these constructs displayed 100% sensitivity (Table 3). However, the reactivities of the linear tripeptide TcD–TcE–PEP-2 or the branched tetrapeptide were greater than that seen for the individual mix of these three peptides[11,36,37], suggesting that the multiepitope constructs can minimize the problems of competition between different peptides for the solid phase.

Apart from the kits described above, other commercial serodiagnostic kits using recombinant or synthetic *T. cruzi* peptides are currently used: an ELISA based on the FRA+CRA mixture (EIE-Recombinante-Chagas®, BioManguinhos, Rio de Janeiro, Brazil); a rapid qualitative immunochromatographic test based on a combination of recombinant antigens (Chagas Stat-Pak®, Chembio Diagnostic Systems, New York, NY, USA); and a qualitative, membrane-based, immunoassay manufactured using *T. cruzi* and *Leishmania* recombinant antigens (Qualicode Chagas/Leishmania® kit, Immunetics, Cambridge, MA, USA). Some of these kits are still under evaluation.

Diagnosis of the acute phase using recombinant antigens

The serological profile in the first month of the acute phase of Chagas disease corresponds to a classical primary immune response[38]. Specific immunoglobulin M (IgM) appears early in the acute phase of *T. cruzi* infection and can be used in the diagnosis of congenital transmission[38]. The recombinant antigen SAPA reacted with IgM and IgG antibodies of sera from acute and chronic chagasic patients[39]. Anti-SAPA antibodies were detected in 90% of serum samples from acute chagasic patients, and in 10–48.7% of chronic patients[22,30–32,39–42]. SAPA reacted with foetal IgM and IgG antibodies present in the cord blood from *T. cruzi*-infected newborns, and it has been suggested that detection of IgM and IgG with SAPA could be used to distinguish congenitally infected infants from uninfected infants[39].

Conclusions and perspectives

The inclusion of recombinant antigens and synthetic peptides for the serological diagnosis of *T. cruzi* infection has been a clear advance in terms of specificity increase. After the first multicentric studies with these new tools, it became clear that single antigens lack the required sensitivity when compared with conventional tests. There have been several successful attempts to increase sensitivity by the use of cocktails of recombinant antigens (both in mixtures and in different spots), mixtures of synthetic peptides or multiepitope antigens. All the products described, commercially available or not, have a higher specificity than conventional tests, and a number of them require fewer steps and hence are faster. The inclusion of one of these tools is recommended if performed in parallel with one of the conventional tests, mainly IIF or ELISA, which will give the desired specificity (given by a recombinant) and the required sensitivity (given by the crude antigenic preparations).

Acknowledgements
Our work was supported by grants from CYTED (Ibero American Project of Biotechnology), FAPESP, CNPq, International Atomic Energy Agency (IAEA) and FMUSP-LIM49. We apologize to those authors whose work could not be cited directly because of space limitations.

References

1 Anon. (1999) Chile and Brazil to be certified free of transmission of Chagas disease. *TDR News* 59, 10 (http://www.who.int/tdr/publications/tdrnews/news59/chagas.htm)
2 Luquetti, A.O. (1997) Etiological treatment for Chagas disease. *Parasitol. Today* 13, 127–128
3 World Health Organization (1991) *Control of Chagas disease* (*WHO Technical Report Series*), 811, 38–47
4 Salles, N.A. *et al.* (1996) Risk of exposure to Chagas disease among seroreactive Brazilian blood donors. *Transfusion* 36, 969–973
5 Camargo, M.E. *et al.* (1986) Collaboration on the standardisation of Chagas disease in the Americas: an appraisal. *Bull. Pan-Am. Health Org.* 20, 233–244
6 Frasch, A.C.C. *et al.* (1991) Comparison of genes encoding *Trypanosoma cruzi* antigens. *Parasitol. Today* 7, 148–151
7 Watson, J.D. *et al.* (1996) The isolation of cloned genes. In *Recombinant DNA* (Watson, J.D. *et al.*, eds), pp. 99–133, Scientific American Books, W.H. Freeman and Company, New York, USA
8 Cotrim, P.C. *et al.* (1995) Characterisation of a high molecular weight immunodominant antigen associated to the cytoskeleton of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 71, 89–98
9 Gruber, A. and Zingales, B. (1993) *Trypanosoma cruzi*: characterization of two recombinant antigens with potential application in the diagnosis of Chagas disease. *Exp. Parasitol.* 76, 1–12
10 Peralta, J.M. *et al.* (1994) Serodiagnosis of Chagas disease by enzyme-linked immunosorbent assay using two synthetic peptides as antigens. *J. Clin. Microbiol.* 32, 971–974
11 Burns, J.M. *et al.* (1992) Identification and synthesis of a major conserved antigenic epitope of *Trypanosoma cruzi*. *Proc. Natl. Acad. Sci. U. S. A.* 89, 1239–1243
12 Houghton, R.L. *et al.* (1999) A multi-epitope synthetic peptide and recombinant protein for the detection of antibodies to *Trypanosoma cruzi* in radioimmunoprecipitation-confirmed and consensus-positive sera. *J. Infect. Dis.* 179, 1226–1234
13 Krautz, G.M. *et al.* (1995) Use of a 24-kilodalton *Trypanosoma cruzi* recombinant protein to monitor cure of human Chagas disease. *J. Clin. Microbiol.* 33, 2086–2090
14 Abate, T. *et al.* (1993) Cloning and partial characterization of a 28 kDa antigenic protein of *Trypanosoma cruzi*. *Biol. Res.* 26, 121–130
15 Lesenechal, M. *et al.* (1997) Cloning and characterisation of a gene encoding a novel antigen of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 87, 193–204
16 Krautz, G.M. *et al.* (1998) Human antibody responses to *Trypanosoma cruzi* 70-kD heat-shock proteins. *Am. J. Trop. Med. Hyg.* 58, 137–143
17 Centron, M.S. *et al.* (1992) Evaluation of recombinant trypomastigote surface antigens of *Trypanosoma cruzi* in screening sera from a population in rural Northeastern Brazil endemic for Chagas disease. *Acta Trop.* 50, 259–266
18 Jazin, E.E. *et al.* (1995) *Trypanosoma cruzi* exoantigen is a member of a 160 kDa gene family. *Parasitology* 110, 61–69
19 Norris, K.A. *et al.* (1997) Identification of the gene family encoding the 160-kilodalton *Trypanosoma cruzi* complement regulatory protein. *Infect. Immun.* 65, 349–357
20 Telles, S. *et al.* (1999) *Trypanosoma cruzi* and human ubiquitin are immunologically distinct proteins despite only three amino acid difference in their primary sequence. *FEMS Immunol. Med. Microbiol.* 24, 123–130
21 Moncayo, A. and Luquetti, A.O. (1990) Multicentre double blind study for evaluation of *Trypanosoma cruzi* defined antigens as diagnostic reagents. *Mem. Inst. Oswaldo Cruz* 85, 489–495
22 Levin, M.J. *et al.* (1991) Recombinant antigens and Chagas disease diagnosis: analysis of a workshop. *FEMS Microbiol. Immunol.* 89, 11–20
23 Umezawa, E.S. *et al.* (1999) Evaluation of recombinant antigens for Chagas disease serodiagnosis in South and Central America. *J. Clin. Microbiol.* 37, 1554–1560
24 Almeida, E. *et al.* (1990) Use of recombinant antigens for the diagnosis of Chagas disease and blood bank screening. *Mem. Inst. Oswaldo Cruz* 85, 513–517
25 Krieger, M.A. *et al.* (1992) Use of recombinant antigens for the accurate immunodiagnosis of Chagas disease. *Am. J. Trop. Med. Hyg.* 46, 427–434
26 Carvalho, M.R. *et al.* (1993) Chagas disease diagnosis: evaluation of several tests in blood bank screening. *Transfusion* 33, 830–834
27 Pastini, A.C. *et al.* (1994) Immunoassay with recombinant *Trypanosoma cruzi* antigens potentially useful for screening donated blood and diagnosing Chagas disease. *Clin. Chem.* 40, 1893–1897
28 Oelemann, W.M. *et al.* (1999) A recombinant peptide antigen line immunoassay optimized for the confirmation of Chagas disease. *Transfusion* 39, 711–717
29 Saez-Alquezar, A. *et al.* (2000) Serological confirmation of Chagas disease by a recombinant peptide antigen line immunoassay: INNO-LIA chagas. *J. Clin. Microbiol.* 38, 852–854
30 Vergara, U. *et al.* (1991) Assay for detection of *Trypanosoma cruzi* antibodies in human sera based on reaction with synthetic peptides. *J. Clin. Microbiol.* 29, 2034–2037
31 Lorca, M. *et al.* (1992) Immunodetection of antibodies in sera from symptomatic and asymptomatic Chilean Chagas disease patients with *Trypanosoma cruzi* recombinant antigens. *Am. J. Trop. Med. Hyg.* 46, 44–49
32 Vergara, U. *et al.* (1992) Evaluation of an enzyme-linked immunosorbent assay for the diagnosis of Chagas disease using synthetic peptides. *Am. J. Trop. Med. Hyg.* 46, 39–43
33 Pereira, C.M. *et al.* (1998) Mapping of B cell epitopes in an immunodominant antigen of *Trypanosoma cruzi* using fusions to *Escherichia coli* lamB protein. *FEMS Microbiol. Lett.* 164, 125–131
34 Peralta, J.M. *et al.* (1994) Serodiagnosis of Chagas disease by enzyme-linked immunosorbent assay using two synthetic peptides as antigens. *J. Clin. Microbiol.* 32, 971–974
35 Rabelo, A. *et al.* (1999) Serodiagnosis of *Trypanosoma cruzi* infection using the new particle gel immunoassay ID-PaGIA Chagas. *Mem. Inst. Oswaldo Cruz* 94, 77–82
36 Betônico, G.N. *et al.* (1999) Evaluation of a synthetic tripeptide as antigen for detection of IgM and IgG antibodies to *Trypanosoma cruzi* in serum samples from patients with Chagas disease or viral diseases. *Trans. R. Soc. Trop. Med. Hyg.* 93, 603–606
37 Houghton, R.L. *et al.* (2000) Multiepitope synthetic peptide and recombinant protein for detection of antibodies to *Trypanosoma cruzi* in patients with treated or untreated Chagas disease. *J. Infect. Dis.* 181, 325–330
38 Camargo, M.E. and Amato Neto, V. (1974) Anti-*Trypanosoma cruzi* IgM antibodies as serological evidence of recent infection. *Rev. Inst. Med. Trop. São Paulo* 16, 200–202
39 Reyes, M.B. *et al.* (1990) Fetal IgG specificities against *Trypanosoma cruzi* antigens in infected newborns. *Proc. Natl. Acad. Sci. U. S. A.* 87, 2846–2850
40 Lorca, M. *et al.* (1995) Diagnostic value of detecting specific IgA and IgM with recombinant *Trypanosoma cruzi* antigens in congenital Chagas disease. *Am. J. Trop. Med. Hyg.* 52, 512–515
41 Gonzalez, J. *et al.* (1996) Serum antibodies to *Trypanosoma cruzi* antigens in Atacamenos patients from highland of northern Chile. *Acta Trop.* 60, 225–236
42 Breniere, S.F. *et al.* (1997) Immune response to *Trypanosoma cruzi* shed acute phase antigen in children from an endemic area for Chagas disease in Bolivia. *Mem. Inst. Oswaldo Cruz* 92, 503–507

PLoS one

Genomic Analyses, Gene Expression and Antigenic Profile of the Trans-Sialidase Superfamily of *Trypanosoma cruzi* Reveal an Undetected Level of Complexity

Leandro M. Freitas[1☯], Sara Lopes dos Santos[1☯], Gabriela F. Rodrigues-Luiz[1], Tiago A. O. Mendes[1], Thiago S. Rodrigues[2], Ricardo T. Gazzinelli[3], Santuza M. R. Teixeira[3], Ricardo T. Fujiwara[1], Daniella C. Bartholomeu[1]*

1 Departamento de Parasitologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, **2** Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Brazil, **3** Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract

The protozoan parasite *Trypanosoma cruzi* is the etiologic agent of Chagas disease, a highly debilitating human pathology that affects millions of people in the Americas. The sequencing of this parasite's genome reveals that trans-sialidase/trans-sialidase-like (TcS), a polymorphic protein family known to be involved in several aspects of *T. cruzi* biology, is the largest *T. cruzi* gene family, encoding more than 1,400 genes. Despite the fact that four TcS groups are well characterized and only one of the groups contains active trans-sialidases, all members of the family are annotated in the *T. cruzi* genome database as trans-sialidase. After performing sequence clustering analysis with all TcS complete genes, we identified four additional groups, demonstrating that the TcS family is even more heterogeneous than previously thought. Interestingly, members of distinct TcS groups show distinctive patterns of chromosome localization. Members of the TcSgroupII, which harbor proteins involved in host cell attachment/invasion, are preferentially located in subtelomeric regions, whereas members of the largest and new TcSgroupV have internal chromosomal locations. Real-time RT-PCR confirms the expression of genes derived from new groups and shows that the pattern of expression is not similar within and between groups. We also performed B-cell epitope prediction on the family and constructed a TcS specific peptide array, which was screened with sera from *T. cruzi*-infected mice. We demonstrated that all seven groups represented in the array are antigenic. A highly reactive peptide occurs in sixty TcS proteins including members of two new groups and may contribute to the known cross-reactivity of *T. cruzi* epitopes during infection. Taken together, our results contribute to a better understanding of the real complexity of the TcS family and open new avenues for investigating novel roles of this family during *T. cruzi* infection.

Citation: Freitas LM, dos Santos SL, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS, et al. (2011) Genomic Analyses, Gene Expression and Antigenic Profile of the Trans-Sialidase Superfamily of *Trypanosoma cruzi* Reveal an Undetected Level of Complexity. PLoS ONE 6(10): e25914. doi:10.1371/journal.pone.0025914

Editor: Mauricio Martins Rodrigues, Federal University of São Paulo, Brazil

Received July 22, 2011; **Accepted** September 13, 2011; **Published** October 19, 2011

Funding: This study was funded by Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), Instituto Nacional de Ciência e Tecnologia de Vacinas (INCTV), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: daniella@icb.ufmg.br

☯ These authors contributed equally to this work.

Introduction

The protozoan parasite *Trypanosoma cruzi* is the etiologic agent of Chagas disease, a debilitating illness that is a major cause of morbidity and mortality in several Latin America countries. Approximately 10 million people carry the parasite, which causes 10,000 deaths annually [1]. During its life cycle, *T. cruzi* passes through three developmental stages. In its insect vectors, the parasite multiplies as extracellular epimastigotes, and in the hindgut it differentiates into non-dividing trypomastigotes. These infective forms are excreted in the feces after a blood meal and may contaminate the puncture site or mucous membranes of a mammalian host, where they can invade a variety of cell types. Inside host cells, trypomastigotes differentiate into amastigotes, which, after a limited number of cell divisions, differentiate into trypomastigotes that are released into circulation upon host cell rupture. This form can then infect another mammalian host cell or be taken by the insect vector during the blood meal, where it differentiates as epimastigotes.

The ability of *T. cruzi* to survive in the mammalian host is in part due to the presence of a diverse surface membrane coat. In fact, a remarkable feature of the *T. cruzi* genome is the massive expansion of genes that encode polymorphic surface proteins, which include the trans-sialidase and trans-sialidase like superfamily (hereafter called TcS), MASP (mucin-associated surface protein), and TcMUC mucins [2]. The TcS is the largest *T. cruzi* gene family, which has more than 1,400 genes, half of which are apparently functional. One of the most well-studied members of the TcS superfamily is the trans-sialidase (TcTS) enzyme. *T. cruzi* is unable to synthesize sialic acids de novo [3], a sugar modification present in

T. cruzi proteins implicated in several key aspects of the *T. cruzi*-host interaction. The sialylation of the parasite surface is possible due to the activity of a modified sialidase that, instead of hydrolyzing sialic acid, transfers alpha (2–3)-linked sialyl residues from sialoglycoconjugates and proteins from the host to the parasite cell-surface mucin proteins (TcMUC) [4–6]. The rapid sialylation of TcMUC proteins upon cell rupture confers a negatively charged coat that protects the extracellular trypomastigotes from being killed by human anti-alpha galactosyl antibodies [7].

The TcS gene family is highly polymorphic, and only a few members have critical residues necessary for catalytic activity [8]. So far, four groups of TcS have been described based on sequence similarity and functional properties. Group I contains active trans-sialidases, namely TCNA and SAPA (shed acute-phase antigen), and TS-epi proteins expressed in the trypomastigote and epimastigote forms, respectively. Group II comprises members of the gp85 surface glycoproteins TSA-1, SA85, gp90, gp82 and ASP-2, which have been implicated in host cell attachment and invasion. FL-160, a representative of group III, is a complementary regulatory protein that inhibits the alternative and classical complement pathways. TsTc13, whose function is unknown, is the representative of group IV and is included in the TcS superfamily because it contains the conserved VTVxNVxLYNR motif, which is shared by all known TcS members [8–13].

The TcS family was identified in the 1980s and, after the publication of the *T. cruzi* genome [2], no comprehensive analysis of its sequences has been performed. Here, by analyzing all the full-length predicted TcS proteins present in the *T. cruzi* genome, we identified four new groups. The TcS groups were characterized based on presence of key TcS motifs, chromosomal localization, expression profile and antigenic properties. Implications of the TcS diversity for *T. cruzi* biology are discussed.

Materials and Methods

Sequence diversity of the *T. cruzi* TcS family

Genome information and sequences were retrieved from TriTrypDB (http://TriTrypDB.org). Only complete TcS sequences totaling 508 sequences were analyzed. The DNA and the translated sequences were aligned using ClustalW 2.0 software with the default parameters [14]. These alignments were used to calculate the total (mean) nucleotide and protein diversity using MEGA4 [15] with three different methods: p-distance (nucleotide and protein sequences), Kimura-2-parameter (nucleotide sequences) and Poisson correction (protein sequences). The diversity error was estimated using bootstrap resampling with 1,000 replications.

Spatial projection and hierarchical clustering

To identify the clusters formed by the TcS protein sequences and by the 3′ sequences flanking the TcS coding regions (300 nucleotides downstream to the stop codons), we calculated the pairwise distance and generated the distance matrixes. The distances between the sequences were generated using the package PHYLIP [16,17]. To provide a visual representation of each distance matrix, we used the multidimensional scaling (MDS) plot with two dimensions (2D). The K-means method [18] was used to define ten clusters. The MDS, hierarchical clustering, statistical analyses and graphing were performed using the R software platform [19].

TcS cluster distribution on *T. cruzi* chromosomes and protein representation

To define the chromosomal distribution of the TcS groups, we used as reference the genome assembly reported in [20], where pairs of homologous chromosomes were arbitrarily built as having the same size. The chromosomal coordinates of the TcS genes, regardless from each homologous chromosome they are derived, were retrieved from the TriTrypDB (http://TriTrypDB.org) and plotted on the chromosomes. The colors of each coding region were the same as the colors used in the MDS protein clusters. The relative positions in the chromosomes were calculated by dividing the start codon coordinate of each gene by the total length of the chromosome. The values found were used to produce a histogram and to compare the distribution of each cluster and the pseudogenes on the chromosomes.

FRIP coordinates were found using the motif xRxP as a query. Only those occurrences located before the Asp-box and/or closer to the N-terminal extremity were considered. The Asp-box was found using the motif SxDxGxTW as a query, allowing up to 1 mismatch, and the TcS signature motif was searched using the VTVxNVxLYNR sequence as a query. In all query motifs, x represents any amino acid. The motifs were searched using the software PatMatch [21]. The signal peptide and the GPI anchor additional site were predicted using the software SignalP [22] and GPI-SOM [23], respectively. Repetitive sequences were identified using the AA-repeatFinder developed by our group (http://gicab.decom.cefetmg.br/bio-web). Only repeats with more than 10 amino acids were reported. The figures depicting the TcS genome distribution and the protein sequences were constructed using Perl (Practical Extraction and Report Language) scripts and the Bio::Graphics module, part of the Bioperl toolkit (http://www.bioperl.org).

Parasite cultures and RNA extraction

Epimastigotes of the CL Brener clone of *T. cruzi* were maintained in a logarithmic growth phase at 28°C in liver infusion tryptose (LIT) medium supplemented with 10% fetal bovine serum. Amastigote and trypomastigote forms were obtained from infected L6 cells grown in Dulbecco's Modified Eagle Medium supplemented with 5% fetal bovine serum, at 37°C and 5% CO_2, as described [24]. Total RNA was isolated using the RNeasy kit (Qiagen).

Real-time RT-PCR

Primers specific for each cluster were designed using Allele ID 7 (Premier Biosoft, Demo version), and the primer specificity was verified using e-PCR and the entire parasite genome as a template. The primers selected are listed in Table S1. Real-time PCR reactions were performed in an ABI 7500 sequence detection system (Applied Biosystems). Reactions in triplicate were prepared containing 1 mM forward and reverse primers, SYBR Green Supermix (Bio-Rad), and each diluted template cDNA. Standard curves were performed for each experiment for each pair of primers using serially diluted *T. cruzi* CL Brener genomic DNA and were used in the calculation of the relative quantity (Rq) values for each sample. qRT-PCRs for the constitutively expressed GAPDH gene were performed to normalize the expression of the TcS genes. Results were analyzed with an ANOVA test, and graphics were constructed in GraphPad Prism 5.0 (GraphPad Inc.).

Epitope prediction, spot peptide array and immunoblot

The 508 complete TcS proteins were submitted for linear B-cell epitope prediction using the Bepipred algorithm [25]. Peptides with 15 amino acids and with prediction scores above 1.3 were selected. Peptides with 70% identity over 70% of the peptide length with *T. cruzi* proteins other than TcS were excluded. For synthesis, we selected those peptides with higher occurrences within a group and with higher prediction scores. The peptides

synthesized are listed in Table S2. The peptides were covalently synthesized in pre-activated cellulose membranes according to the SPOT synthesis technique [26]. Membranes were blocked with 5% BSA and 4% sucrose in PBS and were incubated for one hour and 30 minutes with diluted mice sera (1:500) in blocking solution. After washing, the membrane was incubated with secondary antibody IgG (Sigma) diluted to 1:2000 in blocking solution and, after a second washing, revealed by *ECL Plus Western blotting* (GE Healthcare). The spots were visualized by fluorescence scanning. The membrane was submitted to the same experimental conditions using sera from uninfected mice. Densitometry measures and analysis of each peptide was performed using Image Master Platinum (GE), and the relative density (Rd) cut-off for positivity was determined as 2.0. Graphics were constructed in GraphPad Prism 5.0 (GraphPad Inc.).

Ethics Statement

All animal procedures were approved by the animal care ethics committee of the Federal University of Minas Gerais (Protocol # 143/2009).

Results

Sequence clustering reveals eight groups of the trans-sialidase/trans-sialidase-like superfamily (TcS) of *T. cruzi*

Despite the fact that four TcS groups were previously described [8,13,27], and only one group corresponds to the active trans-sialidase proteins, a much larger number of members of this gene family was annotated in public databases as trans-sialidases. To sort out which members correspond to the previously defined groups and to eventually identify new groups, we performed cluster analysis on all predicted TcS proteins identified in the CL Brener genome, excluding those annotated as partial and/or pseudogenes. A total of 508 TcS members were used to perform pairwise alignments resulting in a distance matrix that was used to generate a multidimensional scaling (MDS) plot (Figure 1). K-means method was then used to define ten clusters or groups (Figure 1A). Clustering with larger numbers of groups resulted in the fragmentation of previous clusters, without shuffling the members among them, indicating the robustness of the clustering of the family in ten groups (data not shown). Three members were located far from the others in the spatial distribution and therefore are the most divergent members of the family. One of them, Tc00.1047053505699.10, is the only representative of the group shown in black, and the Tc00.1047053509265.120 and Tc00.1047053507699.230 formed the brown group. Manual inspection of these three proteins revealed that their N-terminal regions are longer or shorter compared to the other TcS sequences: Tc00.1047053505699.10 contains an extra 260 amino acids at its N-terminal, whereas Tc00.1047053509265.120 and Tc00.1047053507699.230 have a deletion of approximately 160 and 450 amino acids, respectively, in their N-terminal region. The truncated sequences of these two proteins were due to the location of these genes in contig ends. Because gene prediction regarding the initial start codon could not be corrected for these three anomalous sequences, both black and brown groups were excluded from further analysis. The list of proteins belonging to each group is available in the supporting material (Table S3).

Protein and DNA sequences of the eight remaining groups were then aligned and the intra-cluster diversity was calculated using the p-distance, the Kimura-2-parameter and the Poisson correction methods, as described in the material and methods section. The groups are formed from different numbers of members and show distinct diversity indexes (Table 1). Groups labeled in red and dark green are the largest groups, with 227 and 117 members, respectively, totaling 68% of the TcS members. No clear correlation between the number of members and the diversity indexes was found. For instance, small groups (blue and orange) have similar diversity indexes of the largest ones (Table 1).

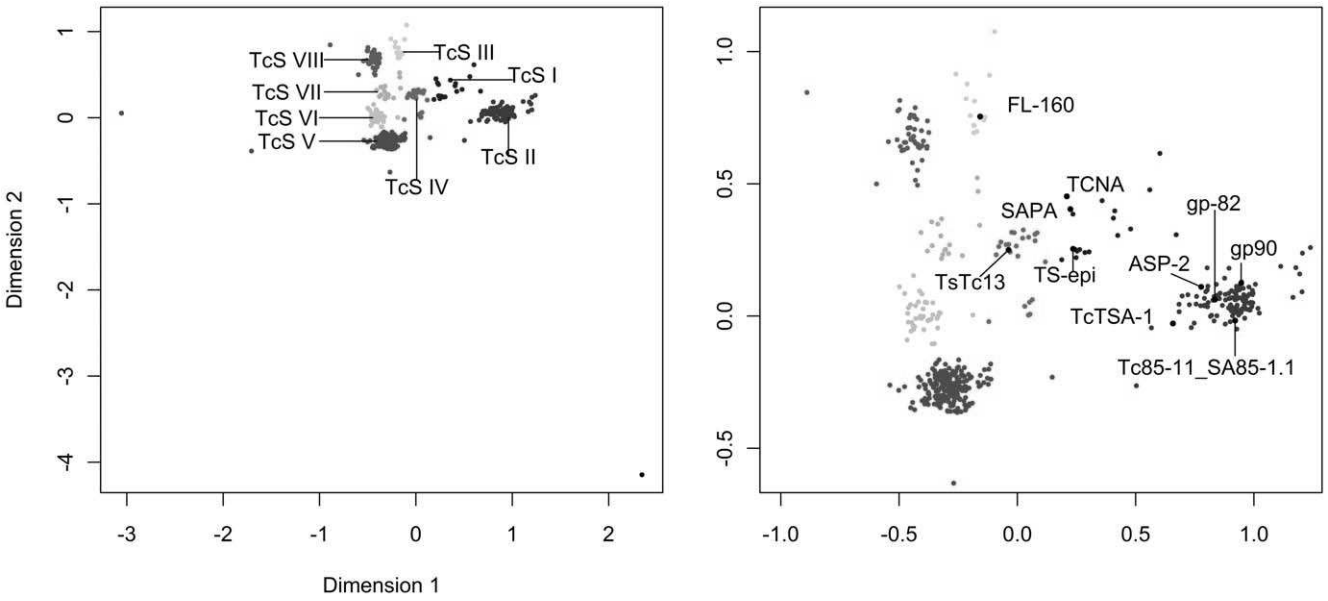


Figure 1. Multidimensional scaling (MDS) plot of the TcS protein sequences. The pairwise alignments of the 508 TcS complete members were performed and the distance matrix was used to generate a multidimensional scaling (MDS) plot. K-means method was used to define the clusters or groups. (A) Pattern of dispersion of all 508 TcS protein sequences resulting in 10 TcS groups. (B) Pattern of dispersion of 505 TcS protein sequences in eight TcS groups. Previously characterized TcS sequences were mapped on the MDS. TcSgroupI - blue; TcSgroupII - dark green; TcSgroupIII - light blue; TcSgroupIV - magenta; TcSgroupV - red; TcSgroupIV - gray; TcSgroupVII - orange and TcSgroupVIII - purple.
doi:10.1371/journal.pone.0025914.g001

Table 1. Diversity indexes of nucleotide, protein and 3′UTR sequences of the TcS family.

		DNA		Protein	
	Number of members	p-distance	K2p	p-distance	Poisson correction
TcSgroupI Blue	19	0.371/0.004	0.690/0.029	0.494/0.009	0.881/0.027
TcSgroupII Dark green	117	0.264/0.004	0.340/0.007	0.419/0.010	0.558/0.018
TcSgroupIII Light blue	15	0.209/0.005	0.263/0.008	0.366/0.010	0.492/0.023
TcSgroupIV Magenta	25	0.179/0.003	0.226/0.005	0.250/0.008	0.320/0.012
TcSgroupV Red	227	0.252/0.004	0.316/0.006	0.396/0.009	0.513/0.015
TcSgroupVI Gray	39	0.246/0.004	0.312/0.007	0.394/0.009	0.513/0.016
TcSgroupVII Orange	17	0.298/0.004	0.425/0.009	0.448/0.009	0.651/0.020
TcSgroupVIII Purple	46	0.215/0.004	0.270/0.006	0.353/0.009	0.453/0.013
TcS family	508	0.413/0.004	0.662/0.011	0.574/0.090	0.912/0.023
3′UTR	495	0.573/0.007	1.086/0.029	-	-

P-distance was used to measure the diversity of the coding regions, proteins and 3′ flanking sequences, with kimura-2-parameters and Poisson correction only to DNA coding and protein sequences, respectively.
doi:10.1371/journal.pone.0025914.t001

We next mapped on the MDS plot the TcS proteins representative from each of the four previously known groups (Figure 1B). As expected, the characterized TcS members mapped into different MDS clusters. TCNA, SAPA and TS-epi, all active trans-sialidase proteins belonging to the previously defined group I, clustered together in the blue group (hereafter named TcSgroupI). From a total of 19 TcSgroupI members, 11 have the critical catalytic residues (Figure S1). GP82, GP90, Tc85-11_SA85-1.1 and ASP-2, all representatives of the previously defined group II, mapped onto the dark green cluster (TcSgroupII). Finally, FL-160 and Ts13, which belong to sialidase groups III and IV, mapped onto the light blue (TcSgroupIII) and magenta (TcSgroupIV) clusters, respectively. None of the TcS proteins previously characterized mapped onto the clusters that are red (the largest TcS group), gray, orange or purple, hereafter named TcSgroup V, VI, VII and VIII, respectively.

Identifying key sialidase signature motifs in the eight MDS clusters

To characterize each of the eight groups, we initially searched for all the key signature motifs as they are described in the literature [8,13] and mapped them into the MDS plot (Figure S2). The canonical VTVxNVxLYNR motif was found in only 328 of 508 TcS sequences used in this study. This result prompted us to investigate whether the other proteins annotated as TcS have a degenerate form of this motif or do not have this motif at all. To this end, we performed ClustalW alignment of all 508 TcS proteins and retrieved the alignment block containing this motif. Visual inspection of this region reveals that 159 sequences have a degenerate version of this motif. Hence, 487 (96%) of the TcS sequences have the canonical or degenerate forms of the VTVxNVxLYNR motif. The remaining sequences do not contain this motif because they have a truncated C-terminal region resulting from premature stop codons and/or frameshifts. Therefore, as previously described, this motif is a signature of the TcS family that is found in all its members. As shown in Figure 2, although variations on the VTVxNVxLYNR motif are observed, the motif is highly conserved within each cluster.

We also searched for the Asp box motif found in bacterial and viral sialidases [28], using as query the SxDxGxTW sequence, where x is any amino acid. A total of 135 sequences have this motif, of which 133 belong to the previously described TcS groups I (blue), II (dark green) and IV (magenta) (Figure 2). The two other sequences having this motif belong to the TcSgroupV (red) and TcSgroupVI (gray). This result is in agreement with previous reports showing that TcSgroupIII (light blue) does not have this motif [29]. To investigate whether TcSgroupIII as well as sequences from the other four new groups have a degenerate form of the Asp box, we searched for a degenerated version of this motif, as described in the material and methods section. This search increased the number of positive Asp box sequences to 383. Only one additional degenerate position was found, resulting in the consensus SxDxGxxW. Although the majority of them have one (220) or two Asp boxes (154), a few (9) have three. Considering this new consensus motif, the Asp box is found in a large majority of the members from TcSgroupI (blue, 17 of 19 members), TcSgroupII (dark green, 114/117), and TcSgroupIV (magenta, 24/25) and is also present in the new groups TcSgroupV (red, 188/227) and TcSgroupVI (gray, 36/39). On the other hand, as previously described, it is missing in TcSgroupIII (lightblue) and has only a few occurrences in the new groups TcSgroupVII (orange, 1/17) and TcSgroupVIII (purple, 3/46) (Figure 2).

The FRIP motif was searched using the pattern xRxP (where x is any amino acid). Because this is a small and degenerate sequence, we considered only those occurrences that are before the Asp-box and/or closest to the N-terminal region [30]. A total of 205 TcS proteins contain the FRIP motif, which is found in the majority of the members of TcSgroupI (blue, 68%), TcSgroupIII

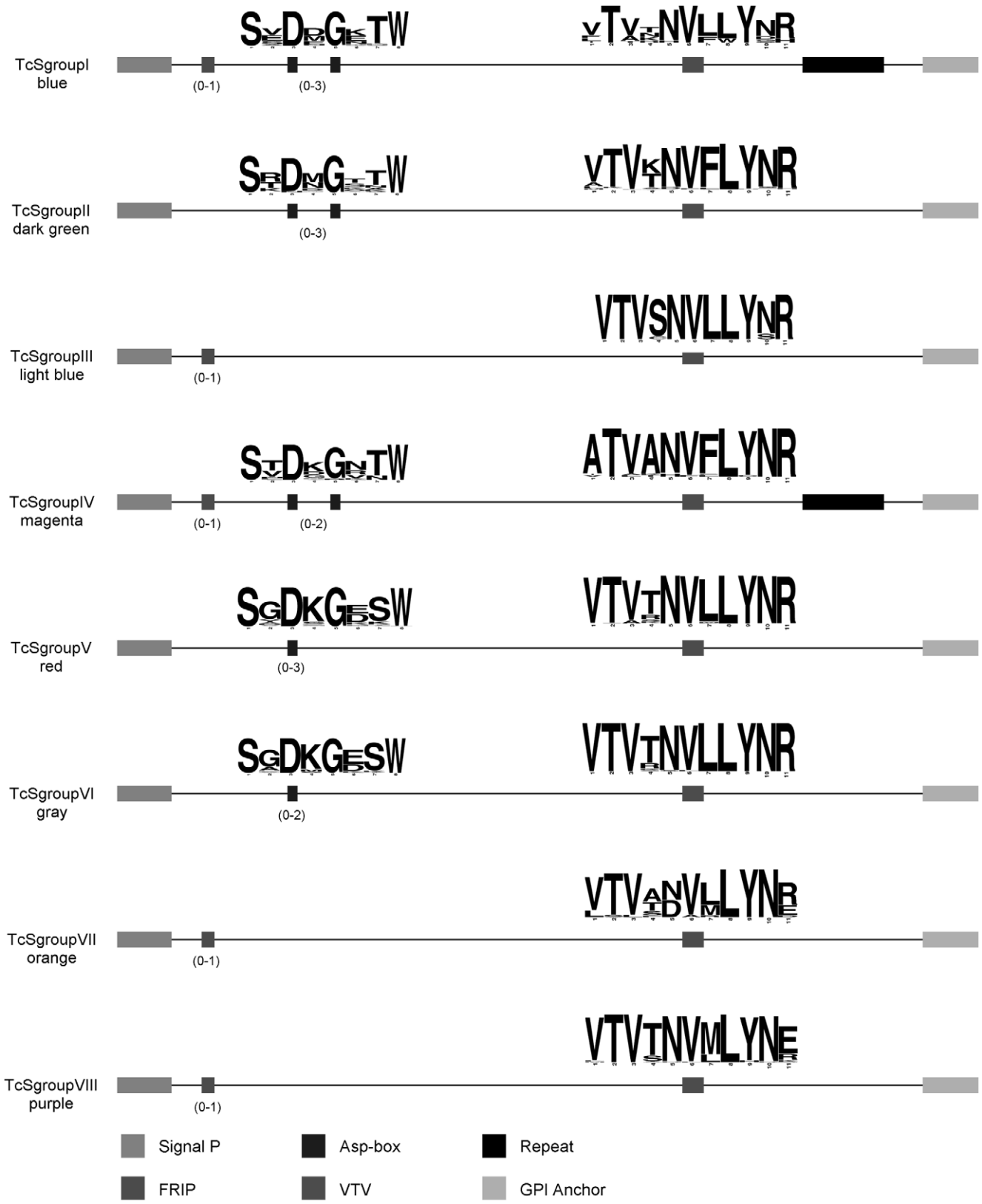


Figure 2. Prototype of each TcS group. The motifs are shown only when they occur in the majority of the proteins within the group. The Asp-box and VTVxNVxLYNR *logos* are shown above each motif. The numbers within parentheses indicate the number of occurrences of a given motif. The length of the proteins within the groups may vary. Graphical representations are not to scale.
doi:10.1371/journal.pone.0025914.g002

(light blue, 87%), TcSgroupIV (magenta, 88%), TcSgroupVII (orange, 76%) and TcSgroupVIII (purple, 87%).

To identify repetitive regions on TcS sequences, we used the AA-repeat finder program (http://gicab.decom.cefetmg.br/bioweb). Only repeats with more than 10 amino acids were considered. We found that repeats are more frequent in the TcSgroupI (blue) and TcSgroupIV (magenta) clusters. These two groups have the largest repetitive regions, which encompass up to 884 amino acids. In fact, although we identified new repeats in these two groups, the largest repeats are those corresponding to the known DSSAH(S/G)TPSTP(A/V) repeat found in TS SAPA and the TcTs13 EPKSA-repeat. On the other hand, TcSgroupV (red), TcSgroupVI (gray) and TcSgroupVII (orange) groups have only 1, 2.5 and 6% of their members, respectively, with repetitive domains whereas no repeat was found in members of the TcSgroupIII (light blue). All repeats identified in this study are shown in Table S4.

In the prototype representation of the eight TcSgroups shown in Figure 2, it is possible to identify three patterns of motif occurrence. The TcSgroupI (blue) and TcSgroupIV (magenta) clusters have the most complex structure, with the FRIP, Asp box and VTVxNVxLYNR motifs and the C-terminal repeats, although the sequences of the VTVxNVxLYNR motif and the C-terminal tandem repeats are distinct. TcSgroupII (dark green), TcSgroupV (red) and TcSgroupVI (gray) clusters contain the Asp box and VTVxNVxLYNR motifs. TcSgroupIII (light blue), TcSgroupVII (orange) and TcSgroupVIII (purple) clusters only have the FRIP and VTVxNVxLYNR motifs, which have a consensus sequence that is group-specific. This pattern of motif occurrence is in agreement with the space distribution of the TcS groups in the MDS (Figure 1). TcS groups I and IV that have all motifs are centered in the MDS, whereas TcSgroups II, V and VI are clustered in the bottom and TcSgroups III, VII and VIII are clustered in the left top region. A graphical representation for each of the 508 TcS proteins can be found in Figure S3.

Mapping the TcS groups on *T.cruzi* chromosomes

It is known that TcS genes can be found in *T. cruzi* subtelomeric regions or in internal positions in the chromosomes that are associated with other genes that encode surface proteins [2]. Subtelomeric regions are defined here as sequences extending from the telomeric hexamer repeats to the first nonrepetitive sequence. We investigated whether there is any bias on the chromosome localization of the TcS clusters. Figure 3 shows the chromosomal distribution of the TcS groups. A total of 60 complete TcS genes (not including partial or pseudogenes) can be found associated with the subtelomeric regions. One of them belongs to the brown cluster, which, as mentioned above, was excluded from our analysis. The majority of the subtelomeric TcS genes (36 members, 61%) belongs to TcSgroupII (dark green), 7 members from TcSgroupIV (magenta) and 10 from TcSgroupVIII (purple) (Figures 3 and 4). No TcSgroupIII (light blue) or TcSgroupVI (gray) genes are located at these regions. Interestingly, with one exception, all members of the largest TcS cluster (TcSgroup V, red) are at internal locations in the chromosomes (Figures 3, 4A and 4B). We have also found that the subtelomeric regions are enriched for TcS pseudogenes (Figure 4C), which is in agreement with the hypothesis that these regions were subject to intense rearrangement [54].

Expression profile of the TcS genes belonging to distinct groups

To characterize the expression profile of TcS genes belonging to distinct groups, we have performed real-time RT-PCR using member-specific primers, designed as described in the material and methods section. The expression of 12 TcS genes derived from six TcS groups was evaluated throughout the three parasite developmental stages using GAPDH mRNA levels, whose expression is constitutive throughout the parasite life cycle, for internal normalization (Figure 5). As a control, we used primers to amplify the cDNAs from the alpha-tubulin and amastin genes, whose mRNA levels we have previously shown to be up-regulated in epimastigotes and amastigotes, respectively [24,31]. The majority of the TcS transcripts are expressed in trypomastigotes and/or amastigote forms. Interestingly, within a group, the expression profile may be highly variable. For example, the TcS5 gene that belongs to TcSgroupII is highly expressed in trypomastigote forms, whereas the TcS27 from the same group shows a much lower level of expression in the trypomastigote and amastigote forms and is barely detected in epimastigotes. Also, TcS9 and TcS33 from TcSgroupIV are more expressed in trypomastigotes and amastigotes; however, TcS34, which is from the same group, is scarcely expressed in all the development stages. The new groups also display a variable expression profile. A very low level of expression was verified for the two genes analyzed from TcSgroupV in all the developmental stages (Figure 5) as well as in the blood trypomastigotes (data not shown). On the other hand, the gene TcS32 from TcSgroupVII is more expressed in the trypomastigotes. The two members of TcSgroupVIII show a variable expression profile, with TcS24 more expressed in trypomastigotes and TcS25 more expressed in amastigotes.

Analyzing sequence conservation of the 3′ flanking region of TcS groups

It is well established that, in Trypanosomatids, the 3′UTR regions are involved in post-transcriptional control mechanisms that confer stage-specific gene expression. To investigate whether the 3′ flanking sequences of TcS genes that belong to the same groups are conserved, we performed pairwise alignments of the 300 nt downstream of the stop codon of the TcSs, and the distance matrix was used to generate the MDS projection. We decided to analyze 300 nt downstream from the stop codon because this is the mean average length of the *T. cruzi* 3′UTRs [32]. The sequences were then color-coded according to the protein clusters showed in Figure 1. TcS genes already characterized as well as those genes whose expression levels were analyzed by real-time RT-PCR (Figure 5) were then mapped onto the MDS projection (Figure 6). We could not find a very clear association between the protein and the 3′ flanking region distances. For example, members of the TcSgroupV (red) form a robust cluster at the protein level and are much more variable according to the analysis of the 3′ flanking region. Also, the 3′ flanking regions of TcSgroupII (dark green) members are scattered in three MDS areas. On the other hand, the 3′ flanking regions of the TcSgroupVIII (purple) members clustered together, which suggests that similar mechanisms may control the expression of some of their genes. Interestingly, the 3′ flanking regions of SAPA and TCNA, both active trans-sialidase enzymes expressed in the trypomastigote forms (TcSgroupI), are clustered very close. Also, the 3′ flanking region of the TS-epi, an active trans-sialidase that is expressed in the epimastigote stage that also belongs to TcSgroupI, is located farther away from the SAPA and TCNA sequences. Moreover, the 3′ flanking region of gp90 and gp82, both expressed in the metacyclic trypomastigotes, and ASP-2, expressed in the amastigote stage, all belong to TcSgroupII and are very close in the MDS projection. Interestingly, although Tc85-11_SA85-1.1 and TsTc13 are expressed in the trypomastigote stage, they are divergent at the protein level (Figure 1) and belong to different TcS groups (II and

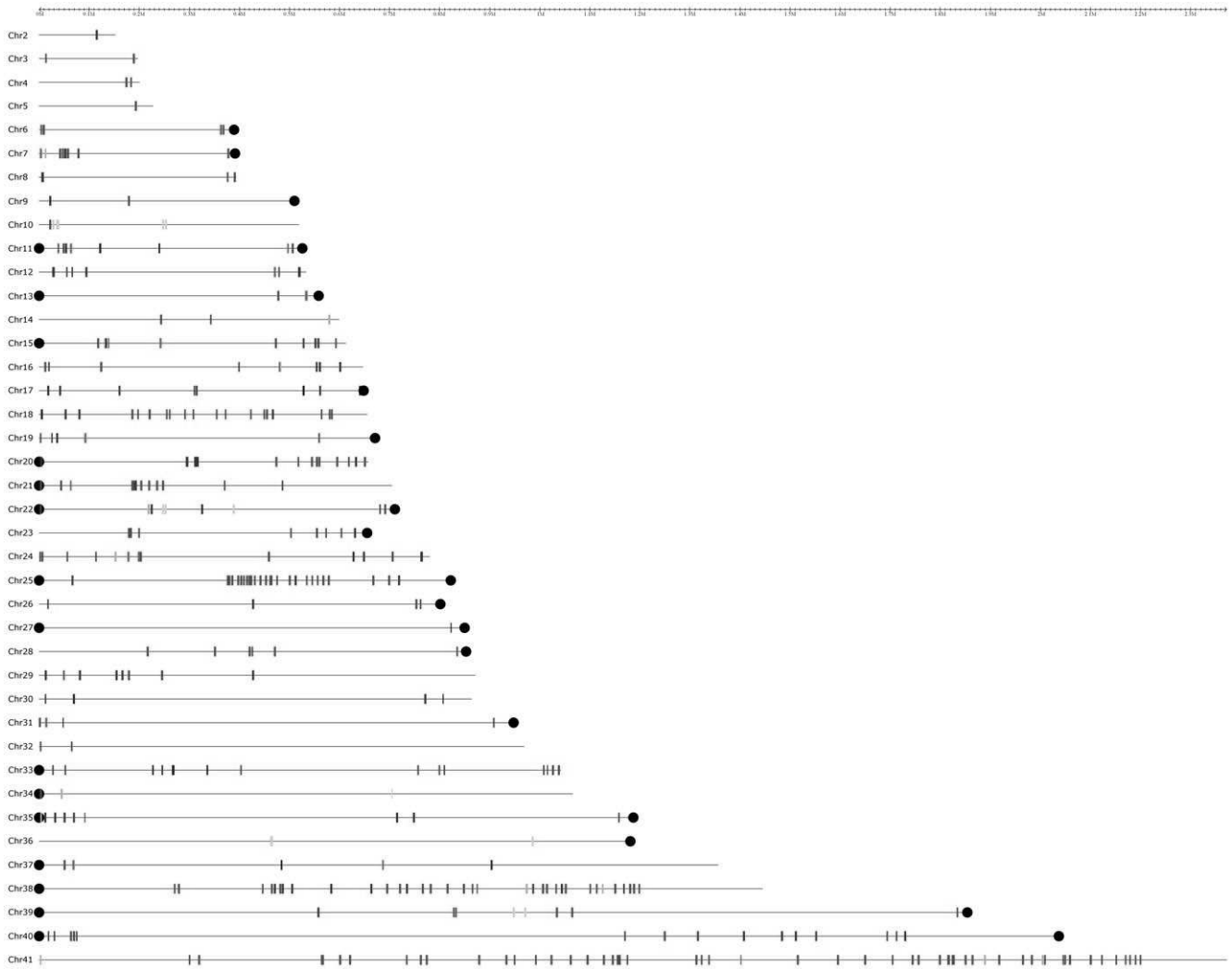


Figure 3. Mapping of TcS genes on *T. cruzi* chromosomes. Each CL Brener chromosome is comprised of 2 homologous chromosomes as proposed by [20]. The genes are color coded according to the color of the corresponding clusters of Figure 1. A total of 374 TcS genes could be mapped on the chromosomes. The remaining genes belong to contigs that could not be assigned to a specific chromosome, according to Weatherly et al., 2009, and are not represented in the figure. Only chromosomes containing TcS genes are shown. Black dots represent telomeric repeats. TcSgroupI - blue; TcSgroupII - dark green; TcSgroupIII - light blue; TcSgroupIV - magenta; TcSgroupV - red; TcSgroupIV - gray; TcSgroupVII - orange and TcSgroupVIII - purple.
doi:10.1371/journal.pone.0025914.g003

IV, respectively); they have similar 3′ flanking regions, which suggests that similar mechanisms for gene regulation may act on both genes.

Antigenicity of the TcS groups

Because the antigenicity of some members of the sialidase family was already reported [33,34], we decided to investigate whether other peptides derived from the TcS family are also antigenic. To this end, we have performed linear B-cell epitope prediction on all 508 complete members of the TcS family. A total of 40 peptides with 15 residues, high prediction scores and high occurrences within the TcS group were synthesized in a solid support by the spot synthesis technique and screened with sera from animals infected with *T. cruzi*. The list of all peptides used in this study is shown in the Table S2. As shown in Figure 7, 11 TcS peptides derived from distinct groups displayed antigenic properties based on a cut-off signal well above background. In agreement with previous studies, peptides corresponding to the SAPA (D5 and D8) [33] and to the TsTc13 repeats (B5) [34] are highly antigenic. We have also identified new epitopes specific to the previously characterized TcSgroups I and IV (D9 and D10, and B10, respectively). At least one peptide from each of the new TcSgroups -V, VI, VII and VIII - was recognized by sera of infected animals (A1, C3, A10 and B4, and A5, respectively). The peptide C3 occurs in the largest number of members (60 in total) from the new TcS groups V and VI and from the previously characterized TcS groups II and IV.

Discussion

The TcS superfamily, the largest *T. cruzi* multigene family [2], was described more than 20 years ago and, after the *T. cruzi* genome release, no comprehensive analysis of the diversity of this gene family was reported. Here, by analyzing all the 508 TcS complete genes present in the *T. cruzi* CL Brener genome [2], we demonstrated that this family displays an even greater variability

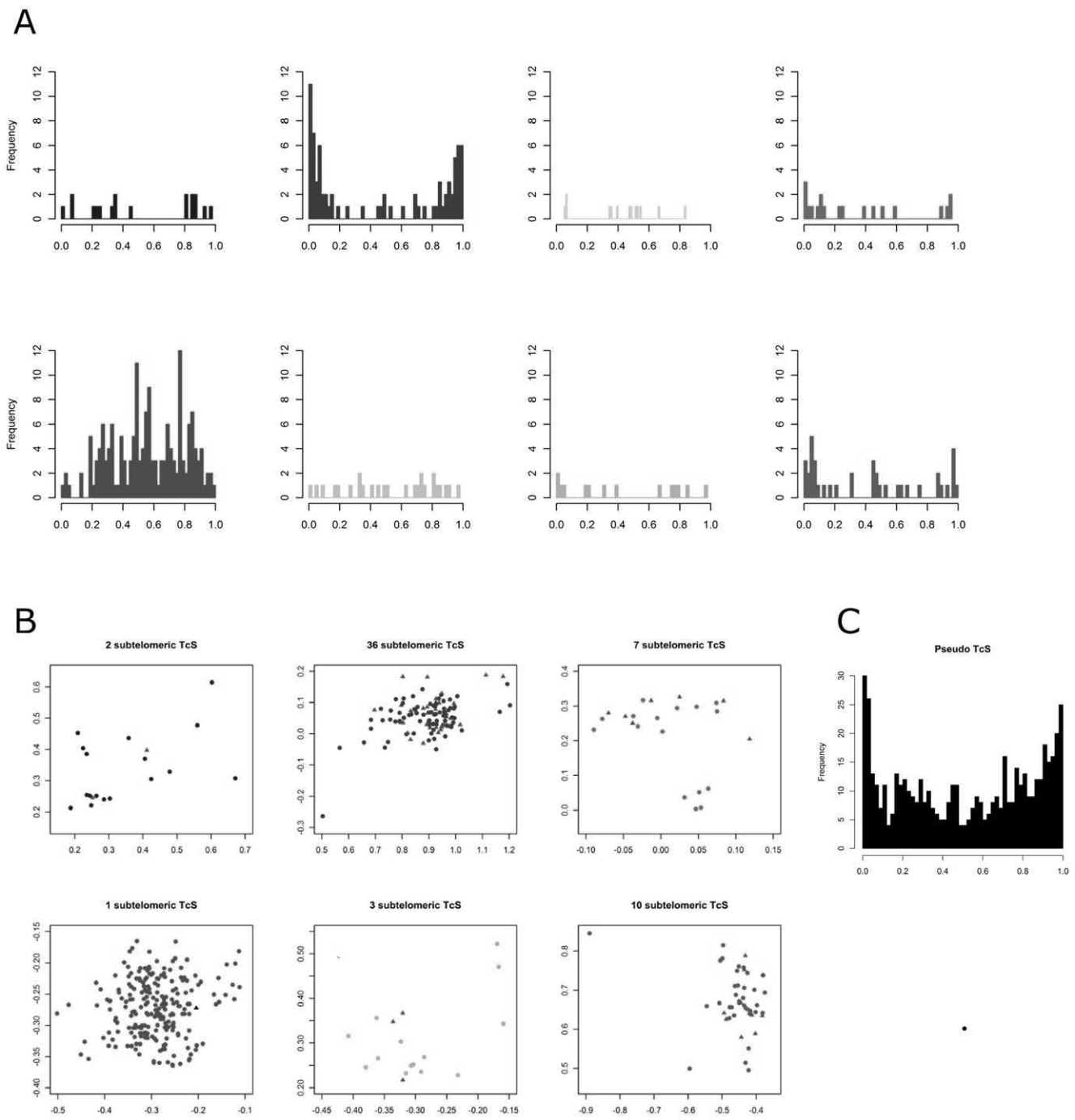


Figure 4. Distribution of each TcS group along the *T. cruzi* chromosomes. (A) Histograms showing the frequency of the TcS genes along the chromosomes. The length of all chromosomes was normalized as 1. The relative position of each gene was calculated by dividing the coordinate of the first nucleotide of the open reading frame by the length of the chromosome. (B) Representation of each group in Figure 1, showing the genes that localize in telomeric regions (black dots for the TcSgroupV and red dots for the other TcS groups). TcSgroupI - blue; TcSgroupII - dark green; TcSgroupIII - light blue; TcSgroupIV - magenta; TcSgroupV - red; TcSgroupIV - gray; TcSgroupVII - orange and TcSgroupVIII - purple. (C) Histogram showing the distribution of TcS pseudogenes along the *T. cruzi* chromosomes.
doi:10.1371/journal.pone.0025914.g004

than previously thought, as shown by means of the diversity indexes and the MDS projection. Based on their pattern of dispersion, we identified eight groups of TcS sequences, four of which were never described before (Figure 1). The distances among the clusters are consistent with the level of similarity and function of the previously described TcS sequences. All proteins that display trans-sialidase activity clustered together (TcSgroupI, blue). Another cluster was formed (TcSgroupII, dark green) from TcS proteins that have no trans-sialidase activity but that are capable of binding to β-galactose, laminin [35], fibronectin [36], collagen [37,38], and cytokeratin [39] and are involved in cell adhesion and invasion. The third TcS group encompasses proteins

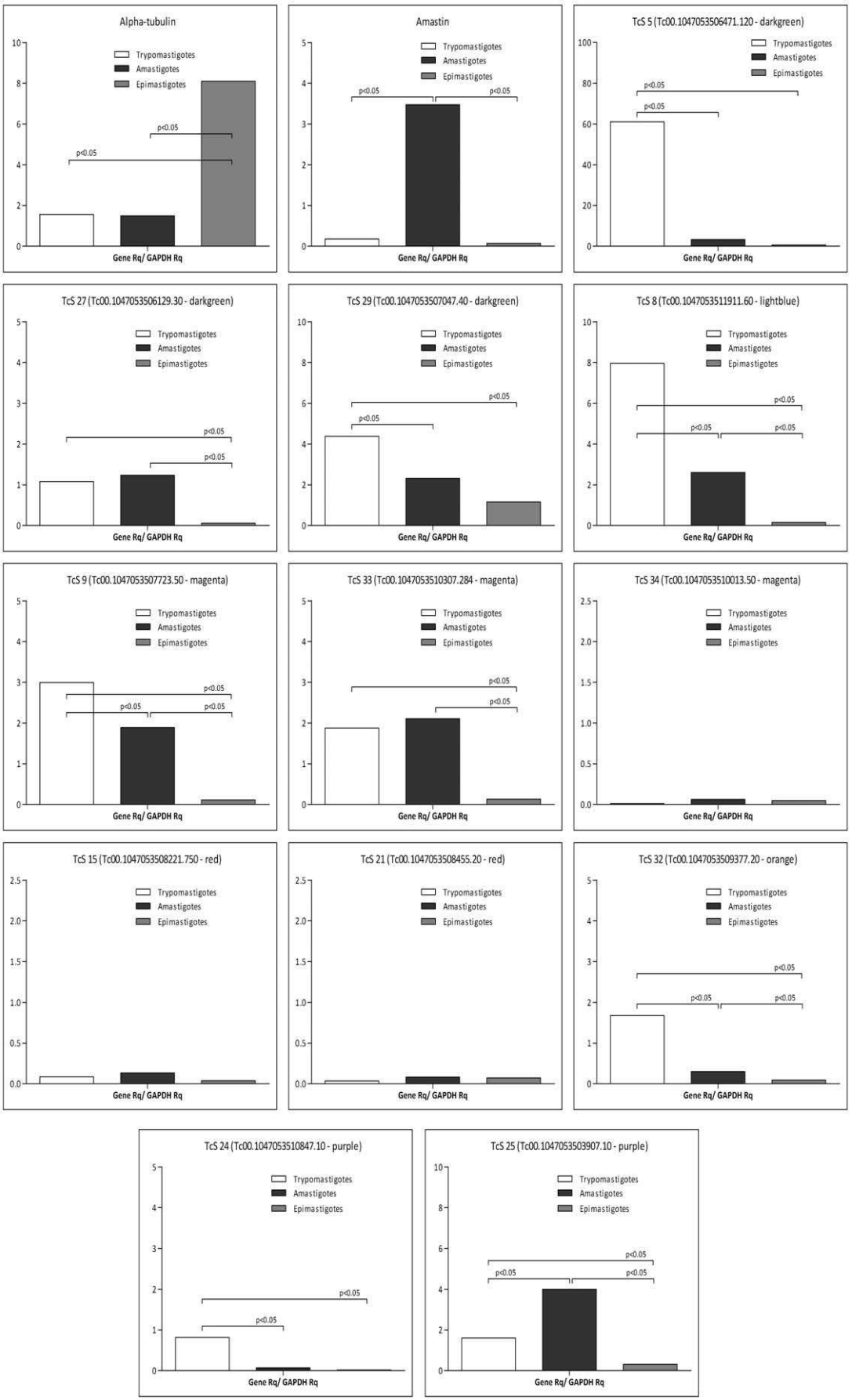

Alpha-tubulin
Amastin
TcS 5 (Tc00.1047053506471.120 - darkgreen)
TcS 27 (Tc00.1047053506129.30 - darkgreen)
TcS 29 (Tc00.1047053507047.40 - darkgreen)
TcS 8 (Tc00.1047053511911.60 - lightblue)
TcS 9 (Tc00.1047053507723.50 - magenta)
TcS 33 (Tc00.1047053510307.284 - magenta)
TcS 34 (Tc00.1047053510013.50 - magenta)
TcS 15 (Tc00.1047053508221.750 - red)
TcS 21 (Tc00.1047053508455.20 - red)
TcS 32 (Tc00.1047053509377.20 - orange)
TcS 24 (Tc00.1047053510847.10 - purple)
TcS 25 (Tc00.1047053503907.10 - purple)
Trypomastigotes
Amastigotes
Epimastigotes
p<0.05
Gene Rq/ GAPDH Rq

Figure 5. Expression profile of TcS genes by qRT-PCR. Relative quantity (Rq) calculations were based on specific standard curves for each TcS gene. Rq values of each cDNA sample (TcS Rq) were normalized with the GAPDH gene (GAPDH Rq), a gene constitutively expressed throughout the parasite life cycle. Alpha-tubulin and amastin were used as controls for genes more expressed in epimastigote and amastigote stages.
doi:10.1371/journal.pone.0025914.g005

involved in the regulation of the complement system (CRP - complement regulatory proteins). Previously characterized members of this group are the CRPs [29,40], which include the FL-160 [41]. Recently, using data from the *T. cruzi* CL Brener genome project, Beucher and Norris (2008) identified CRP paralogs based on sequence similarity with a functional characterized CRP (GenBank accession number AAB49414). Also, these authors divided the CRPs into two groups, HSG (high similarity group, with more than 80% identity with AAB49414) and LSG (low-similarity group, with sequence identity between 54 and 62% with AAB49414) [29]. Here we could verify that all HSGs, and excluding two exceptions, the LSGs, fell into TcSgroupIII (light blue). These two members, which do not belong to TcSgroupIII, were clustered within TcSgroupVII (orange). In fact, they are the two most divergent sequences of the LSG subgroup [29] and correspond to members of the TcSgroupVII that are closest to the TcSgroupIII (Figure S4). Further investigation is necessary to verify whether these two proteins as well as other members of the TcSgroupVII have complement regulatory activity. Finally, a member of the TcSgroupIV that was previously described corresponds to the TsTc13 family, whose function is unknown. Based on the pattern of dispersion of the TcS groups in the MDS projection and the occurrence and sequence of key TcS motifs, we hypothesize that the new groups V and VI and the previously described TcS groupII are more related among each other when compared to the other groups. The same is valid for the new groups VII and VIII and the TcS group III. For instance, TcS groups II, V and VI are the only ones that do not have the FRIP motif and their consensus sequences of the VTVxNVxLYNR motif are very similar. Also, TcS groups III, VII and VIII share the same pattern of motif occurrence and are clustered in a similar region in the MDS projection.

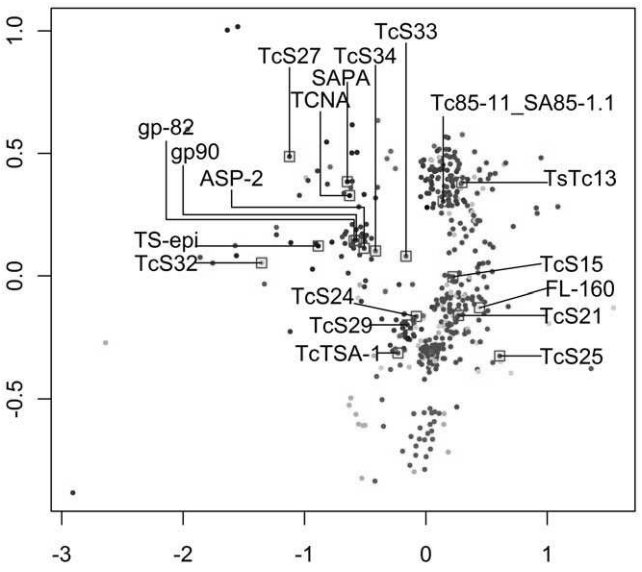
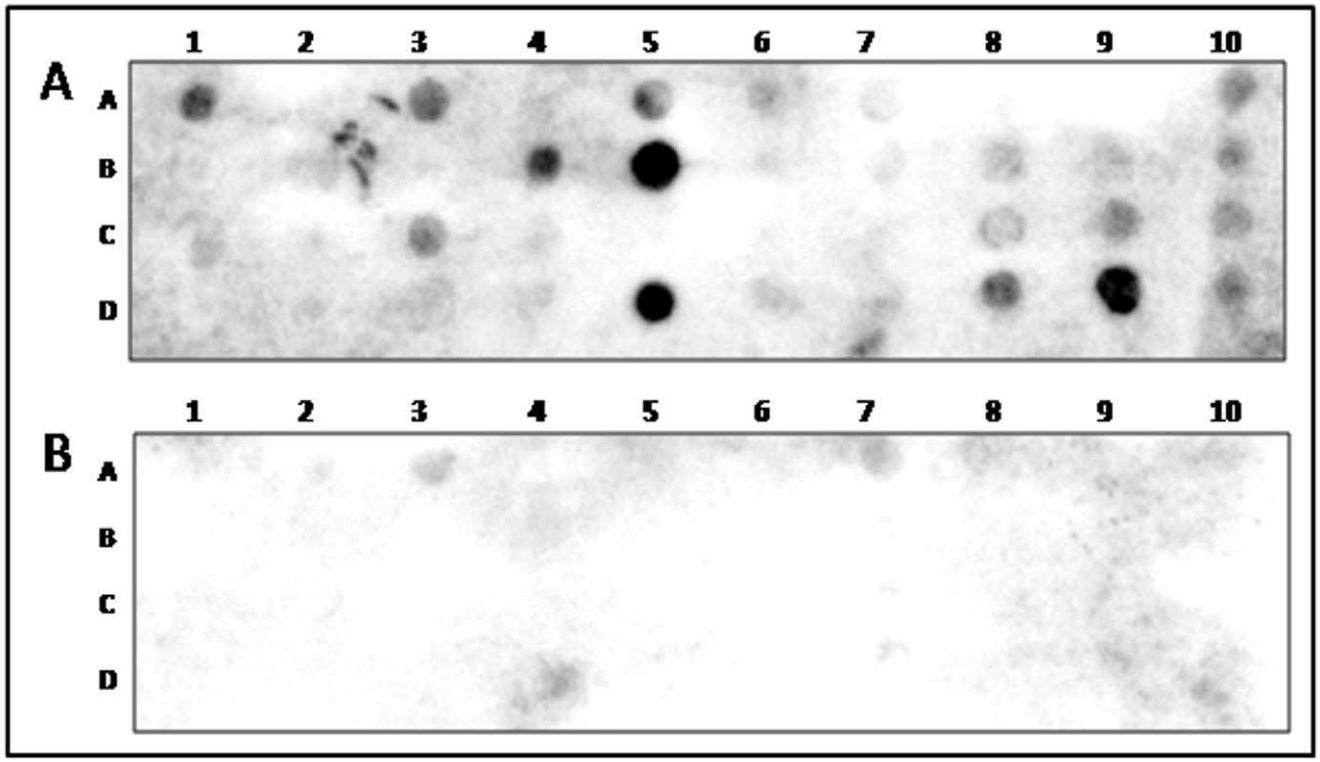


Figure 6. Multidimensional scaling (MDS) plot of the 3′ flanking regions of the TcS genes. A total of 300 nucleotides downstream from the stop codon of each gene were analyzed. Sequences smaller than 300 nucleotides were excluded. Previously characterized genes were mapped on the MDS.
doi:10.1371/journal.pone.0025914.g006

Trypanosoma brucei genome encodes active trans-sialidases expressed in the insect form of the parasite [42]. Although no active trans-sialidase was identified in *Trypanosoma rangeli*, sialidases/sialidase-like proteins similar to TcS groups I, II and III were found, and several of these members are expressed in the epimastigote and trypomastigote forms of the parasite [43,44]. The evolution of the TcS family suggests a gene ancestor encoding an active trans-sialidase expressed in insect forms of the genus *Trypanosoma* and several rounds of duplication and diversification would give rise to trans-sialidases expressed in mammalian forms [45]. Later in evolution *T. rangeli*, would have lost the active trans-sialidase, retaining the sialidase activity. These evidences along with the centered location of TcSgroupI in the MDS projection suggest that extensive expansion and sequence diversification of trans-sialidases similar to TcSgroupI would have originated other groups and functions.

Although the TcS family displays a high degree of sequence variation (Table 1), several motifs are conserved. The most conserved is the VTVxNVxLYNR motif, which is located upstream from the carboxyl terminus of all the TcS full-length members (Figure 2). Recently, it has been demonstrated that a version of this motif (VTVTNVFLYNRPLN), referred to as the FLY motif, may act as a virulence factor [46,47]. BALB/c mice administered with FLY-synthetic peptide are more susceptible to *T. cruzi* infection, displaying increased systemic parasitaemia and mortality [47]. Also, it has been shown that the FLY motif binds to endothelial cells of the heart, suggesting that it might contribute to the parasite tropism to this organ [10]. We identified the exact sequence of the FLY peptide in 28 members of TcSgroupII. Because a very similar version of this motif (**A**TV**A**NV-FLYNRPLN, in which mismatches are indicated in bold and are underlined) is also found in 23 members of TcSgroupIV, we speculate that, as several TcSgroupII members, this group may also participate in host cell attachment/invasion.

Two other motifs, FRIP (xRxP) and Asp box, can be found in various groups of the TcS family. The FRIP motif, which is closest to the N-terminal, is involved in binding the carboxylate group of sialic acid [48]. This motif is found not only in TcSgroupI, but also in the majority of the members of the TcS groups III, IV, VII and VIII (Figure 2). Although this motif is involved in binding sialic acid, it has been shown that enzymatically inactive members of the sialidase family in *T. cruzi* still preserve carbohydrate binding properties [49,50]. The Asp box follows the FRIP motif and can be repeated up to five times in the sequences of viral, bacterial, trypanosomatid and mammalian sialidases. Although its function is unknown, it is worth noting that the Asp box occurs in secreted proteins and in proteins that act on, or interact with, carbohydrates [51]. Recently, it has been shown that at least some inactive trans-sialidases act as lectin-like proteins able to interact with the carbohydrate portion of glycoconjugates, only if they are sialylated [52]. The authors hypothesized that these inactive trans-sialidase proteins could bind to host surfaces that are rich in sialyl-donor glycoconjugates (functioning as anchors), facilitating the active enzyme to more efficiently undertake the sialyl-transferring activity. Here, we have shown that, in addition to TcSgroupI, members of TcSgroupIV have both FRIP and Asp box motifs



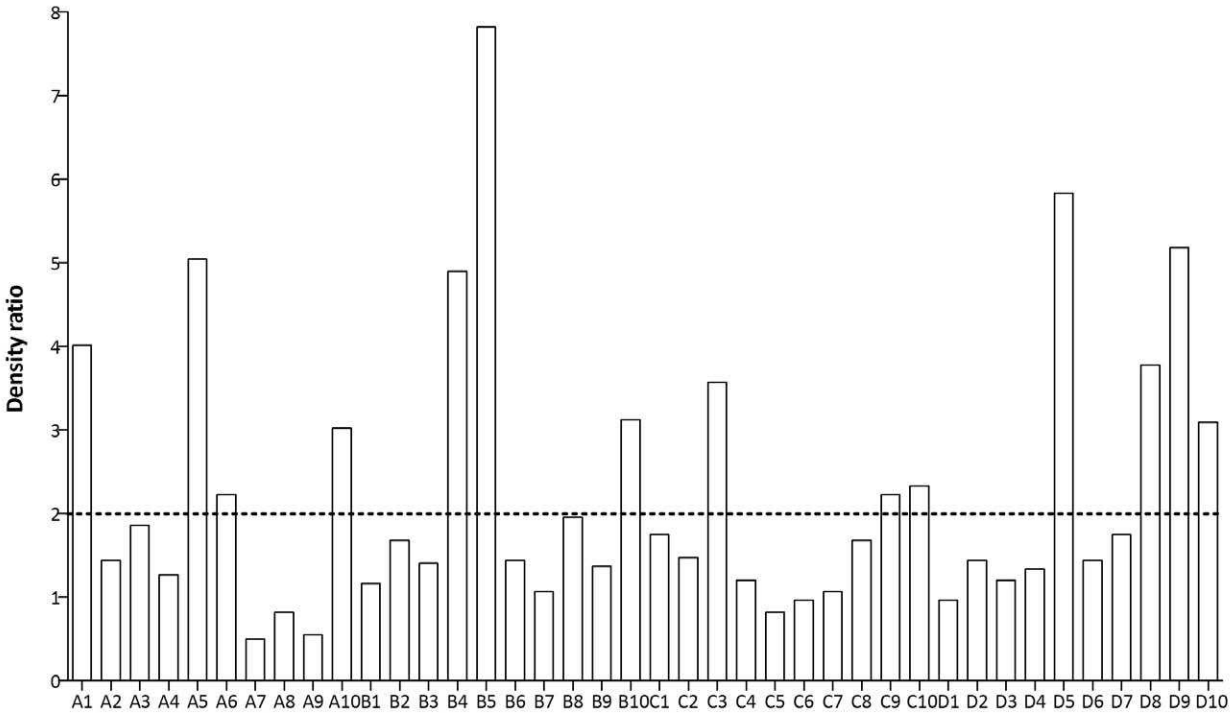


Figure 7. Antigenic profile of TcS peptides. The top panel shows a representative result of immunoblot employing a SPOT synthesis membrane and pools of sera from *T. cruzi*-infected mice (A) and from control uninfected mice (B). The reaction was revealed with secondary anti-total IgG antibody. The bottom panel shows the relative intensity of the signal of each spot estimated based on a comparison of the reactivity in immunoblots with sera from *T. cruzi*-infected mice to the background levels, determined by reactivity with sera from uninfected mice. A signal was scored as reactive when relative intensity (RI)$\geq$2. The peptides analyzed for each TcS group are as follows: TcS group I, D5–D10; TcS group III, C9-D4; TcS group IV, B5-C1, C3; TcS group V, A1, C2, C3, C7; TcS group VI, C2–C8; TcS group VII, A9-B4; TcS group VIII, A2–A8.
doi:10.1371/journal.pone.0025914.g007

(Figure 2) and therefore may also display carbohydrate binding properties.

After mapping all TcS groups on the *T. cruzi* chromosomes [20], we found no association between a group and a specific chromosomal location (Figure 3). Interestingly, we found a distinctive pattern of gene distribution along the chromosomes for members of the TcS groups II and V, with the former clearly enriched at the end of the chromosomes, whereas the latter is concentrated in the middle of the chromosomes (Figure 4). *Trypanosoma brucei* and *Plasmodium falciparum* have a sophisticated strategy for immune evasion, known as antigenic variation, which allows the parasites to adapt to the host environment through exposing and changing specific variable antigenic surface proteins [53–58]. In these parasites, the genes that encode surface proteins that are involved in antigenic variation are preferentially located at subtelomeric regions because these are favorable genomic environments that facilitate gene switching, expression, expansion and generations of new variants [53,55]. Because we found an enrichment of TcS pseudogenes within subtelomeres (Figure 4C), we speculate that these *T. cruzi* regions have also been subjected to intense rearrangement. *T. cruzi* does not undergo antigenic variation but instead co-expresses several variable surface proteins, among which is TcS [59]. Nevertheless, the subtelomeric location of TcSgroupII may facilitate the generation of new variants. In fact, *in silico* simulations suggested that both mutation and gene conversion may contribute to the generation of diversity in the TcS family [60,61]. Gene conversion may be frequent in subtelomeric regions, and therefore could promote a faster diversification of TcSgroupII. This scenario may be particularly important for this group because several of its members have been implicated in host cell attachment/invasion, and *T. cruzi* has the ability to infect a broad range of host cells. Therefore, it is possible that the large repertoire of peptides derived from TcSgroupII may contribute to this phenomenon.

Co-expression of several members of the TcS family has been described in the mammalian stages of the parasite [59]. Here, we show that the levels of expression are not homogeneous between and within the TcS groups (Figure 5). It is well known that the 3′UTRs are implicated in the control of the gene expression of several *T. cruzi* genes that are regulated during the life cycle. Although we have not mapped the 3′UTRs of the genes selected for expression analysis, for a few genes, it was possible to find a correlation between the expression profile and the sequence similarity in their 3′ flanking regions. For example, SAPA and TCNA genes, which are both active trans-sialidases expressed in trypomastigotes, have almost identical 3′ flanking regions (Figure 6). On the other hand, the 3′ flanking sequences of the genes TcS8 and TcS25 are quite similar (75% identity) despite the fact that their pattern of expression is very distinct (Figures 5 and 6). In this case, it is possible that cis-acting regulatory elements present in regions other than the 3′UTR may modulate their expression. It is also unclear what the proportion of the total TcS repertoire is expressed and whether the repertoire and/or the level of expressed genes may change during the parasite infection. High-throughput RNA sequencing approaches will clarify these questions.

We have also investigated the antigenic profile of peptides derived from distinct groups of the TcS family (Figure 7). Besides the known epitopes derived from the repetitive sequences of the SAPA and TcTS13 proteins, new B-cell epitopes were identified in members from both previously described and new TcS groups. Nine of the 14 reactive peptides are found in more than one TcS member. Specifically, the highly reactive peptide C3 (Figure 7) occurs in the largest number of proteins (60 in total) including members of the two new TcS groups V and VI. Also, similar but not identical sequences of this peptide were found in more than 150 TcS members. The cross-reaction among several epitopes and the sequence variability of the TcS family might contribute to the simultaneous presence of B-cell related epitopes during an infection. In fact, it has been proposed that cross-reactivity among the *T. cruzi* epitopes could be an evasion mechanism that drives the immune system into a series of spurious and non-neutralizing antibody responses [62]. In this regard, it has been shown that subtle differences at amino acid positions in or around the active site of the TcS proteins that have trans-sialidase activity might delay the immune response and avoid inhibiting the complete enzymatic makeup of the parasite [63]. This scenario may represent an evolutionary pressure driving the diversification of TcSgroupI, which harbors the active trans-sialidases. Whether a similar mechanism is involved in the diversification of the other TcS groups remains to be addressed.

The diversity of the TcS family may be even greater than reported here since the current assembly of the CL Brener genome is fragmented [2,20], and therefore additional TcS genes may not be part of the dataset analyzed in this study. Nevertheless, based on the nearly complete repertoire of TcS sequences, we can now design probes and antibodies specific for each group, to be employed in more assertive strategies to investigate the role of this complex family during *T. cruzi* infection.

Supporting Information

Figure S1 Partial alignment of active trans-sialidase proteins. FRIP and Asp-box motifs, and critical amino acids residues involved in trans-sialidase activity are shaded in gray. The amino acid positions are relative to the first methionine. Only N-terminal region of the active trans-sialidase proteins is shown.
(DOCX)

Figure S2 Multidimensional scaling plot of the TcS proteins indicating the presence of characteristic TcS motifs. TcS proteins with the motifs are represented by red dots. (A) SXDXGXTW motif; (B) VTVXNVXLYNR motif; (C) SXDXGXTW motif allowing 1 mismatch; (D) sequences with VTVXNVXLYNR motif found in the alignment block of the 505 TcS derived from the eight clusters identified in this study; (E) FRIP (XRXP) motif. X represents any amino acid.
(DOCX)

Figure S3 Prototype of each TcS protein. The peptide signal is represented in gray, FRIP in green, Asp-box in blue, VTVXNVXLYNR in red, repeats in black and GPI anchor addition site in orange.
(TIF)

Figure S4 Divergent CRP – complement regulatory proteins. Protein sequences involved in the regulation of complement system identified by Beucher and Norris (2008). Sequences were mapped on the MDS showed in Figure 1. HSG sequences (high similarity group) and LSG sequences (low-similarity group) are indicated by red and black squares, respectively.
(DOCX)

Table S1 Primers used in the Real-time RT-PCR reactions.
(DOC)

Table S2 TcS peptides analyzed by immunoblotting.
(DOCX)

Table S3 List of members of each TcS group.
(XLS)

Table S4 List of TcS repeats.
(XLSX)

Acknowledgments

We thank Michele Silva de Matos and Jefferson Bernardes for technical assistance.

Author Contributions

Conceived and designed the experiments: LMF SLS DCB. Performed the experiments: LMF SLS GRL TAOM. Analyzed the data: LMF SLS DCB. Contributed reagents/materials/analysis tools: TSR RTG SMRT RTF DCB. Wrote the paper: LMF SLS SMRT DCB.

References

1. World Health Organization (2002) Control of Chagas Disease. Second report of the WHO Expert Committee. WHO Technical Report Series 905.
2. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309: 409–415.
3. Previato JO, Andrade AF, Pessolani MC, Mendonça-Previato L (1985) Incorporation of sialic acid into *Trypanosoma cruzi* macromolecules. A proposal for a new metabolic route. Mol Biochem Parasitol 16: 85–96.
4. Mucci J, Risso MG, Leguizamón MS, Frasch AC, Campetella O (2006) The trans-sialidase from *Trypanosoma cruzi* triggers apoptosis by target cell sialylation. Cell Microbiol 8: 1086–1095.
5. Vercelli CA, Hidalgo AM, Hyon SH, Argibay PF (2005) *Trypanosoma cruzi* trans-sialidase inhibits human lymphocyte proliferation by nonapoptotic mechanisms: implications in pathogenesis and transplant immunology. Transplant Proc 37: 4594–4597.
6. Frasch AC (2000) Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. Parasitol Today 16: 282–286.
7. Pereira-Chioccola VL, Acosta-Serrano A, Correia de Almeida I, Ferguson MA, Souto-Padron T, et al. (2000) Mucin-like molecules form a negatively charged coat that protects *Trypanosoma cruzi* trypomastigotes from killing by human anti-alpha-galactosyl antibodies. J Cell Sci 113(Pt 7): 1299–1307.
8. Schenkman S, Eichinger D, Pereira ME, Nussenzweig V (1994) Structural and functional properties of Trypanosoma trans-sialidase. Annu Rev Microbiol 48: 499–523.
9. Souza W, Carvalho TM, Barrias ES (2010) Review on *Trypanosoma cruzi*: Host Cell Interaction. Int J Cell Biol 2010: 1–18.
10. Tonelli RR, Giordano RJ, Barbu EM, Torrecilhas AC, Kobayashi GS, et al. (2010) Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo. PLoS Negl Trop Dis 4: e864.
11. Tzelepis F, de Alencar BC, Penido ML, Claser C, Machado AV, et al. (2008) Infection with *Trypanosoma cruzi* restricts the repertoire of parasite-specific CD8+ T cells leading to immunodominance. J Immunol 180: 1737–1748.
12. Rubin-de-Celis SS, Uemura H, Yoshida N, Schenkman S (2006) Expression of trypomastigote trans-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole. Cell Microbiol 8: 1888–1898.
13. Cross GA, Takle GB (1993) The surface trans-sialidase family of *Trypanosoma cruzi*. Annu Rev Microbiol 47: 385–411.
14. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
15. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.
16. Felsenstein J (1989) Phylogeny Inference Package (Version 3.2). Cladistics 5: 164–166.
17. Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Free program distributed by the authors over the internet from http://evolution.genetics.washington.edu/phylip.html: Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
18. Hartigan JA, Wong MA (1979) Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society Series C (Applied Statistics) 28: 100–108.
19. R Development Core Team (2011) R: A language and environment for statistical computing, reference index version 2.13.0. (R Foundation for Statistical Computing, Vienna).
20. Weatherly DB, Boehlke C, Tarleton RL (2009) Chromosome level assembly of the hybrid *Trypanosoma cruzi genome*. BMC Genomics 10: 255.
21. Yan T, Yoo D, Berardini TZ, Mueller LA, Weems DC, et al. (2005) PatMatch: a program for finding patterns in peptide and nucleotide sequences. Nucleic Acids Res 33: W262–266.
22. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783–795.
23. Fankhauser N, Mäser P (2005) Identification of GPI anchor attachment signals by a Kohonen self-organizing map. Bioinformatics 21: 1846–1852.
24. Bartholomeu DC, Silva RA, Galvão LM, el-Sayed NM, Donelson JE, et al. (2002) *Trypanosoma cruzi*: RNA structure and post-transcriptional control of tubulin gene expression. Exp Parasitol 102: 123–133.
25. Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2: 2.
26. Frank R (1992) Spot-synthesis: an easy technique for the position- ally addressable, parallel chemical synthesis on a membrane support. Tetrahedron 48: 9217–9232.
27. Colli W (1993) Trans-sialidase: a unique enzyme activity discovered in the protozoan *Trypanosoma cruzi*. FASEB J 7: 1257–1264.
28. Roggentin P, Rothe B, Kaper JB, Galen J, Lawrisuk L, et al. (1989) Conserved sequences in bacterial and viral sialidases. Glycoconj J 6: 349–353.
29. Beucher M, Norris KA (2008) Sequence diversity of the *Trypanosoma cruzi* complement regulatory protein family. Infect Immun 76: 750–758.
30. Todeschini AR, Mendonça-Previato L, Previato JO, Varki A, van Halbeek H (2000) Trans-sialidase from *Trypanosoma cruzi* catalyzes sialoside hydrolysis with retention of configuration. Glycobiology 10: 213–221.
31. Teixeira SM, Russell DG, Kirchhoff LV, Donelson JE (1994) A differentially expressed gene family encoding "amastin," a surface protein of *Trypanosoma cruzi* amastigotes. J Biol Chem 269: 20509–20516.
32. Campos PC, Bartholomeu DC, DaRocha WD, Cerqueira GC, Teixeira SM (2008) Sequences involved in mRNA processing in *Trypanosoma cruzi*. Int J Parasitol 38: 1383–1389.
33. Pollevick GD, Affranchino JL, Frasch AC, Sánchez DO (1991) The complete sequence of a shed acute-phase antigen of *Trypanosoma cruzi*. Mol Biochem Parasitol 47: 247–250.
34. Burns JM, Shreffler WG, Rosman DE, Sleath PR, March CJ, et al. (1992) Identification and synthesis of a major conserved antigenic epitope of *Trypanosoma cruzi*. Proc Natl Acad Sci U S A 89: 1239–1243.
35. Giordano R, Chammas R, Veiga SS, Colli W, Alves MJ (1994) *Trypanosoma cruzi* binds to laminin in a carbohydrate-independent way. Braz J Med Biol Res 27: 2315–2318.
36. Ouaissi A, Cornette J, Taibi A, Velge P, Capron A (1988) Major surface immunogens of *Trypanosoma cruzi* trypomastigotes. Mem Inst Oswaldo Cruz 83 Suppl 1: 502.
37. Velge P, Ouaissi MA, Cornette J, Afchain D, Capron A (1988) Identification and isolation of *Trypanosoma cruzi* trypomastigote collagen-binding proteins: possible role in cell-parasite interaction. Parasitology 97(Pt 2): 255–268.
38. Santana JM, Grellier P, Schrével J, Teixeira AR (1997) A *Trypanosoma cruzi* secreted 80 kDa proteinase with specificity for human collagen types I and IV. Biochem J 325(Pt 1): 129–137.
39. Magdesian MH, Giordano R, Ulrich H, Juliano MA, Juliano L, et al. (2001) Infection by *Trypanosoma cruzi*. Identification of a parasite ligand and its host cell receptor. J Biol Chem 276: 19382–19389.
40. Norris KA, Bradt B, Cooper NR, So M (1991) Characterization of a *Trypanosoma cruzi* C3 binding protein with functional and genetic similarities to the human complement regulatory protein, decay-accelerating factor. J Immunol 147: 2240–2247.
41. Van Voorhis WC, Eisen H (1989) Fl-160. A surface antigen of *Trypanosoma cruzi* that mimics mammalian nervous tissue. J Exp Med 169: 641–652.
42. Engstler M, Reuter G, Schauer R (1992) Purification and characterization of a novel sialidase found in procyclic culture forms of *Trypanosoma brucei*. Mol Biochem Parasitol 54: 21–30.
43. Grisard EC, Stoco PH, Wagner G, Sincero TC, Rotava G, et al. (2010) Transcriptomic analyses of the avirulent protozoan parasite *Trypanosoma rangeli*. Mol Biochem Parasitol 174: 18–25.
44. Buschiazzo A, Campetella O, Frasch AC (1997) *Trypanosoma rangeli* sialidase: cloning, expression and similarity to *T. cruzi* trans-sialidase. Glycobiology 7: 1167–1173.
45. Briones MR, Egima CM, Eichinger D, Schenkman S (1995) Trans-sialidase genes expressed in mammalian forms of *Trypanosoma cruzi* evolved from ancestor genes expressed in insect forms of the parasite. J Mol Evol 41: 120–131.
46. Magdesian MH, Tonelli RR, Fessel MR, Silveira MS, Schumacher RI, et al. (2007) A conserved domain of the gp85/trans-sialidase family activates host cell extracellular signal-regulated kinase and facilitates *Trypanosoma cruzi* infection. Exp Cell Res 313: 210–218.
47. Tonelli RR, Torrecilhas AC, Jacysyn JF, Juliano MA, Colli W, et al. (2011) In vivo infection by *Trypanosoma cruzi*: the conserved FLY domain of the gp85/trans-sialidase family potentiates host infection. Parasitology 138: 481–492.
48. Gaskell A, Crennell S, Taylor G (1995) The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure 3: 1197–1205.
49. Cremona ML, Campetella O, Sánchez DO, Frasch AC (1999) Enzymically inactive members of the trans-sialidase family from *Trypanosoma cruzi* display beta-galactose binding activity. Glycobiology 9: 581–587.
50. Todeschini AR, Dias WB, Girard MF, Wieruszeski JM, Mendonça-Previato L, et al. (2004) Enzymatically inactive trans-sialidase from *Trypanosoma cruzi* binds sialyl and beta-galactopyranosyl residues in a sequential ordered mechanism. J Biol Chem 279: 5323–5328.

51. Copley RR, Russell RB, Ponting CP (2001) Sialidase-like Asp-boxes: sequence-similar structures within different protein folds. Protein Sci 10: 285–292.
52. Oppezzo P, Obal G, Baraibar MA, Pritsch O, Alzari PM, et al. (2011) Crystal structure of an enzymatically inactive trans-sialidase-like lectin from *Trypanosoma cruzi*: The carbohydrate binding mechanism involves residual sialidase activity. Biochim Biophys Acta 1814: 1154–1161.
53. Scherf A, Lopez-Rubio JJ, Riviere L (2008) Antigenic variation in *Plasmodium falciparum*. Annu Rev Microbiol 62: 445–470.
54. Kim D, Chiurillo MA, El-Sayed N, Jones K, Santos MR, et al. (2005) Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. Gene 346: 153–161.
55. Horn D, Barry JD (2005) The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. Chromosome Res 13: 525–533.
56. Chiurillo MA, Peralta A, Ramírez JL (2002) Comparative study of *Trypanosoma rangeli* and *Trypanosoma cruzi* telomeres. Mol Biochem Parasitol 120: 305–308.
57. Cano MI (2001) Telomere biology of trypanosomatids: more questions than answers. Trends Parasitol 17: 425–429.
58. Scherf A, Figueiredo LM, Freitas-Junior LH (2001) Plasmodium telomeres: a pathogen's perspective. Curr Opin Microbiol 4: 409–414.
59. Atwood JA, Weatherly DB, Minning TA, Bundy B, Cavola C, et al. (2005) The *Trypanosoma cruzi* proteome. Science 309: 473–476.
60. Azuaje FJ, Ramirez JL, Da Silveira JF (2007) *In silico*, biologically-inspired modelling of genomic variation generation in surface proteins of *Trypanosoma cruzi*. Kinetoplastid Biology and Disease 6: 6–17.
61. Azuaje F, Ramirez JL, Da Silveira JF (2007) An exploration of the genetic robustness landscape of surface protein families in the human protozoan parasite *Trypanosoma cruzi*. IEEE Transactions on Nanobioscience 6: 223–228.
62. Pitcovsky TA, Buscaglia CA, Mucci J, Campetella O (2002) A functional network of intramolecular cross-reacting epitopes delays the elicitation of neutralizing antibodies to *Trypanosoma cruzi* trans-sialidase. J Infect Dis 186: 397–404.
63. Ratier L, Urrutia M, Paris G, Zarebski L, Frasch AC, et al. (2008) Relevance of the diversity among members of the *Trypanosoma cruzi* trans-sialidase family analyzed with camelids single-domain antibodies. PLoS One 3: e3524.

ELSEVIER

Journal of Immunological Methods 267 (2002) 13–26

JIM
Journal of
Immunological Methods

www.elsevier.com/locate/jim

The SPOT-synthesis technique
Synthetic peptide arrays on membrane supports—principles and applications

Ronald Frank *

AG Molecular Recognition, German Research Centre for Biotechnology (GBF), Mascheroder Weg 1, D-38124 Brunswick, Germany

Abstract

Presented first in 1990 at the 21st European Peptide Symposium in Barcelona, Spain [Frank, R., Güler, S., Krause, S., Lindenmaier, W., 1991. Facile and rapid 'spot synthesis' of large numbers of peptides on membrane sheets. In: Giralt, E., Andreu, D. (Eds.) Peptides 1990, Proc. 21st Eur. Peptide Symp. ESCOM, Leiden, p. 151.], the SPOT-synthesis method opened up countless opportunities to synthesise and subsequently screen large numbers of synthetic peptides as well as other organic compounds arrayed on a planar cellulose support [Tetrahedron 48 (1992) 9217]. Already in 1991, a commercial kit for manual SPOT-synthesis became available through Cambridge Research Biochemicals (CRB, UK), and in 1993, a semi-automated SPOT-synthesiser, the ASP222, was launched by ABIMED Analysen-Technik, Germany. Both made the technique available to many research laboratories, even those not experienced in or equipped for chemistry. Although SPOT-synthesis is not as impressively miniaturised as, e.g. the Affymax photolithographic technique [Science 251 (1991) 767], it fulfils similar demands with the advantage of a reliable and easy experimental procedure, inexpensive equipment needs and a highly flexible array and library formatting. The method permits rapid and highly parallel synthesis of huge numbers of peptides and peptide mixtures (pools) including a large variety of unnatural building blocks, as well as a growing range of other organic compounds. Further advantages are related to the easy adaptability to a wide range of assay and screening methods such as binding, enzymatic and cellular assays, which allow in situ screening of chemical libraries due to the special properties of the membrane supports. Therefore, peptide arrays prepared by the SPOT-technique became quite popular tools for studying numerous aspects of molecular recognition, particularly in the field of molecular immunology.

Keywords: Combinatorial synthesis; Epitope analysis; Functional genomics; High throughput screening; Microarrays; Molecular recognition; Peptide arrays; SPOT-synthesis

Abbreviations: AA, amino acid; BCIP, 5-bromo-4-chloro-indolylphosphate; Boc, *tert*-butyloxycarbonyl; BPB, bromophenol blue; CDR, complementarity determining region; Fmoc, 9-fluorenylmethoxycarbonyl; DKP, diketopiperazine; ELISA, enzyme-linked immunosorbent assay; HMBA, hydroxymethylbenzoic acid; HPLC, high-performance liquid chromatography; HTS, high throughput screening; i.d., inner diameter; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-tetrazolium bromide; NMP, *N*-methylpyrrolidinone; Pfp, pentafluorophenyl; PNA, peptide nucleic acid; PTFE, polytetrafluoroethylene; TFA, trifluoroacetic acid.

* Tel.: +49-531-6181-720; fax: +49-531-6181-795.
E-mail address: frank@gbf.de (R. Frank).

0022-1759/02/$ - see front matter
PII: S0022-1759(02)00137-0

1. Introduction

The currently very successful paradigm in scientific research applying a systematic empirical search rather than an iterative rational design to solve complex scientific questions relies heavily on technologies that permit for a rapid and comprehensive screening of diverse types of molecular probes. Combinatorial chemical and biological syntheses were the pioneering technologies that paved the way, allowing fundamentally new experimental approaches in molecular biology, immunology and drug discovery research (for a discussion on this topic, see: Gallop et al., 1994; Lebl, 1999). Miniaturisation and automation then became central topics by steadily pushing the number of probes and test samples that can be screened in ever shorter times, markedly reducing the assay dimensions and costs, as well as opening access to the analysis of ever smaller sample sizes ranging from very limited biopsy material to even single cells.

In 1988, Ed Southern disclosed his concept of synthesising partial or complete sets of oligonucleotide probes in a microscaled chequered arrangement onto a planar glass surface to be used as a tool to analyse complex nucleic acid samples of genome-wide origins by multiplexed hybridisation (Southern, 1988). This marked another technological breakthrough: microarray technology. The impact of this technology is now being compared to that of the microelectronics revolution. Other methods rapidly complemented Southern's approach such as the photolithographic synthesis of peptides and later oligonucleotides on glass (Pirrung et al., 1990) and the SPOT-synthesis of peptides on membrane supports (Frank and Güler, 1990). Meanwhile, not only chemical in situ synthesis is utilised to generate arrays of sensor probes, but sophisticated chemical printing instrumentation has been developed to dispense minute volumes of solutions of any type of compound for array at densities up to several thousands per cm^2 (Phimister, 1999). Microarray technology is a fast developing field because it enables us to utilise and exploit the enormous amount of information generated by genome and proteome research. With respect to the topic of this special issue of the *Journal of Immunological Methods*, it should be remembered that most of the current high throughput screening methods were initially invented for systematic studies on molecular recognition events in the immune system based on synthetic peptides. This is also true for SPOT-synthesis.

2. Spot-synthesis

2.1. The technique

The basis concept came from observations made previously with combinatorial oligonucleotide (Frank et al., 1983) and peptide synthesis (Frank and Döring, 1988) on separate, labelled membrane (cellulose) disks, indicating that chemical reactions can proceed to completion, when only enough reagent solution is used as can be taken up by the support material itself. This observation suggested that, e.g. individual amino acid coupling reactions in a multiple parallel synthesis scheme could be carried out simultaneously on distinct areas of a continuous membrane sheet (paper). The principle of the method is most easily illustrated by the cartoon of Fig. 1A, which has become our kind of logo for SPOT-synthesis. When dispensing a small droplet of liquid onto the planar surface of a porous membrane, the droplet is absorbed and forms a circular spot. Using a solvent of low volatility containing appropriate reagents, such a spot forms an open reactor for chemical conversions involving the reactive functions anchored to the membrane support, e.g. conventional solid phase synthesis (Merrifield, 1963). The same holds for a droplet deposited onto a nonporous surface where the chemistry then proceeds at the planar interface between liquid and solid phase. A great number of separate spots or lines can be arranged as arrays on a larger surface area and each of these is individually addressable by manual or automated delivery of the corresponding reagent solutions. The volume dispensed and the absorptive capacity of the membrane or the surface tension properties of the support/solvent combination determine the spot size. According to the area specific functionality of the support, the spot size correlates with the particular scale of the synthesis. The spot size also controls the minimal distance between spot positions and the maximum density of the array. Synthetic steps common to all spot reactors are carried out by washing the whole membrane or surface with the respective reagents and solvents.

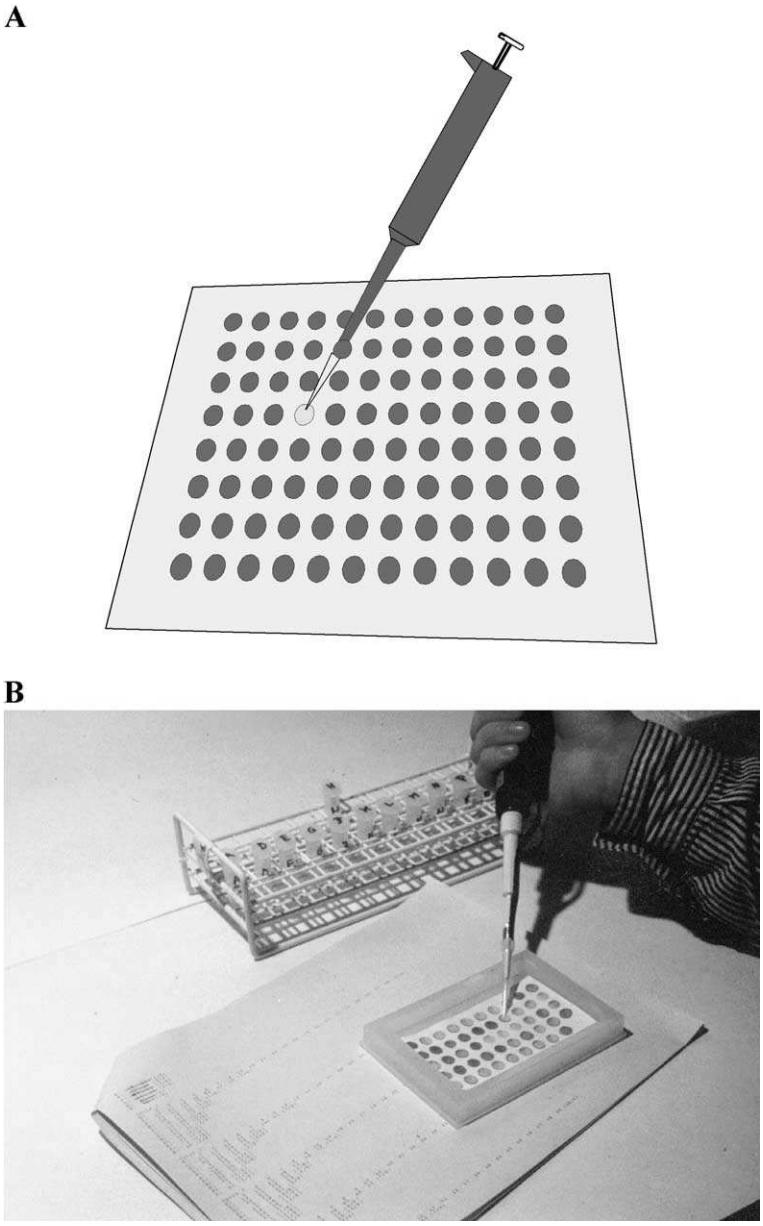


Fig. 1. (A) Cartoon illustrating the principle of SPOT-synthesis. (B) Experimental set-up for manual SPOT-synthesis showing the reaction tray with a membrane displaying blue-stained amino-spots and yellow-stained acylated spots, the rack of small tubes containing the activated AA derivatives, the computer-generated listing of the pipetting operations and the manually operated pipette.

2.2. The supports

The membrane support initially chosen is of specially selected, pure cellulose chromatography paper as produced from cotton linters. This material is well compatible both with mild Fmoc-based peptide synthesis chemistry (Frank and Döring, 1988) and the subsequent physiological media of the biological assay. Membranes of this type are still the major supports used for the assembly and biological screen-

ing of peptide arrays. With an appropriate linker/spacer chemistry, cellulose membranes are applied for the screening of immobilised peptides as well as the preparation of soluble peptides in the 1 to 100 nmol scale (Frank and Overwin, 1996). For other chemistries not compatible with the polysaccharide support, alternative materials may be selected: acrylate-grafted polypropylene fleece (Wenschuh et al., 2001) or membranes (AIMS, see below); acrylate-coated PTFE membranes (Fitzpatrick et al., 1994); polystyrene-grafted PTFE membranes (Thiele et al., in press; AIMS, see below); polystyrene-grafted polyethylene films (Berg et al., 1989). Laboratories experienced in solid phase chemistry may adapt their own special membrane supports. However, many types of carefully designed, derivatised and optimised membrane materials suitable for SPOT-synthesis are now commercially available from, e.g. AIMS Scientific Products in Braunschweig, Germany, http://www.aims-scientific-products.de). Area-specific functionalities are from 0.1 to 1 μmol/cm^2, which is comparable with conventional bead-type supports. Suitable non-porous planar surfaces for microarray fabrication by SPOT-synthesis are presently under investigation (Adler et al., 2001).

2.3. The chemistry

Arrays of spots reactors providing suitable anchor functions for peptide assembly are easily generated by first uniformly converting the surface to an "amino"-support employing a suitable spacer chemistry, e.g. the esterification of a protected amino acid such as Fmoc-βAla-OH to accessible hydroxyl functions on a cellulose membrane (Frank, 1992), or the hydroxylated surface of a whole plastic sheet followed by cleavage of the protecting group. Stimulated by SPOT-synthesis, amino-membranes with a more sophisticated and stable anchor are now commercially available (see above). In the second derivatisation step, the array of spot reactors is generated by spot-wise coupling of another amino acid derivative or a suitable linker compound, and all residual amino functions in between spots are blocked by acetylation. This array formation step requires very accurate pipetting if high density arrays are to be prepared. During peptide assembly, slightly larger volumes are dispensed so that the spots then formed exceed the original ones in order to avoid incomplete couplings at the edges.

Chemical and technical performance of this type of simultaneous parallel solid phase synthesis has been optimised for the assembly of peptide structures (linear, cyclic and branched molecules including D, L, coded and noncoded AA derivatives) by utilising conventional mild Fmoc/tBu chemistry (Frank and Overwin, 1996; Molina et al., 1996; Kramer et al., 1999). Activation of 0.2–0.3 M solutions of protected amino acid derivatives, generally dissolved in NMP, can be carried out by almost all known procedures but most favourably by the use of pentafluorophenyl esters or in situ formed 1-hydroxybenzotriazole esters. The latter two methods are compatible with the use of bromophenol blue (BPB) as an indicator of free amino functions (Krchnák et al., 1988). This blue colour staining very conveniently helps to visualise the array of the amino-spots on the membrane, to monitor the coupling reaction procedure by colour change to yellow (Fig. 1B) and also all other chemical steps of the coupling cycles such as the complete removal of piperidine from the Fmoc-deprotection step. The BPB stain makes SPOT-synthesis a really colourful experience! Final peptide deprotection is carried out with 50–90% of trifluoroacetic acid in dichloromethane with 3% triisobutylsilane and 2% water as scavengers. The cellulose support survives about 3 h of 50% TFA treatment or 1 h of 90% TFA treatment. Longer incubations then gradually weaken the paper strength and could impair the further handling in the bioassay. This is obviously no issue with the polypropylene or PTFE membranes. However, their hydrophobic nature can be problematic with the bioassay.

If solution phase peptides are to be prepared, then synthesis is carried out on correspondingly derivatised membranes by incorporation of a cleavable linker. Particularly with porous membranes, loadings of up to 1 μmol/cm^2 are practical for short peptides (<10 mers) and enough product can be isolated even from a small spot for solution phase assays. Preferably, a two-step deprotection scheme is applied, which leaves the linkages stable during side chain deprotection with TFA plus scavengers and then releases the peptides only after removal of all chemicals. Prior to release, the membrane is dried and cut into pieces each carrying one spot or the spots are punched out with the help of a suitable device. For the standard 96 or 384 formats, a

puncher is commercially available from AIMS Scientific Products that places the cut spots directly into the wells of microtiter plates (Fig. 2A). Peptide release from ester linkages of the hydroxymethylbenzoic acid (HMBA), glycolic acid or similar type can be achieved with 10 mM sodium hydroxide to give the peptide free acids (Baleux et al., 1986) or by ammonia vapour to give the carboxamides (Bray et al., 1991). Other safety-catch type linkers such as the diketopiperazine-forming Lys-Pro linker (Bray et al., 1990) or the 2-(N^{Im}-Boc-imidazol-4-yl)-glycolic acid linker (Hoffmann and Frank, 1994) release the peptides only after

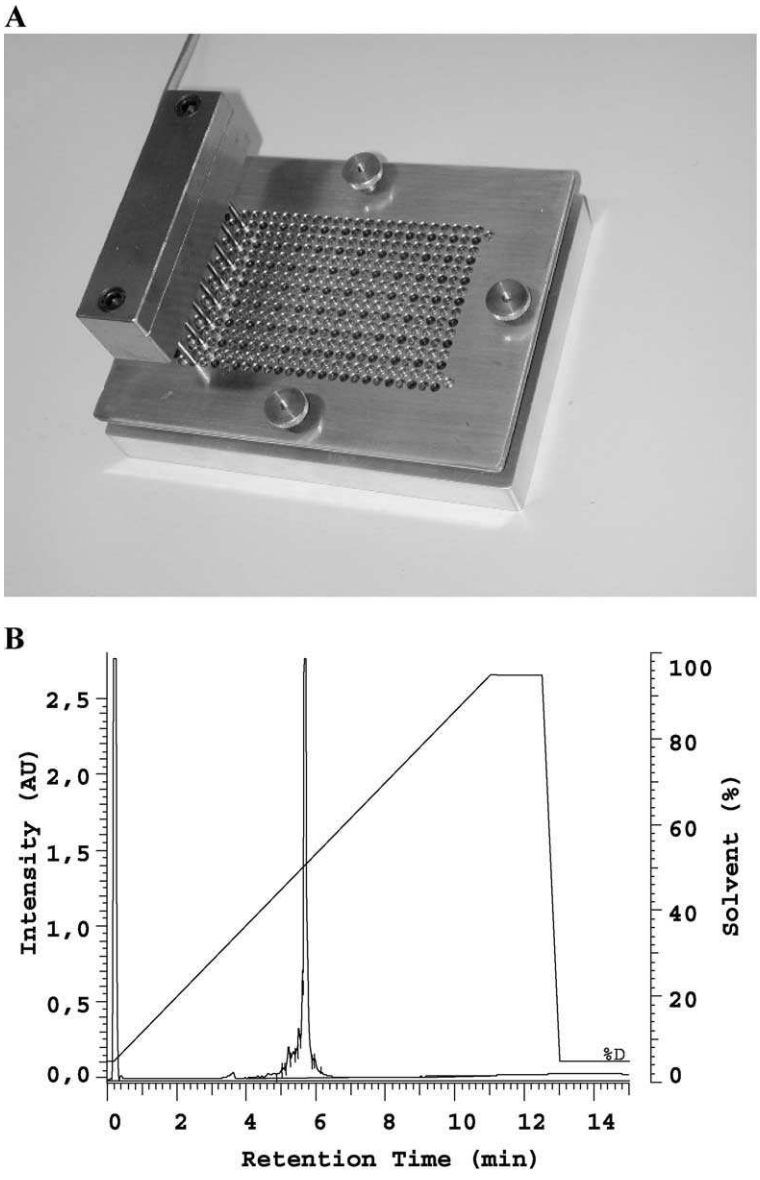


Fig. 2. (A) A puncher device to cut out eight spots at a time from a 384 membrane-bound array into the wells of a 96-well microtiter plate. (B) Reversed-phase HPLC-trace of a crude peptide Ac-AHWGMWGGWPGG-(DKP) recovered from a spot initially functionalised with the Boc-Lys(Fmoc)-Pro-linker. Chromatographic conditions: 50×2 mm i.d. 3 μm C18(2) column LUNA, Phenomenex; flow rate of 0.8 ml/min; solvent C: 0.1% TFA in water; solvent D: 0.1% TFA in acetonitrile; UV-detection at 220 nm.

a simple pH change from the acid deprotection and washing to an appropriate buffer of pH 7–8 for direct transfer into a bioassay.

Despite the simplicity of the SPOT-synthesis procedure, peptide products are of surprisingly high quality (Fig. 2B). SPOT-synthesis can be performed without compromising the chemistry. From the open reactor, the solvent is slowly evaporated; however, this maintains a constant high concentration of reagents. Having dried out after about 15 min, a second aliquot of reagent solution can be applied to the same position without any increase in the size of the spot and the risk of contaminating the neighbouring spots. In this way and guided by the BPB indicator, double to triple spottings may be applied to bring even difficult couplings to completion. Obviously, SPOT-synthesis is not a particularly new solid phase chemistry, so any problems observed with conventional solid phase synthesis will occur also here. However, meaningful results in a bioassay can be obtained with peptides of quite low purity as long as an efficient capping step is included in the synthesis cycle to avoid the formation of any other peptide contaminants except truncations of the target sequence. The majority of peptides synthesised for B- and T-cell epitope screening are 15 to 20 mers, but successful experiments have been described for even 38 mers (Töpert et al., 2001).

Amino acid coupling reactions can be performed also with mixtures of amino acid derivatives to synthesise arrays of peptide pools that cover complex peptide libraries with millions of individual sequences (Geysen et al., 1986). A reliable process for the introduction of randomised positions (X) within a peptide sequence assembled on a spot, which results in an almost equal representation of amino acid residues, is coupling with equimolar amino acid mixtures and applying these at a submolar ratio of 0.5 to 0.8 with respect to available amino functions on the spots (Kramer et al., 1993; Frank, 1994). This is to allow all activated derivatives (also the slower coupling ones) to react quantitatively during a first round of spotting. All peptide elongations are then completed by two to three successive repeats of spotting. By applying this coupling procedure, any position in a peptide sequence can easily be randomised to any degree without special considerations or increase in technical effort. The randomised X-position is just treated as a special type of amino acid derivative. Fig. 6 shows an example of a complex peptide library array probed with a monoclonal antibody. Strategies for the deconvolution of individual sequences by activity screening of randomised pools are discussed below in the context of SPOT-peptide array applications.

2.4. Instrumentation

SPOT-synthesis is particularly flexible with respect to numbers and scales that can be accomplished. The arrays are freely selectable to fit the individual needs of the experiment by variation of membrane quality, thickness, specific anchor loading and spot size, which only requires an adjustment of the spotted volume. This particular advantage was already highlighted in the original publication (Frank, 1992), which demonstrated the accessible scope (Fig. 3) ranging from (A) the standard microtiter plate array as pioneered in multiple peptide synthesis by the "multipin technique" of Geysen et al. (1984) and later adapted by several multiple synthesisers to (C) the currently favoured microscope-slide sized microarray formats. Additionally, several identical copies or different arrays can be prepared on the same membrane sheet and later cut apart for their use in separate assays.

The simple experimental set-up for manual SPOT-synthesis is shown in Fig. 1B. Proper performance of all pipetting steps is assured by taking advantage of pencil markings that can be used to index the spots on the membrane by, e.g. numbers (see also Fig. 3). An $8\times12=96$-spotted membrane, the reaction tray, amino acid derivatives, forceps, as well as a software for peptide editing and generation of a pipetting protocol are parts of the commercial SPOT-synthesis kit initially provided by Cambridge Research Biochemicals and now by Sigma-Genosys (http://www.genosys.com). A semi-automated SPOT-synthesiser has been developed at ABIMED Analysen-Technik (http://www.abimed.de) based on a Gilson pipetting workstation, the ASP222 (Fig. 4A). An almost identical robot is now also available from Intavis Bioanalytical Instruments (http://www.intavis.com). The ASP222 uses a fine, stainless steel capillary needle connected to an electronically driven syringe as the dispensing tool. The movement of the needle from one spot to the next led us to name the synthesiser our

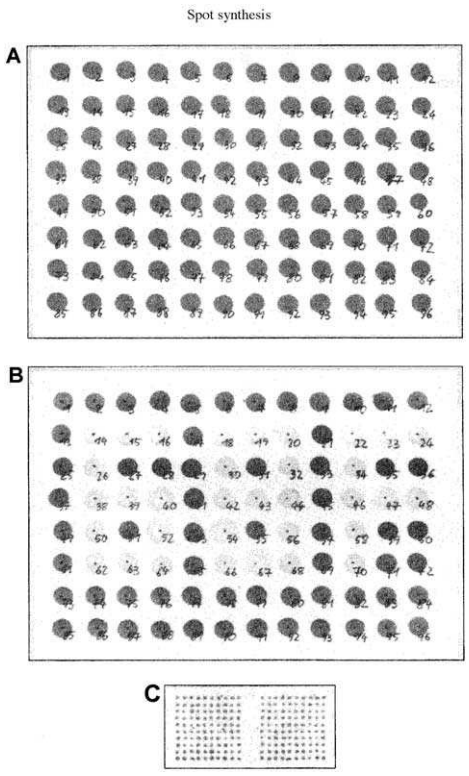


Fig. 3. Different array formats illustrating the scope of SPOT-synthesis (taken from the original publication: Frank, 1992). (A) 96 spots on an 8×12 cm paper sheet labelled with pencil numbers. (B) The same as A after spotting Fmoc-Gly-OPfp to selected spots, which causes a colour change from blue to yellow (see also Fig. 1B). (C) 200 spots manually generated with the help of a microsyringe on a piece of paper membrane having the size of a microscope slide.

"peptide sewing machine". The ASP222 can work on a total area of 30×30 cm or simultaneously on up to six standard 8×12 cm membrane sheets. Moreover, automated spotting can be exploited to reduce the size of the spots and, thus, increase considerably their number per unit area. Otherwise, this is extremely tedious to manufacture by hand. The ASP222 can generate up to 2000 spots (=1 mm at 2 mm intervals) on a standard sheet or up to 25,000 spots on the whole working area of the robot. The dispensed volume is reduced down to 30 nl in this case (Fig. 4B). Miniaturisation is limited by the thickness and homogeneity of the membrane or surface, the size of the needle opening and the capillary forces that operate when the tip of the needle touches the support surface. To date, the ASP222 instrument allows the distribution of preactivated monomer solutions onto the spot positions; all washing steps are carried out manually in separate washing trays gently agitated on a rocker platform. A fully automated process will be possible with a special synthesis block on the *MultiSynTech Syro* synthesiser (http://www.multisyntech.com) and with a novel instrument that is currently under development at Intavis.

Nonporous planar surfaces of plastic materials have been functionalised with a capacity of up to 100 nmol/cm^2 and successfully used in Spot-synthesis (Adler et al., 2001). Such materials have a much smoother surface and thus allow further miniaturisation by spotting as little as 3 nl (current limit of the ASP222 contact dispenser) at intervals as small as 1 mm. Furthermore, a rotating disc with a nonporous planar surface having the size of a CompactDisc will be the support in a novel, fully automated synthesiser (the BioDisk-Synthesiser) that utilises ink-jet heads for reagent distribution (Adler et al., 2001). This instrument will be advanced to generate microarrays of up to 10,000 peptides and other compounds in a special circular r/φ-array format (Fig. 5).

2.5. Scope of applications

As mentioned above, SPOT-synthesis is very flexible as no special equipment is needed to physically separate reactor zones and a variety of support materials covering a broad range of capacities are applicable. Thus, SPOT-synthesis can be readily adapted to many different bioassay/screening formats based on both immobilised and solution phase peptides.

Particularly in its miniaturised format, thousands of peptides or pools of peptides making up complete libraries of up to several billion components can be rapidly prepared and screened as immobilised arrays with the whole membrane or parts of it incubated in the assay. Because of its hydrophilic nature, the cellulose membranes are particularly well suited for the presentation of immobilised peptides to a biological assay system. Any reporter system can be used with the assay that results in a signal which is spatially trapped at the peptide site, e.g. detection of antibody molecules bound to peptide spots can be achieved by conventional solid phase ELISA or "Western blot" procedures (Fig. 6A). Other label-

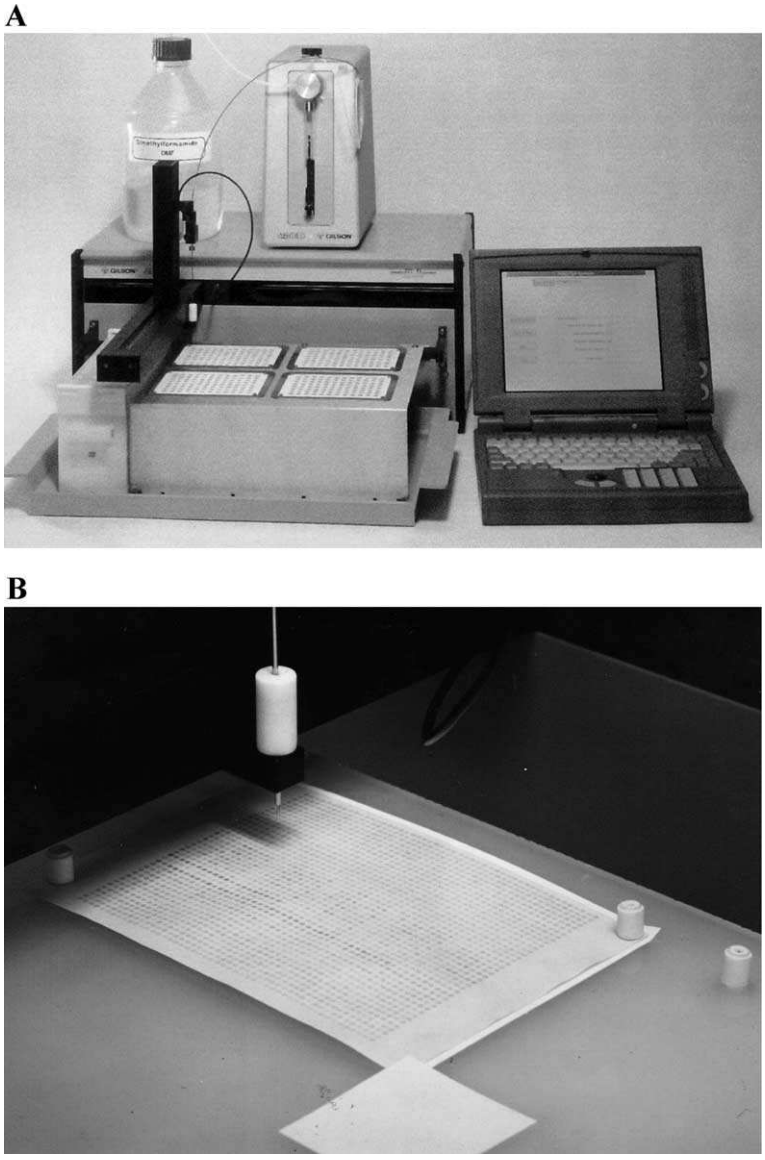


Fig. 4. (A) The ASP222 spotting robot (the peptide sewing machine). (B) The ASP 222 working on a 2*k* array having the size of a microtiter plate.

ling/detection techniques such as chemiluminescence or radioisotopes are also fully compatible. Fluorescent dyes are somewhat problematic as the synthetic arrays show some background fluorescence. Membrane bound peptide arrays are reusable many times (>50) if the peptides are not irreversibly modified by the assay procedure and if bound ligands such as proteins can be removed completely. Signal patterns obtained from peptide arrays can be documented and quantitatively evaluated utilising modern image analysis tools (Fig. 6B).

The technical advantages of the simultaneous handling of many peptide products as in an immobilised array is of course lost upon cleavage into separate

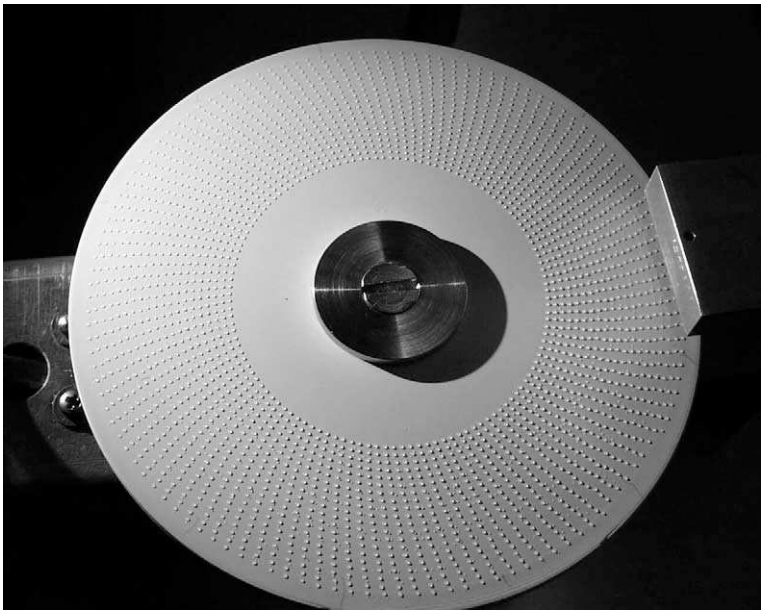


Fig. 5. 2500 droplets of reagent solutions deposited onto a turning surface with the size of a CompactDisk resulting in a circular r/φ-array format.

wells or tubes. These advantages, however, can be regained if the above-mentioned safety-catch linkers are correctly tuned for slow cleavage kinetics. Then, the entire preconditioned membrane-bound array can be overlayed with, e.g. enzymes or cells embedded in a soft agar (to restrict diffusion/movement) at physiological pH where they instantly come into contact with the locally slowly released peptides (Gonzales-Gil et al., 1997). This constitutes a novel type of very flexible high throughput screening technique comparable to the "Continuous Format-HTS (CF-HTS)" method as disclosed by Beutel et al. (1999). The latter method is based on compound arrays prepared by spotting and drying small volumes of dissolved preformed compounds from a library stock onto plastic sheets, which are then overlayed with an agar layer containing the other assay components.

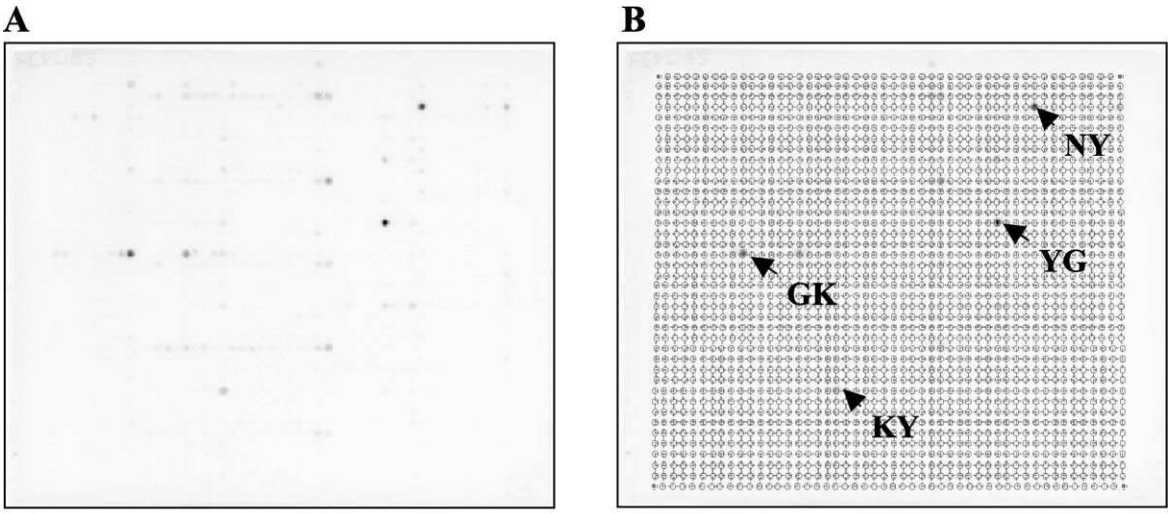
Peptide library formats that have been successfully screened by use of SPOT-peptide arrays exploit a large variety of search strategies to identify peptides that represent natural protein binding sites (epitopes), that are mimics of such sites (mimotopes, as termed by Geysen et al., 1986), or that selectively bind target molecules other than proteins. The corresponding peptide libraries were composed of very small arrays of only a few tens of peptides up to very complex libraries with more than a thousand array elements. This includes:

(A) Generic strategies such as

- *iterative deconvolution* with two defined positions (400 sub-libraries) according to Geysen et al. (1986), with three defined positions (8000 sub-libraries) according to Kramer and Schneider-Mergener (1995) and even with four defined positions of AA clusters (1296 sub-libraries) according to Kramer et al. (1995);
- *positional scanning* according to Pinilla et al. (1993);
- *dual positional scanning* according to Frank (1994);
- *orthogonal (self-deciphering) libraries* according to Déprez et al. (1994).

(B) Dedicated strategies based on available peptide/protein sequence information such as

epitope mapping, the systematic scanning of a protein sequence with overlapping 6 to 20 mers



C

Dual Positional Scan		Strongest Signal
1-2-X-X-X-X	⇒	N-Y-X-X-X-X
X-2-3-X-X-X	⇒	X-Y-G-X-X-X
X-X-3-4-X-X	⇒	X-X-G-K-X-X
X-X-X-4-5-X	⇒	X-X-X-K-Y-X
X-X-X-X-5-6	⇒	-
Natural epitope	⇒	N-Y-G-K-Y-E

Fig. 6. (A) Electronic image from a commercial, high-resolution scanner of an antibody binding experiment according to Frank (1992) with a generic dual positional scanning peptide library. Signal development was carried out with an alkaline phosphatase-conjugated secondary antibody and the BCIP/MTT reagents. (B) Array-mask, generated by the *phoretix*™ array^2 software to quantify the signals from 'A'. (C) Result of the experiment of 'A' that permits reading of the epitope sequence a priori from the overlapping dipeptide signals (Frank, 1994). Glutamic acid (E) in the natural epitope is not relevant for specific recognition as it can be replaced by any other AA (Frank et al., 1996).

having amino acid offsets from one to five residues according to Kazim and Atassi (1980) and Geysen et al. (1984);
duotope scanning, as above but with all combinations of two fragments, connected head-to-tail, for the delineation of assembled, conformational epitopes according to Reineke et al. (2000);
sizing, the definition of the boundary amino acid residues of a linear epitope by multiple length scanning of only the epitope region, according to Geysen et al. (1987);
replacement set analysis, the systematic substitution of each epitope residue by all other AA residues (usually the genetically coded ones) according to Geysen et al. (1987);
iterative transformation from L-peptide to D-peptide via D-replacement sets according to Geysen and Mason (1993) and Kramer et al. (1998);
motif scanning, a systematic editing of an epitope sequence into an all alanine (A) or all random (X) background to define consensus recognition motives according to Frank (1994) and Frank (1999).

A comprehensive and up-to-date literature survey of SPOT's applications has been recently compiled by Frank and Schneider-Mergener (2002). More than

80% of the about 300 published reports are concerned with immunological epitope analysis. However, particularly in the last years, applications in other fields are steadily growing. Below, a brief listing of different types of experiments that have been pursued with SPOT-peptide arrays is given; these are classified according to assay principles in order to highlight the broad methodological scope of the technique.

Analysis by analyte binding to immobilised peptide

Studies of immunologically relevant analytes:

- mapping and analysis of linear epitopes with polyclonal animal sera, human sera, monoclonal antibodies and single-chain Fv-fragments (soluble proteins and presented on phage);
- mapping and analysis of assembled, conformational epitopes recognised by monoclonal antibodies;
- deconvolution by generic library screening of linear epitopes and mimotopes (Geysen et al., 1986) with monoclonal antibodies;
- mapping and analysis of antibody paratopes (CDR-derived peptides) both for antigen recognition and anti-idiotype recognition;
- mapping and analysis of T-cell epitopes by MHC-binding or by T-cell stimulation with immobilised spot-bound peptides.

Analogous studies as above have been carried out for general protein–protein/peptide interaction:

- mapping and analysis of linear binding sites (receptors, chaperones, cell-skeleton proteins, signalling molecules, etc.);
- mapping and analysis of assembled, conformational binding sites;
- deconvolution by generic library screening of linear binding sites and mimics.

Further studies relate to

- inhibition of enzymes (proteases, protein kinases) by mapping and "analoguing" the substrate sites as well as by peptides deconvoluted from generic libraries;
- interactions between protein/peptide and nucleic acids;
- interactions between peptide nucleic acid (PNA) and nucleic acids;
- interactions of peptides with small ligands such as heme or metal ions;
- de novo protein design has been reported for heme and copper-binding 4-helix bundles.

Analysis by chemical or enzymatic transformation of immobilised peptide arrays

- Chemical ligation and glycation selectivity;
- mapping and analysis as well as deconvolution of protein kinase substrates;
- mapping and analysis of protease substrates.

Analysis of bound analyte

These applications are related to the use of peptide arrays as tools in the multiple affinity enrichment and isolation of the bound analyte for further analysis and application.

- Monospecific antibodies from polyclonal sera;
- monospecific Fv-fragments from phage display libraries;
- proteins from cell extracts;
- proteins from mutagenesis or cDNA-expression libraries presented on phage particles.

Analysis by the activity of cleaved, solution-phase peptide arrays

- Mapping and analysis of T-cell epitopes by in vitro T-cell stimulation;
- mapping and analysis as well as deconvolution of inhibitory peptides for proteases and protein kinases;
- mapping and analysis of ligand peptides for receptor signalling on intact cells.

Miscellaneous

SPOT-bound (cellulose-bound) peptides have been used directly in animal immunization.

3. Outlook

At present, an increasing number of organisms are becoming completely known, at least at the level of

the genome structure. This immediately opens opportunities to extend those dedicated peptide library strategies to genome-coded protein sequence information for a whole chromosome, genome or intergenome inspection. Such libraries might cover all combinatorial variants of a bioactive amino acid sequence as displayed by a complete genome or a family of genomes (e.g. the genome-specific patterns of T-cell epitope peptides; potential cross-reactive analogues of an epitope present in the human and a pathogen's genome for deciphering the cause of autoimmune responses), genome-wide species-specific differences in peptide repertoires, or even the entire peptide complement (a genome-wide peptide scan) of an organism for a global protein interaction mapping. Parallel synthesis in an array format is particularly adequate to display these types of genome-dedicated peptide diversity to a bioassay, because this diversity covers large numbers of distinct peptide sequences. This format not only suits rapid identification of novel bioactive peptide fragments and the establishment of protein/protein interaction maps based on peptide epitope recognition, but will also allow global monitoring of the responses of a part or even all peptide specific targets in a complex biological system or organism.

Array technology rapidly evolves and promises to become an important commercial factor in the future biotech industry. SPOT-synthesis contributed to the advancement of this technological development. To keep pace with the exploding amount of sequence information coming from the many genome initiatives (including the human genome), which urgently seek functional characterisation, we can expect more intelligent approaches and more potent array production processes to emerge. Due to the biological potency of peptides, peptide arrays will be important tools for functional genomics and proteomics. Finally, the transfer of such powerful research tools to future medical diagnostics is obvious. Nevertheless, to date, a process that meets the high quality and low cost production criteria for routine clinical diagnosis is unavailable.

Chemical synthesis and screening of non-peptide compound arrays by SPOT-synthesis is not covered in this manuscript as it is outside the topic of this special issue. However, respective applications are of increasing interest and relevance (Borman, 2000; Stockwell, 2000). With respect to molecular immunology, such arrays could have a strong impact on research in the field of innate immunity.

References

Adler, F., Türk, G., Frank, R., Zander, N., Wu, W., Volkmer-Engert, R., Schneider-Mergener, J., Gausepohl, H., 2001. A new array format for the automated parallel combinatorial synthesis by the SPOT-technique. In: Epton, R. (Ed.), Proc. International Symp. on 'Innovation and Perspectives in Solid Phase Synthesis 2000', York, 1999. Mayflower Worldwide, Kingswinford, p. 221.

Baleux, F., Calas, B., Mery, J., 1986. Glycolamidic ester groups as labile linkage in solid phase peptide synthesis: use with FMOC-protected amino acids. Int. J. Pept. Protein Res. 28, 22.

Berg, R.H., Almdal, K., Batsberg-Pedersen, W., Holm, A., Tam, J.P., Merrifield, R.B., 1989. Long-chain polystyrene-grafted polyethylene film matrix: a new support for solid-phase peptide synthesis. J. Am. Chem. Soc. 111, 8024.

Beutel, B.A., Schurdak, M.E., Voorbach, M.J., Burns, D.J., Joseph, M.K., 1999. Continuous format high throughput screening. USA Patent 5976813.

Borman, S., 2000. Combinatorial synthesis hits the spot. Chem. Eng. News 78, 25.

Bray, A.M., Maeji, N.J., Geysen, H.M., 1990. The simultaneous multiple production of solution phase peptides: assessment of the Geysen method of simultaneous peptide synthesis. Tetrahedron Lett. 31, 5811.

Bray, A.M., Maeji, N.J., Jhingran, A.G., Valerio, R.M., 1991. Gas phase cleavage of peptides from a solid support with ammonia vapour. Application in simultaneous multiple peptide synthesis. Tetrahedron Lett. 32, 6163.

Déprez, B., Williard, X., Bourel, L., Coste, H., Hyafil, F., Tartar, A., 1994. Self-deciphering, orthogonal combinatorial libraries of soluble organic compounds: discovery of a potent V2 vasopressin antagonist. In: Maia, H.L.S. (Ed.), Peptides 1994: Proc. 23rd Symp. of the Europ. Peptide Society, Braga, 1994. ESCOM, Leiden, p. 455.

Fitzpatrick, R., Goddard, P., Stankowski, R., Coull, J., 1994. Hydrophilic membrane supports for DNA synthesis. In: Epton, R. (Ed.), Proc. International Symposium on 'Innovation and Perspectives in Solid Phase Synthesis', Oxford, 1993. Mayflower Worldwide, Birmingham, p. 157.

Frank, R., 1992. Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. Tetrahedron 48, 9217.

Frank, R., 1994. SPOT-synthesis: an easy and flexible tool to study molecular recognition. In: Epton, R. (Ed.), Proc. International Symposium on 'Innovation and Perspectives in Solid Phase Synthesis', Oxford, 1993. Mayflower Worldwide, Birmingham, p. 509.

Frank, R., 1999. Motif-scanning: a combinatorial randomisation approach for the elucidation of key residues in linear epitopes. In: Epton, R. (Ed.), Proc. International Symposium on 'Innova-

tion and Perspectives in Solid Phase Synthesis and Combinatorial Libraries 1998'. Mayflower Scientific, Birmingham, p. 295.

Frank, R., Döring, R., 1988. Simultaneous multiple peptide synthesis under continuous flow conditions on cellulose paper discs as segmental solid supports. Tetrahedron 44, 6031.

Frank, R., Güler, S., 1990. Verfahren zur schnellen Synthese von trägergebundenen oder freien Peptiden oder Oligonucleotiden, damit hergestelltes Flachmaterial, dessen Verwendung sowie Vorrichtung zur Durchführung des Verfahrens. German patent application P 40 27 657.9.

Frank, R., Overwin, H., 1996. SPOT-synthesis: epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. In: Morris, G.E. (Ed.), Methods in Molecular Biology. Epitope Mapping Protocols, vol. 66. Humana Press, Totowa, p. 149.

Frank, R., Schneider-Mergener, J., 2002. SPOT-synthesis—scope and applications. In: Koch, J., Mahler, M. (Eds.), Peptide Arrays on Membrane Supports: Synthesis and Applications. Springer Verlag, Heidelberg, p. 1.

Frank, R., Heikens, W., Heisterberg-Moutsis, G., Blöcker, H., 1983. A new general approach for the simultaneous chemical synthesis of large numbers of oligonucleotides: segmental solid supports. Nucleic Acids Res. 11, 4365.

Frank, R., Hoffmann, S., Kieβ, M., Lahmann, H., Tegge, W., 1996. Combinatorial synthesis on membrane supports by the SPOT-technique: imaging peptide sequence and shape space. In: Jung, G. (Ed.), Peptide and Non-Peptide Libraries—A Handbook for Organic Medicinal and Biochemists. VCH, Weinheim, p. 363.

Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A., Gordon, E.M., 1994. Applications of combinatorial technologies to drug discovery: 1. Background and peptide combinatorial libraries. J. Med. Chem. 37, 1233;
Applications of combinatorial technologies to drug discovery: 2. Combinatorial organic synthesis, library screening strategies, and future directions. J. Med. Chem. 37, 1385.

Geysen, H.M., Mason, T.J., 1993. Screening chemically synthesised peptide libraries for biologically-relevant molecules. Bioorg. Med. Chem. Lett. 3, 397.

Geysen, H.M., Meloen, R.H., Barteling, S.J., 1984. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. Proc. Natl. Acad. Sci. U. S. A. 81, 3998.

Geysen, H.M., Rodda, S.J., Mason, T.J., 1986. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. Mol. Immunol. 23, 709.

Geysen, H.M., Rodda, S.J., Mason, T.J., Tribbick, G., Schoofs, P.G., 1987. Strategies for epitope analysis using peptide synthesis. J. Immunol. Methods 102, 259.

Gonzales-Gil, G., Volkmer-Engert, R., Germeroth, L., Piossek, C., Schleuning, W.-D., Schneider-Mergener, J., 1997. A novel high capacity in vivo screening system using inducible peptide libraries bound to continuous cellulose membranes. In: Epton, R. (Ed.), Peptides 1996: Proc. 24th Symp. of the Europ. Peptide Society, Edinburgh, 1996 Mayflower Scientific, Kingswinford, p. 427.

Hoffmann, S., Frank, R., 1994. A new safety-catch peptide–resin linkage for the direct release of peptides into aqueous buffers. Tetrahedron Lett. 35, 7763.

Kazim, A.L., Atassi, M.Z., 1980. A novel and comprehensive synthetic approach for the elucidation of protein antigenic structures. Biochem. J. 191, 261.

Kramer, A., Schneider-Mergener, J., 1995. Highly complex combinatorial cellulose-bound peptide libraries for the detection of antibody epitopes. In: Maia, H.L.S. (Ed.), Peptides 1994: Proc. 23rd Symp. of the Europ. Peptide Society, Braga, 1994. ESCOM, Leiden, p. 475.

Kramer, A., Volkmer-Engert, R., Malin, R., Reineke, U., Schneider-Mergener, J., 1993. Simultaneous synthesis of peptide libraries on single resin and continuous cellulose membrane supports: examples for the identification of protein, metal and DNA binding peptide mixtures. Pept. Res. 6, 314.

Kramer, A., Vakalopoulou, E., Schleuning, W.-D., Schneider-Mergener, J., 1995. A general route to fingerprint analyses of peptide–antibody interactions using a clustered amino acid peptide library: comparison with a phage display library. Mol. Immunol. 32, 459.

Kramer, A., Stigler, R.D., Knaute, T., Hoffmann, B., Schneider-Mergener, J., 1998. Stepwise transformation of a cholera toxin and a p24 (HIV-1) epitope into D-peptide analogs. Protein Eng. 11, 941.

Kramer, A., Reineke, U., Dong, L., Hoffmann, B., Hoffmuller, U., Winkler, D., Volkmer-Engert, R., Schneider-Mergener, J., 1999. Spot synthesis: observations and optimisations. J. Pept. Res. 54, 319.

Krchnák, V., Vágner, J., Safár, P., Lebl, M., 1988. Noninvasive continuous monitoring of solid-phase peptide synthesis by acid–base indicator. Collect. Czechoslov. Chem. Commun. 53, 2542.

Lebl, M., 1999. Parallel personal comments on classical papers in combinatorial chemistry. J. Comb. Chem. 1, 3.

Merrifield, R.B., 1963. Solid phase peptide synthesis: I. The synthesis of a tetrapeptide. J. Am. Chem. Soc. 85, 2149.

Molina, F., Laune, D., Gougat, C., Pau, B., Granier, C., 1996. Improved performance of spot multiple peptide synthesis. Pept. Res. 9, 151.

Phimister, B., 1999. The Chipping Forecast. Nat. Genet. 21, Supplement.

Pinilla, C., Appel, J., Houghten, R.A., 1993. Positional scanning synthetic peptide combinatorial libraries. In: Schneider, C.H., Eberle, A.N. (Eds.), Peptides 1992. Proc. 22nd Eur. Peptide Symp. ESCOM Science Publishers, Leiden, p. 65.

Pirrung, M.C., Read, J.L., Fodor, S.P.A., Stryer, L., 1990. Large scale photolithographic solid phase synthesis of polypeptides and receptor binding screening thereof. United States Patent application 492, 462.

Reineke, U., Sabat, R., Hoffmüller, U., Schmidt, M., Kurzhals, D., Wenschuh, H., Volk, H.-D., Germeroth, L., Schneider-Mergener, J., 2000. Identification of miniproteins using cellulose-bound duotope scans. In: Fields, G.B., et al. (Eds.), Peptides for the New Millenium, Proc. 16th Amer. Peptide Symp. 1999. Kluwer Academic Publishing, Dordrecht, p. 167.

Southern, E.M., 1988. Analysing polynucleotide sequences. Great Britain Patent Application GB 8810400.5.

Stockwell, R.B., 2000. Chemical genetics: ligand-based discovery of gene function. Nat. Rev. 1, 116.

Thiele, J., Atuegbu, A.E., de Francisco, M.N., Antonenko, V.V., 2000. Novel supports for solid-phase organic synthesis (SPOS): polystyrene-grafted porous Teflon membranes. In: Epton, R. (Ed.), Proc. Int. Symp. on Combinatorial Chemistry 2000. Mayflower Worldwide, Kingswinford, in press.

Töpert, F., Pires, R., Landgraf, C., Oschkinat, H., Schneider-Mergener, J., 2001. Synthesis of an array comprising 837 variants of the hYAP WW protein domain. Angew. Chem., Int. Ed. Engl. 40, 897.

Wenschuh, H., Schmidt, M., Germeroth, L., Reineke, U., Scharn, D., Heine, N., Hummel, G., Jobron, L., Schneider-Mergener, J., Matuschewski, H., Ulbricht, M., Schedler, U., Schulz, M., 2001. Spatially addressed SPOT-synthesis on novel polymeric membranes. In: Epton, R. (Ed.), Proc. International Symp. on 'Innovation and Perspectives in Solid Phase Synthesis 2000', York, 1999. Mayflower Worldwide, Kingswinford, p. 383.

B cell epitope prediction

Article · January 2009

CITATIONS
60

READS
1,256

2 authors:

Julia Ponomarenko
University of California, San Diego
72 PUBLICATIONS **3,733** CITATIONS

SEE PROFILE

Marc Van Regenmortel
Medical University of Vienna
430 PUBLICATIONS **13,511** CITATIONS

SEE PROFILE

35

B-CELL EPITOPE PREDICTION

Julia V. Ponomarenko and Marc H.V. van Regenmortel

INTRODUCTION

When a living organism encounters a pathogenic virus or microbe, the B cells of the immune system recognize the pathogen's antigens by their membrane-bound immunoglobulin receptors and, in response, produce antibodies specific to these antigens. The term *antigen* refers to any entity—a cell, a macromolecular assembly, or a molecule—that may be bound by either a B-cell receptor or an antibody molecule. The binding portion of an antigen is called a *B-cell epitope* or an antigenic determinant. If an antigen is a protein, an epitope may be either a short peptide from the protein sequence or a patch of atoms on the protein surface in the three-dimensional space. Since other types of epitopes, such as T-cell epitopes, are not discussed in this chapter, B-cell epitopes will be referred to as *epitopes*.

The property of an antigen to bind specifically complementary antibodies is known as the antigen's *antigenicity*; likewise, the ability of an antigen to induce an immune response is called its *immunogenicity*. Neither antigenicity nor immunogenicity is an intrinsic feature of the antigen. Antigenicity is defined with respect to a specific antibody, and an epitope thus acquires an identity only because an antibody is able to bind to it. The entire accessible surface of an antigen is likely to be recognized by a panel of antibodies that is large enough (Berzofsky, 1985). Although an antigen's antigenicity is defined by antibody–antigen interactions, its immunogenicity depends on characteristics of the immune system of the organism encountering the antigen. For example, rabbit albumin is immunogenic in the mouse but not in the rabbit.

The main topic of this chapter is the problem of epitope prediction in protein antigens. The story began more than 40 years ago when Anderer in Germany showed that short C-terminal peptides of the coat protein of tobacco mosaic virus could elicit antibodies that recognized the virus and neutralized its infectivity. At that time, there were no data on three-dimensional structures of viral proteins; therefore, attempts to discover peptides that could mimic protein epitopes and possess the same immunogenicity as the whole protein

Structural Bioinformatics, Second Edition Edited by Jenny Gu and Philip E. Bourne

remained entirely empirical. Subsequently, theoretical methods for epitope prediction have been developed leading to synthesis of such peptides that are important for development of immunodiagnostic tests and vaccines.

THE PROBLEM OF B-CELL EPITOPE PREDICTION

The main objective of epitope prediction is to design a molecule that can replace an antigen in the process of either antibody production or antibody detection. Such a molecule can be synthesized or, in case of a protein, its gene can be cloned into an expression vector. Designed molecules are preferable to use because they are inexpensive and noninfectious in contrast to viruses or bacteria, which may be harmful to a researcher or experimental animal. Since scientists have mainly focused on peptides as molecules mimicking antigen epitopes, the applications of epitope prediction and identification methods will be discussed here for peptides only.

A synthetic peptide may correspond to a short continuous stretch from a protein sequence and bind an antibody raised against a protein; such a peptide is called a *continuous epitope* of the protein. However, as crystallographic studies of antibody–protein complexes have shown, most epitopes are *discontinuous*, since they consist of atoms from distant residues joined on the protein surface in the three-dimensional space. A synthetic peptide can also be designed to mimic the structure of a discontinuous epitope of the protein. A synthetic peptide need not necessarily mimic either continuous or discontinuous protein epitopes since it can also mimic nonprotein epitopes—polysaccharides, DNA, glycoproteins, and other molecules—thereby expanding the range of potential pathogens that can be targeted by vaccines and diagnostics. Antibodies and synthetic peptides representing epitopes have many applications in the diagnosis of infectious and autoimmune diseases. In addition, many attempts have been made to develop peptide-based synthetic vaccines.

Diagnostic immunoassays utilize either peptides or antipeptide antibodies for detection, isolation, and characterization of molecules associated with various disease states. Synthetic peptides can be used for detection of antibodies produced as a result of infections, allergies, autoimmune diseases, or cancers. To be used as a diagnostic tool, a peptide should be antigenic, that is, able to bind a specific antibody. The prediction of such peptides is one of the goals of the epitope prediction methods. Another goal is the prediction of immunogenic peptides that can be used for production of antipeptide antibodies. Antibodies can detect proteins and various disease marker molecules, including viral proteins and bacterial lyposaccharides, present at the early stages of infections. Antipeptide antibodies used in diagnostic immunoassays can be obtained either *in vivo* by immunizing animals or *in vitro* by developing hybridoma cell lines, that is, engineered cells designed to produce high volumes of antibodies, or using combinatorial libraries. For example, a panel of recombinant antibodies can be obtained using a phage display selection of antibody fragments that bind a certain protein or peptide. Practically, any 10–15 residue-long peptide an stimulate the immune system of an animal to produce antibodies that bind the peptide. The problem is to find a peptide possessing *cross-reactive immunogenicity*, that is, a peptide that is able to elicit cross-reactive antibodies binding the specific molecule from which the peptide was derived. The prediction of such peptides is vital not only for design of diagnostics but also for vaccines.

Design of potential synthetic vaccines is another major application of epitope prediction and identification methods. There are concerns about the safety of many of today's effective vaccines as they are live attenuated or killed bacteria or viruses (examples include polio, measles, cholera, pertussis); if they are incompletely attenuated or killed, they may revert their pathogenicity or cause undesirable immune reactions. In contrast, synthetic peptides are considered as candidates for safe and inexpensive vaccines. Methods predicting immunogenic peptides, which can elicit antibodies neutralizing a pathogen, could lead to rational vaccine design. However, scientists do not have sufficient knowledge of how the immune system responds to a particular pathogen; therefore, it is difficult to predict which peptides are likely to possess cross-neutralizing immunogenicity and provide protection against the pathogen (Van Regenmortel, 2006). Synthetic vaccine candidates must still be tested experimentally to demonstrate their ability to generate neutralizing antibodies.

Although a theoretical prediction of epitopes is a highly challenging task, significant progress has been made in finding synthetic peptides representing epitopes that are useful diagnostic and research tools. Research efforts are also underway to develop synthetic peptide vaccines against HIV, human T-cell leukemia virus type 1 (HTLV-1), and *Streptococcus pyogenes* infections, malaria, and severe acute respiratory syndrome (SARS). Also, synthetic peptides are considered as therapeutic vaccines to cure cancer, Alzheimer, and autoimmune diseases. Bioinformatics input into further development of epitope prediction methods is highly desired and anticipated, and it may play a major role in helping the development of synthetic vaccines.

ANTIBODY STRUCTURE AND FUNCTION

Among distinct classes of antibodies, IgG immunoglobulin is the most common class in higher mammals. IgG contains two identical heavy chains (length of about 500 amino acids), each comprised of one variable (V_H) and three constant (C_{H1}, C_{H2}, and C_{H3}) immunoglobulin domains, and two identical light chains (length of about 250 amino acids), each comprised of one variable (V_L) and one constant (C_L) immunoglobulin domains (Figure 35.1). Each IgG molecule possesses two identical antigen binding sites (*paratopes*) situated in variable regions of the molecule and comprised of *complementarity determining regions* (*CDRs*). The CDRs comprise a total of about 50–70 amino acid residues that form six loops (three in both heavy and light chains) and vary greatly in sequence and length among antibodies that bind different antigens (Collis, Brouwer, and Martin, 2003). Hypervariability of CDRs gives the more than 10^9 different antibodies circulating in an organism the capacity to bind virtually any antigen likely to be encountered.

Antibodies demonstrate high specificity for antigens, which means that an antibody raised against an antigen may sometimes not recognize it even after a single residue substitution. At the same time, antibodies are commonly cross-reactive, that is, bind different antigens. For example, cross-reactivity with structurally similar and structurally distinct antigens has been demonstrated for the monoclonal antibodies raised against hen egg-white lysozyme (HEL) (Bentley, Boulot, and Chitarra, 1994). In the case of eight different avian lysozymes that are structurally similar antigens, the variable fragment of D11.15 antibody bound all lysozymes with high affinity, while the variable fragment of D1.3 antibody was highly specific to only two lysozymes, HEL and bobwhite quail

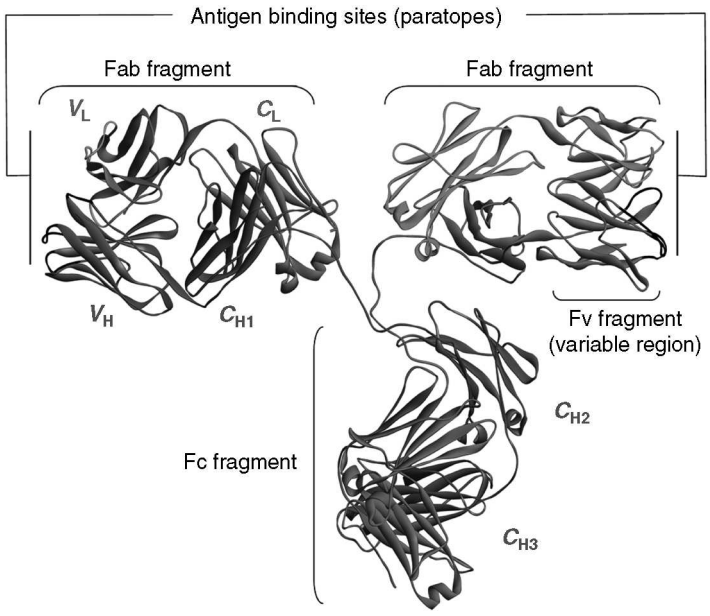


Figure 35.1. The structure of intact anticanine lymphoma monoclonal IgG2A mouse antibody [PDB : 1IGT]. The antibody heavy chains are shown in blue and red, light chains are colored in green and magenta; CDR loops are black on heavy chains and cyan on light chains. The image of the molecule was produced by J. Ponomarenko using the WebLabViewer software (Accelrys Inc.). This figure also appears in Color Figure section.

egg-white lysozyme. Antibody cross-reactivity with structurally distinct antigens has been demonstrated in complexes of D1.3 antibody with HEL and E225 antibody. Two D1.3 paratopes shared half of their residues, thus allowing other residues to make exclusive contacts with either E225 or HEL (Bentley, Boulot, and Chitarra, 1994). Usually, only about 10–25 of the 50–70 CDR residues of a certain antibody participate in the interaction with any given epitope, meaning that an antibody may harbor a large number of different paratopes.

Another additional factor mediating antibody cross-reactivity is its conformational diversity, as demonstrated, for example, by analyses of X-ray structures and binding kinetics of SPE7 antibody raised against DNP hapten (James, Roversi, and Tawfik, 2003). CDRs of SPE7 antibody adopted different conformations in its free states while binding different antigens. ''Conformational diversity, whereby one sequence adopts multiple structures and multiple functions, increases the effective size of the antibody repertoire'' (James, Roversi, and Tawfik, 2003).

Unique structural and functional properties of immunoglobulins allow an antibody to bind many different antigens. Also, an individual organism produces billions of different unique antibodies. Hence, the immune system in higher mammals can combat virtually any infection. For more detailed reviews on antibody structure, kinetics, and energetics of antibody–protein interactions, and structural basis of antibody affinity maturation see (Sundberg, 2004) and the reviews by the authors of the first X-ray structures of an intact antibody Fab fragments in free state and in complex with lysozyme (Braden et al., 1998; Braden and Poljak, 2000).

EXPERIMENTAL METHODS USED FOR B-CELL EPITOPE IDENTIFICATION

Experimental approaches for B-cell epitope identification can be divided into two categories—structural and functional. They lead to two different perceptions of what constitutes an epitope (Van Regenmortel, 1989). Structural methods include X-ray crystallography, nucleic magnetic resonance (NMR), and electron microscopy (EM) of antibody–antigen complexes. Functional methods for detecting and characterizing antibody–antigen binding utilize various techniques, such as surface plasmon resonance (Pattnaik, 2005), mass spectrometry (Hager-Braun and Tomer, 2005), and NMR spectroscopy (Johnson and Pinto, 2004), as well as immunoassays, including ELISA, ELISPOT, Western blot, and so on. The full list of assays used for epitope identification is available at the Immune Epitope Database web site (Peters et al., 2005). For a review of experimental methods used for B-cell epitope identification see (Morris, 1996).

Structural studies, particularly X-ray crystallography of antibody–antigen complexes, identify the so-called *structural epitopes* defined by the set of antigen atoms considered to be in contact with atoms of the antibody. If a structural epitope is defined as consisting of antigen residues in which any atom is separated from any antibody atom by a distance ≤ 4 Å, epitopes encompass 10–22 residues (average of 16 residues, calculated on a data set of 59 epitopes from structures of one-chain proteins in complex with two-chain antibody fragments (Ponomarenko and Bourne, 2007)). Examples of structural epitopes known in influenza A virus hemagglutinin HA1 chain are shown in Figure 35.2.

Functional epitopes are usually delineated by functional assays that measure the change in antibody–antigen binding affinity resulting from antigen modifications, such as site-directed mutagenesis (Benjamin and Perdue, 1996; Sundberg, 2004). Functional epitopes in proteins are usually smaller than structural epitopes: only three–five residues of the structural epitope contribute significantly to the antibody–antigen binding energy (Cunningham and Wells, 1993).

Functional epitopes derived from mutagenesis studies and structural epitopes obtained from structures of antibody–protein complexes are shown in Figure 35.3 for NC10 and NC41 epitopes of influenza virus N9 neuraminidase (Tulip et al., 1994). Seldom do all residues included in the structural epitope contribute to the energy of antibody–protein interaction. Thus, NC10 and NC41 structural epitopes of N9 neuraminidase include 16 and 21 residues, respectively. However, only three epitope residues contributed significantly in binding either NC10 or NC41 (Figure 35.3, residues in red), while mutations of many other residues of the structural epitopes had no effect on antibody binding (Figure 35.3, residues in cyan). At the same time, NC10 functional epitope included K432 residue that was not involved in contacts with the antibody (Figure 35.3a, residue in magenta) (Tulip et al., 1994).

Conformational changes in antibody and antigen molecules are often observed in the structures of antibody–antigen complexes in comparison to uncomplexed structures of antibody and antigen (Berger et al., 1999; Bosshard, 2001). Conformational changes take place when antibodies bind both short peptides and whole proteins. For example, Fab 17/9 undergoes major conformational rearrangements in CDR H3 upon binding a peptide from influenza virus hemagglutinin HA1(75–110) (Rini, Schulze-Gahmen, and Wilson, 1992). Another example is the structure of HEL lysozyme in complex with D44.1 Fab [PDB : 1MLC] in which both molecules show conformational changes when compared with uncomplexed structures; moreover, two complexes in the crystal unit

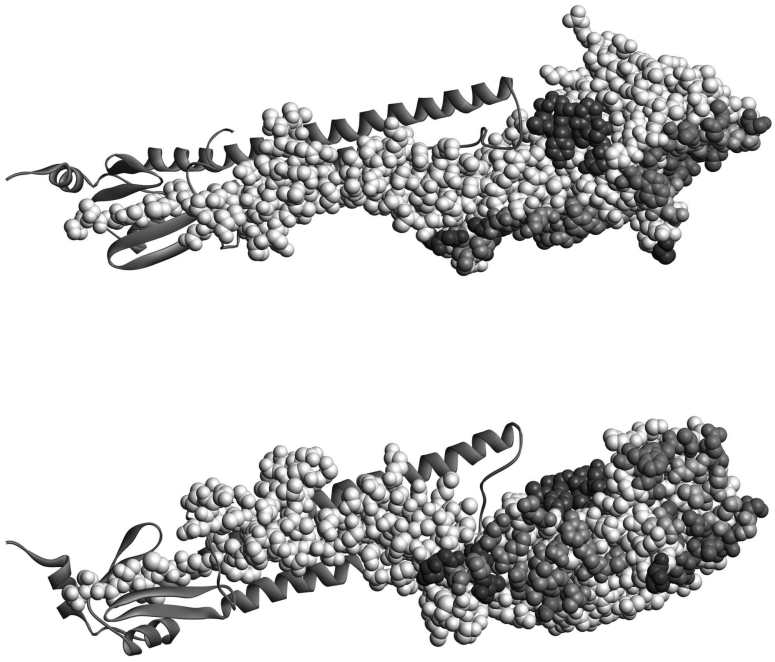


Figure 35.2. Two orthogonal views of the structure of influenza A virus (strain A/Aichi/2/68 H3N2 (X31)) hemagglutinin HA1 chain [PDB : 1EO8] with epitopes known from X-ray structures in complexes with antibodies. Chain HA1 is shown in light gray upon which are mapped residues of one linear B-cell epitope 100-YDVPDYASL-108 recognized by 17/9 Fab [PDB : 1HIM] (teal) and four structural B-cell epitopes inferred from protein structures in complexes with antibody fragments: HC45 Fab [PDB : 1QFU] (blue), BH151 Fab [PDB : 1EO8] (magenta), HC63 Fab [PDB : 1KEN] (green), and HC19 Fab [PDB : 2VIR] (red). The hemagglutinin HA2 chain is shown in cyan. Residues common to HC45 and BH151 epitopes are shown in orange; residues common to HC63 and HC19 epitopes are shown in yellow. Structural epitope consists of protein antigen residues in which any atom of the epitope residue is separated from any antibody atom by a distance $\leq$4 Å. The images were produced by J. Ponomarenko using the WebLabViewer software (Accelrys Inc.). This figure also appears in Color Figure section.

have slightly different antibody–antigen contacts and also slightly different paratope conformations (Braden et al., 1994). Investigating differences in complexes present in the same crystal asymmetric unit, Decanniere and colleagues (Decanniere et al., 2001) found that two independent complexes of the variable fragment of cAb-Lys3 antibody and HEL show a significant difference in relative orientation between antibody and antigen, and this difference reflects the flexibility of the antibody–antigen interactions rather that an artifact of crystallization.

Antigen and antibody residues that are not part of a paratope or an epitope may also undergo conformational changes during antibody–antigen binding. For example, the main chain of the residue Thr87, which is a part of neither CDRs nor the paratope of the antibody TP7, moved 3.2 Å when TP7 bound Taq DNA polymerase I [PDB : 1BGX] (Murali et al., 1998). The structure of human angiogenin in complex with 26-2F Fab [PDB : 1H0D] demonstrates an example of conformational changes of antigen residues

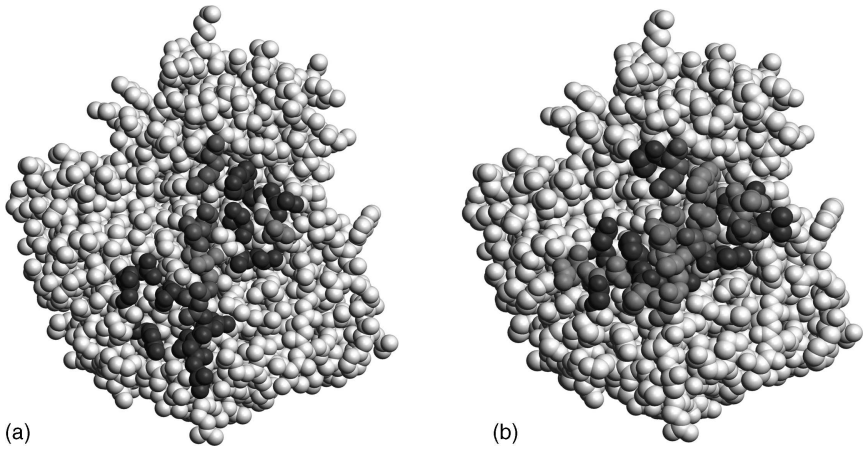


Figure 35.3. Structural and functional B-cell epitopes on influenza virus N9 neuraminidase. (a) NC10 structural epitope derived from [PDB : 1NMB]. Residue K432, which was not a part of the structural epitope but has markedly reduced antibody binding after substitution, is shown in magenta. Residues of the structural epitope N329, A369, and S370, which contributed significantly in binding of NC10 antibody, are shown in red. Residues of the structural epitope for which mutations I368R and N400K have not markedly reduced antibody binding are shown in cyan. (b) NC41 structural epitope derived from [PDB : 1NCA]. Residues of the structural epitope S367, N400, and K432, which contributed significantly in binding of NC41 antibody, are shown in red. Residues of the structural epitope, for which mutations P328K, N329D, N329K, N344I, N345G, I368R, I368Y, T401L, and W403G have not markedly reduced antibody binding, are shown in cyan. Residues A369, S370, and S372, for which mutations to smaller residues (Gly, Ala) have not markedly reduced antibody binding, are shown in green. All colored residues are part of the structural epitopes defined as antigen residues in which any atom is separated from any antibody atom by a distance $\leq$4 Å. Residues, for which mutations have markedly reduced antibody binding, are shown in red; the others are shown in cyan. Residues, for which mutagenesis data are not available, are shown in blue. Residues are colored according to data provided in (Tulip et al., 1994). The images were produced by J. Ponomarenko using the WebLabViewer software (Accelrys Inc.). This figure also appears in Color Figure section.

that are not a part of the epitope (Chavali et al., 2003). The 26-2F epitope of angiogenin is located at two loops (residues 34–41 and 85–91) and their conformations do not change when the antibody binds angiogenin. Meanwhile, conformation of the cell binding region located between these two loops (residues 59–68) changes dramatically: rmsd for 10 Cα atoms is 13.2 Å (Chavali et al., 2003). In free angiogenin [PDB : 1B1I], residues 62–68 are part of a β-hairpin formed by the 2nd and 3rd β-strands and the intervening loop 66–68; residues 59–61 lie on a loop that connects the hairpin structure with the 3rd α-helix. In the complex with antibody, the 2nd β-strand is lost, and the entire 58–67 segment forms an extended loop. Hence, the antibody binding changes the native angiogenin conformation that is critical for the functioning of the protein (Chavali et al., 2003).

In case the three-dimensional structure of the protein or its homologue is known, a functional epitope can be mapped on the protein structure for further study of antigenic properties of the epitope and antigen. Thus, using the homology modeling approach, Kolaskar and colleagues modeled the structure of the envelope glycoprotein of the

Japanese encephalitis virus and mapped discontinuous epitopes known from functional assays on the modeled structure (Kolaskar and Kulkarni-Kale, 1999). Analyzing 49 functional epitopes identified by site-directed mutagenesis, Blythe (Blythe, 2006) mapped them to known 3D structures of 22 proteins and demonstrated that buried residues prevailed in functional epitopes. It was also observed that most residues in mapped epitopes were separated too far from each other to form a single patch on the surface of the protein (Blythe, 2006). This observation highlights the difference between functional and structural epitopes.

A frequently applied method of epitope identification is the screening of peptide libraries for antibody binding (Geysen, Meloen, and Barteling 1984; Geysen et al., 1987; Folgori et al., 1994; Tribbick, 2002; Rowley, O'Connor, and Wijeyewickrema, 2004). A variety of inexpensive methods of peptide synthesis allow production of large libraries of peptides corresponding to overlapping or randomized regions of a certain protein. When antigenic peptides are selected, they can be mapped on the protein three-dimensional structure obtained from either X-ray analysis or homology modeling. For example, 17B and CG10 discontinuous epitopes of HIV-1 envelope protein gp120 and 13b5 epitope of HIV-1 capsid protein p24 have been reconstituted using 12mer peptides selected by screening the combinatorial phage display peptide libraries (Enshell-Seijffers et al., 2003). Thus reconstituted 17B and 13b5 epitopes corresponded to the structural epitopes. The structure of CG10 antibody in complex with gp120 was not available; however, using the ELISA assay, the authors demonstrated that the reconstituted CG10 epitope corresponded to the functional epitope (Enshell-Seijffers et al., 2003). Examples of epitopes mapped using libraries of overlapping peptides include epitopes of the *M. bovis* secretory protein MPB70 (Radford et al., 1990) and the Sta58 major outer membrane protein of *R. tsutsugamushi* (Lachumanan et al., 1993). The algorithms used for reconstituting discontinuous epitopes from the selected peptides are discussed later in this chapter.

Q1

Experimental methods allow the identification of either structural or functional epitopes that may not correspond exactly to the antigenic or immunogenic epitopes of a protein. The reason for this is that antibodies may recognize not only the epitopes against which they were elicited but also bind to a variety of related epitopes that possess some structural similarity with the immunogen. Antibodies raised against a native protein may thus cross-react with a linear peptide that only partially mimics the epitope present in the protein. Furthermore, it is possible that the protein preparation used to raise antibodies contained at least some denatured molecules and that the observed cross-reactivity with peptides was due to antibodies specific to the denatured protein. Many investigators believe that the majority of continuous epitopes described in literature actually correspond to unfolded regions of the denatured proteins and are not genuine epitopes present in native proteins (Laver et al., 1990; Van Regenmortel, 1996). In the case of functional epitopes, residues attributed to them on the basis that their replacement by other residues abolishes the antigenic activity may in fact not be a part of the epitope since the residue substitution may have caused conformational changes in the protein altering the binding activity. It was demonstrated, for example, that substitutions of charged and buried residues in E. coli toxin CcdB, can lead to the protein misfolding (Bajaj, Chakrabarti, and Varadarajan, 2005). Experimental evidence for individual residues to be a part of an epitope should therefore be considered with caution, taking into account the types of immunological assays used as well as possible conformational changes that may occur in the antigen.

HISTORY OF ATTEMPTS AT B-CELL EPITOPE PREDICTION

Reliable prediction of potentially immunogenic epitopes in a given protein or a whole pathogen genome may significantly reduce experimental efforts in discovering epitopes needed for the design of vaccines and immunodiagnostics. Prediction of discontinuous epitopes and design of synthetic mimics of discontinuous epitopes require knowledge of protein structures. In 1980, there were only 57 structures of proteins in the Protein Data Bank (PDB) (Berman et al., 2000). Consequently, experimental techniques and prediction methods have focused primarily on continuous epitopes. Currently, when more than 40,000 protein structures are available, including more than 250 structures of antibody–protein complexes, further development of structure-based methods for epitope prediction is anticipated. The historical landmarks in the study of B-cell epitopes are listed in Table 35.1.

Prediction of Continuous Epitopes

The rationale behind attempts at continuous epitope prediction lies in the following assumption that synthetic peptides corresponding to linear segments in proteins may elicit antipeptide antibodies that, in turn, may cross-react with the native protein antigen (so-called cross-reactive antibodies) and, in addition, neutralize the infectivity of the pathogen

TABLE 35.1. Selected Landmarks in the Study of B-cell Epitopes

Date	Landmark	Reference
1958	The first 3D structure of protein myoglobin	(Kendrew et al., 1958)
1963	First antipeptide antibody raised against the C-terminus of tobacco mosaic virus coat protein and able to neutralize the infectivity of the virus	(Anderer and Schlumberger, 1965)
1973	The first high-resolution (2.8 Å) structure of an intact immunoglobulin Fab fragment	(Poljak et al., 1973)
1980	Antipeptide cross-reactive antibodies were produced without *a priori* knowledge of discontinuous epitope in proteins and proteins 3D structures; antipeptide antibodies have become a standard tool in biology	(Walter, 1986)
1980	3D structures are known for 57 proteins	(Berman et al., 2000)
1981	The first method for linear epitope prediction based on amino acid properties	(Hopp, 1981)
1984	First attempts at epitope prediction based on 3D protein structure	(Tainer et al., 1984; Westhof et al., 1984)
1986	The first 3D structure of antibody-protein complex (Fab-lysozyme, 2.8 Å)	(Amit et al., 1986)
1990	The first 3D structure of peptide–antibody complex (2.8 Å, the antigen is a synthetic 19-amino acid peptide homologue of the C-helix of myohemerythrin (Mhr), antipeptide antibody Fab' fragment)	(Stanfield et al., 1990)
2000	The first database collecting B-cell epitopes: FIMM	(Schonbach et al., 2000)
2005	AntiJen database	(Toseland et al., 2005)
2005	Bcipepe database	(Saha, Bhasin, and Raghava, 2005)
2005	Immune Epitope Database and Analysis Resource	(Peters et al., 2005)

harboring the antigen (so-called neutralizing cross-reactive antibodies (Van Regenmortel, 2006)). In 1965, the first antipeptide neutralizing antibody was produced, but widespread use of the technique of antipeptide antibodies production occurred only 15 years later (Lerner, 1984; Walter, 1986). Antipeptide cross-reactive and neutralizing antibodies can be used as immunochemical reagents for protein identification in research, medical diagnostics, and therapies. A peptide possessing cross-neutralizing immunogenicity can be used as a vaccine. In either case, the presence of cross-reactive antigenicity or immunogenicity does not mean that the peptide exactly reproduces the structure of the epitope in the protein—the only requirement is a sufficient degree of epitope resemblance to allow cross-reactivity to be observed. More details and examples of the design of synthetic peptide-based vaccines are discussed in the section ''Applications.''

The history of attempts at linear epitope prediction began in 1981, when Hopp and Woods (Hopp and Woods, 1981) proposed a method for predicting epitopes based on calculation of residue hydrophilicity and on the assumption that hydrophilic regions in the protein are predominantly located on the surface and are therefore potentially antigenic. In 1983, Hopp and Woods published the first computer program for epitope prediction implementing the method (Hopp and Woods, 1983). The predictive power of the method was demonstrated by the authors on 12 proteins for which immunochemical data existed. In the case of the hepatitis B surface protein antigen (HBsAg), the synthesized peptide corresponding to residues 138–149 of the protein, which was predicted to be an epitope, was able to bind antibodies directed against HBaAg and to elicit an anti-HBsAg immune response in animals (Hopp and Woods, 1983). Among eight epitopes predicted in influenza hemagglutinin HA1 subunit, one epitope was later shown to be recognized by the anti-HA monoclonal antibody 17/9 (Rini, Schulze-Gahmen, and Wilson, 1992) (Figure 35.3).

Since 1981, various attempts at linear epitope prediction were published that implemented different amino acid properties, such as hydrophilicity (Hopp and Woods, 1981; Parker, Guo, and Hodges, 1986), hydrophobicity (Kyte and Doolittle, 1982; Eisenberg, Weiss, and Terwilliger, 1984), solvent accessibility (Emini et al., 1985), secondary structure (Chou and Fasman, 1974; Garnier, Osguthorpe, and Robson, 1978; Levitt, 1978; Pellequer, Westhof, and Van Regenmortel, 1993), antigenicity (Welling et al., 1985; Jameson and Wolf, 1988), flexibility (Karplus and Schulz, 1985), and many others. Subsequently, attempts were made to compare the success rate of these scale-based methods (Van Regenmortel and Daney de Marcillac, 1988; Pellequer, Westhof, and Van Regenmortel, 1991; Hopp, 1993). A comparison of 12 scales applied to 85 identified linear epitopes in 14 proteins showed that most scales gave 50–62% correct predictions (Pellequer, Westhof, and Van Regenmortel, 1991), while a β-turn scale was slightly better (70%) (Pellequer, Westhof, and Van Regenmortel, 1993). Attempts to combine amino acid properties did not significantly improve the prediction success rate (Pellequer and Westhof, 1993; Alix, 1999; Odorico and Pellequer, 2003).

Recently, several methods using machine learning approaches have been published. For example, BepiPred method is based on a combination of Parker hydrophilicity scale (Parker, Guo, and Hodges, 1986), Levitt secondary structure scale (Levitt, 1978), and hidden Markov models (HMM) (Larsen, Lund, and Nielsen, 2006). The machine learning classifier developed by Sollner (Sollner, 2006; Sollner and Mayer, 2006) uses the common single amino acid propensity scales together with parameters reflecting the probability of close neighborhood for a given stretch of amino acids. ABCpred algorithm (Saha and Raghava, 2006) that applied artificial neural network (ANN) achieved a maximum accuracy of 66%.

Methods using single amino acid propensity scales give area under the receiver operating characteristic curve (A_{ROC}) values of 0.60 (Greenbaum et al., 2007). A recent assessment of 484 amino acid scales (in the range of window sizes and cutoff values) on 50 proteins with linear epitopes from AntiJen database (Toseland et al., 2005) demonstrated that ''single-scale amino acid propensity profiles cannot be used to predict epitope location reliably'' (Blythe and Flower, 2005). ''The combination of scales and experimentation with several machine learning algorithms showed little improvement over single scale-based methods, which were considered to perform inadequately'' (Greenbaum et al., 2007). The conclusions demonstrate the need for evaluation of methods based on more reliable data sets of immunogenic epitopes and for methods applying more sophisticated propensity scales.

Prediction of Discontinuous Epitopes

Prediction of discontinuous epitopes in a protein is based on the knowledge of the protein three-dimensional structure. First attempts at epitope prediction based on 3D protein structure began in 1984 when a correlation was established between crystallographic temperature factors and several known continuous epitopes of tobacco mosaic virus protein, myoglobin, and lysozyme (Westhof et al., 1984). For myohemerythrin, highly mobile regions were determined based on X-ray crystallographic temperature factors, and it was shown that antibodies raised against peptides, which corresponded to mobile regions, reacted strongly with myohemerythrin, while antibodies raised against peptides from less mobile regions did not (Tainer et al., 1984). Thornton and colleagues (Thornton et al., 1986) proposed a method to locate potential discontinuous epitopes based on a protrusion of protein regions from the protein's globular surface. The protruded parts of proteins corresponded to the experimentally determined continuous epitopes in myoglobin, lysozyme and myohemerythrin (Thornton et al., 1986). A correlation between antigenicity, solvent accessibility, and flexibility of antigenic regions in proteins was also found (Novotny et al., 1986). However, at the time, protein structural data were mostly used for prediction of continuous epitopes rather than discontinuous epitopes.

When the first X-ray structure of an antibody–protein complex was solved in 1986 (Amit et al., 1986), a completely different approach to epitope prediction became possible. The first attempts to predict epitopes were based on structural properties of a whole discontinuous epitope, which was considered to be an area on the protein surface favorable for antibody binding. Thus, Jones and Thornton, using their method for predicting protein–protein binding sites (Jones and Thornton, 1997), were able to successfully predict epitopes on the surface of the β-subunit of human chorionic gonadotropin (βhCG) (Jones and Thornton, 2000).

At the same time, scientists began exploiting protein–protein docking algorithms for epitope prediction. Most docking algorithms are based on the ''lock-and-key'' model, which assumes the existence of predefined complementarity of three-dimensional shapes of epitope and paratope. However, conformational changes observed in the structures of antibody–antigen complexes may be the main reason why unbound docking (when antibody and antigen structures are taken from different structures) poorly performs on antibody–protein complexes (Halperin et al., 2002), and bound docking (when antibody and antigen structures are taken from the same structure of their complex) substantially outperforms unbound docking (Duhovny, Nussinov, and Wolfson, 2002). Simulated docking of paratopes based on their X-ray structures did not reveal the specific atomic interactions

responsible for antibody–antigen recognition (Carneiro and Stewart, 1994). For the detailed review of docking algorithms and their performance on antibody–protein complexes see (Halperin et al., 2002).

Discontinuous epitopes in a protein with a known 3D structure can be reconstituted from the antibody binding peptides selected from randomized peptide libraries. Interpreting such peptides may be problematic when there is no sequence similarity between peptides and the protein. Several bioinformatics tools address the problem: 3D-Epitope-Explorer (3DEX) (Schreiber et al., 2005), MIMOX (Huang et al., 2006), Epitope Mapping Tool (EMT) (Batori et al., 2006), EPIMAP (Mumey et al., 2003), MIMOP (Moreau et al., 2006), PepSurf (Mayrose et al., 2007), and Mapitope (Bublil et al., 2007). Also, to facilitate a structure-based design of peptides representing the whole surface or a particular region of a protein, the SUPERFICIAL method (Goede, Jaeger, and Preissner, 2005) and the MEPS server (Castrignano et al., 2007) have been developed. These peptides can be further selected for binding a specific antibody and mapped on the protein structure to reconstitute the epitope.

Although many attempts to predict epitopes have been made, scientists are still unable to reliably predict epitopes, both continuous and discontinuous. Despite some successful predictions achieved for particular epitopes in certain proteins, existing methods perform poorly overall even if the three-dimensional structure of a protein is known.

BIOINFORMATICS METHODS FOR B-CELL EPITOPE PREDICTION

To discuss bioinformatics methods for epitope prediction, 9 web servers available on the Internet and 14 different epitope prediction methods have been chosen. The methods have been tested on six epitopes of influenza A H3N2 hemagglutinin HA1 (strains A/Aichi/2/68 and X31) including four structural discontinuous epitopes (Ponomarenko and Bourne, 2007) and the only two known protective continuous epitopes (Bui et al., 2007). The epitopes are shown in Figure 35.2 and listed in Tables 35.2 and 35.3. Other web servers for B-cell epitope prediction, which were not evaluated in this chapter, include ePitope (http://epitope-informatics.com), Antigenic (http://liv.bmc.uu.se/cgi-bin/emboss/antigenic), and VaxiJen (http://www.jenner.ac.uk/VaxiJen/).

Sequence-Based Methods

Sequence-based methods are limited to the prediction of continuous epitopes. The majority of the sequence-based methods assume that epitopes have to be accessible for antibody binding and, hence, are based on using epitope properties related to surface exposure. In the current review, sequence-based methods have been tested on prediction of two protective epitopes known in influenza A virus hemagglutinin HA1 (Bui et al., 2007). The first continuous epitope is the 91–108 epitope (SKAFSNCYPYDVPDYASL), which is a protective epitope in rabbit; that is, an epitope able to elicit in a rabbit antibodies neutralizing infectivity of influenza viruses of the strains A/Achi/2/68(H3N2) and A/Texas/1/77(H3N2) (Muller, Shapira, and Arnon, 1982). The 100–108 epitope known from the X-ray structure in complex with 17/9 antibody, which is a part of the protective epitope, is shown in Figure 35.2 in teal. The second continuous epitope is the 127–133 epitope (WTGVTQN) protective against the influenza strain A/Achi/2/68 (H3N2) in mouse (Naruse et al., 1994).

TABLE 35.2. Prediction of Protective Continuous B-Cell Epitopes in HA1

Method	Method Parameters	Number of Predicted Epitopes in HA1	Positions: Score (Rank) (If Applied) of the Best Prediction		
			91–108 Epitope	100–108 Epitope	127–133 Epitope
Chou and Fasman's β-turn scale	Window size 7	313 (maximum score 1.39, minimum score 0.72)	95–101: 1.35 (7)	101–107: 1.17 (57)	127–133: 1.07 (128)
Chou and Fasman's β-turn scale	Window size 9	311 (maximum score 1.28, minimum score 0.86)	96–104: 1.28 (6)	100–108: 1.1 (86)	n/a
Chou and Fasman's β-turn scale	Window size 18	303 (maximum score 1.28, minimum score 0.86)	91–108: 1.14 (41)	n/a	n/a
Emini's surface accessibility scale	Window size 6	314 (minimum score 0.08, maximum score 5.53)	100–105: 1.95 (37)	100–105: 1.95 (37)	128–133: 1.06 (129)
Karplus and Schulz's scale	Window size 7	312 (minimum score 0.9, max score 1.13)	100–106: 1.01 (138)	100–106: 1.01 (138)	127–133: 1.04 (74)
Kolaskar's antigenicity scale	Window size 7	313 (minimum score 1.02, maximum score 1.17)	97–103: 1.16 (2)	102–108: 1.11 (25)	127–133: 0.97 (260)
Parker's hydrophobicity scale	Window size 7	313 (min score 4.19, maximum score 6.24)	101–107: 3.59 (49)	101–107: 3.59 (49)	127–133: 2.2 (130)
Bepipred	Score threshold 0.35	16 (maximum score 1.94, minimum score 1.28)	99–105	99–105	127–146
Bepipred	Score threshold 0.90	5 (maximum score 1.94, minimum score 1.28)	100–103	100–103	130–144
Bepipred	Score threshold 1.30	4 (maximum score 1.94, minimum score 1.28)	Not predicted	Not predicted	Not predicted
ABCPred	Window size 16	31 (max score 0.93, threshold 0.51)	93–109: 0.70 (24)	93–109: 0.70 (24)	123–139: 0.90 (4)
ABCPred	Window size 10	28 (maximum score 0.80, threshold 0.51)	95–105: 0.70 (14)	99–109: 0.63 (20)	125–135: 0.61 (23)
CEP	default	17	91–96, 101–107	101–107	121–137

TABLE 35.3. Prediction of Structural Discontinuous B-Cell Epitopes in HA1

Epitope	BH151 Epitope [PDB : 1EO8]			HC63 Epitope [PDB : 1KEN]			HC45 Epitope [PDB : 1QFU]			HC19 Epitope [PDB : 2VIT]		
Epitope Size	15 Residues			16 Residues			19 Residues			18 Residues		
Method	Sensitivity	ppv	Specificity	Sensitivity	ppv	Specificity	Sensitivity	ppv	Specificity	Sensitivity	ppv	Specificity
CEP	**0.27**	**0.22**	**0.96**	**0.56**	**0.17**	**0.86**	**0.47**	**0.24**	**0.91**	**0.72**	**0.3**	**0.9**
DiscoTope (−3.1)	0	0	0.99	0	0	0.99	0	0	0.99	0	0	0.99
DiscoTope (−7.7)	0	0	0.78	**0.56**	**0.13**	**0.8**	0.11	0.03	0.78	0.17	0.04	0.78
DiscoTope (−10.5)	0.07	0.01	0.54	**0.81**	**0.09**	**0.58**	0.21	0.03	0.55	0.5	0.06	0.56
PatchDock 1st model	0.06	0.04	0.91	0.05	0.04	0.87	0	0	0.89	0	0	0.89
PatchDock best model of 10	0	0	0.89	0	0	0.88	**0.21**	**0.17**	**0.94**	**0.22**	**0.02**	**0.36**
DOT 1st model	0	0	0.95	0	0	0.97	0	0	0.95	0	0	0.95
DOT best model of 10	0	0	0.95	0	0	0.97	**0.21**	**0.27**	**0.96**	**0.22**	**0.33**	**0.97**
ProMate	0	0	0.99	0	0	0.99	0	0	0.99	0	0	0.99
PPI-PRED (1st patch)	0	0	0.85	**0.69**	**0.23**	**0.88**	0	0	0.85	**0.33**	**0.13**	**0.87**
PPI-PRED (best patch)	**0.87**	**0.23**	**0.86**	**0.69**	**0.23**	**0.88**	**0.84**	**0.29**	**0.87**	**0.33**	**0.13**	**0.87**

Sensitivity (*recall*) = $TP/(TP + FN)$—a proportion of correctly predicted epitope residues (TP) with respect to the total number of epitope residues (TP + FN).
Specificity = $1 - FP/(TN + FP)$—a proportion of correctly predicted nonepitope residues (TN) with respect to the total number of nonepitope residues (TN + FP).
Positive predictive value (*ppv*) (*precision*) = $TP/(TP + FP)$—a proportion of correctly predicted epitope residues (TP) with respect to the total number of predicted epitope residues (TP + FN).
The statistical significance of a prediction, that is, the difference between observed and expected frequencies of an actual epitope/nonepitope residue in the predicted epitope/nonepitope was determined by Fisher's exact test (right tailed). The prediction was considered significant if the significance level was $\geq$95%, that is, the *P-value* was $\leq$0.05. Significant predictions are shown in bold font.

Amino acid scale-based methods apply amino acid scales to compute the scores of a residue i in a given protein sequence. The $i-(n-1)/2$ neighboring residues on each side of residue i are used to compute the score for residue i in a window of size n. The final score for residue i is the average of the scale values for n amino acids in the window.

Chou and Fasman's method (Chou and Fasman, 1978) is based on calculating the probability of a stretch of residues to be a part of secondary structure β-turn. The first of the two tested protective epitopes was predicted with ranks seven and six when window size was seven and nine, respectively (Table 35.2).

Emini's solvent accessibility scores (Emini et al., 1985) were calculated based on surface accessibility scale, which has been determined by Janin and Wodak (Janin and Wodak, 1978) and reflected surface exposure probabilities for amino acids as calculated on X-ray structures of 28 proteins. The method defines a surface probability (S_n) at a sequence position n as a product of fractional surface probabilities for amino acids at positions from $n-2$ to $n+3$. Thus, S_n for a random hexapeptide is equal to 1.0 with probabilities greater than 1.0 indicating an increased probability for being found on the surface. In HA1, epitope 100–105 had score 1.951 (rank 37), while 128–133 epitope had score 1.058 (rank 129) (Table 35.2).

Karplus and Schulz's flexibility scale (Karplus and Schulz, 1985) was constructed on the basis of the mobility of protein segments derived from known temperature B factors of the C⟨ atoms of 31 proteins of known structure. The method did not predict the protective epitopes in HA1 (Table 35.2).

Kolaskar and Tongaonkar (Kolaskar and Tongaonkar, 1990) calculated the frequency of occurrence of each type of amino acid (f_{Ag}) in 156 experimentally determined epitopes from 34 different proteins. Then, using Parker's (Parker, Guo, and Hodges, 1986) scales, averaged values for hydrophilicity, accessibility, and flexibility were calculated for every overlapping heptapeptide in the 156 epitopes, and the frequency of occurrence of amino acids on the surface (f_{s}) was calculated. The antigenic propensity (A_p) value for each amino acid was calculated as $A_p = f_{\mathrm{Ag}}/f_{\mathrm{s}}$. About 75% of epitopes have been correctly predicted applying the antigenic propensity scale to the epitopes on which the scale has been built (Kolaskar and Tongaonkar, 1990). In HA1, the method predicted the 91–108 epitope, while the second protective epitope was not predicted (Table 35.2).

Parker's hydrophobicity scale (Parker, Guo, and Hodges, 1986) was determined experimentally using high-performance liquid chromatography (HPLC) on a set of 20 synthetic peptides accounting for each of the 20 amino acids. Hydrophobicity value for each amino acid was assigned according to the retention time for each peptide. In the method of calculating a score of a residue i, a window of seven residues was used. The corresponding value of the scale was introduced for each of the seven residues and the arithmetical mean of the seven residues values was assigned to the fourth, $(i+3)$, residue in the segment. In HA1, the method predicted the 101–107 epitope with the rank 49 (of 313 predictions) and the 127–133 epitope with the rank 130 (Table 35.2).

The aforementioned five scale-based methods have been evaluated on the data sets from Pellequer (Pellequer, Westhof, and Van Regenmortel, 1993), AntiJen (Toseland et al., 2005), and HIV (Los Alamos NL) databases. The values of the area under the receiver operating characteristic curve (A_{ROC}) for these methods did not exceed 0.60 (Blythe and Flower, 2005; Greenbaum et al., 2007). In addition, there was no significant correlation found between the location of known continuous epitopes and the sequence profiles generated using either single-scale-based methods or their combinations (Blythe and Flower, 2005).

The web server that implements the considered five scale-based methods is available at the IEDB web site (Peters et al., 2005).

ABCpred (Saha and Raghava, 2006) method applies artificial neural network for predicting continuous epitopes in protein sequences. The method has been trained on a nonredundant set of 700 epitopes from Bcipep database (Saha, Bhasin, and Raghava, 2005) and 700 randomly derived peptides from Swiss-Prot database. Tested on the trained data sets using fivefold cross-validation, the method achieved an accuracy of ∼66% (Saha and Raghava, 2006). The same accuracy has been observed for the independent data set of 187 epitopes from the structural database of allergenic proteins (SDAP) (Ivanciuc, Schein, and Braun, 2003) and epitopes excluded from training. In comparison to other methods, ABCpred performed significantly better than the methods based on Karplus and Schulz's flexibility scale (Karplus and Schulz, 1985) and Parker's hydrophobicity scale (Parker, Guo, and Hodges, 1986); the last two methods demonstrated the best accuracies of 59.4% and 61.5%, respectively, on the independent data set (Saha and Raghava, 2006). ABCpred predicted both HA1 protective epitopes among 31 and 28 predictions at the window sizes 16 and 10, respectively (Table 35.2).

BepiPred (Larsen, Lund, and Nielsen, 2006) method is based on a combination of hidden Markov model and two amino acid scales, Parker's hydrophilicity scale (Parker, Guo, and Hodges, 1986) and Levitt's secondary structure scale (Levitt, 1978). These scales have been chosen as the best performers on the Pellequer's (Pellequer, Westhof, and Van Regenmortel, 1993) data set among five scales: Parker's, Levitt's, Emini's accessibility (Emini et al., 1985), hydrophibicity (Kyte and Doolittle, 1982), and antigenicity (Welling et al., 1985). The estimated *P-values* for the hypothesis that Parker's and Levitt's scale-based methods performed as random were below 0.1%. The HMM method built on the epitopes from AntiJen (Toseland et al., 2005) database gave $A_{\text{roc}} = 0.66 \pm 0.01$ on the Pellequer's data set. On the HIV data set, BepiPred performed the best among amino acid scale-based methods, HMM, and BepiPred. Thus, Levitt's and Parker's scales alone and their combination gave areas under ROC curve of 0.57, 0.59, and 0.58, respectively; HMM gave 0.59; and BepiPred gave 0.6 that was significantly better than random prediction (Larsen, Lund, and Nielsen, 2006).

BepiPred assigns a score value to each protein residue; for example, for HA1 the maximum score was 1.94, and the minimum score was −1.28 (the average score was 0.25). The method's specificity and sensitivity depend on the score threshold. Thus, on the benchmark data set of 85 epitopes and 85 random sequences (nonepitopes), the method gave at thresholds of −0.20, 0.35, and 1.30 sensitivities and specificities, 75% and 50%, 49% and 75%, and 13% and 96%, respectively. Sensitivity was calculated as a proportion of correctly predicted epitopes with respect to the number of all epitopes; specificity was calculated as a proportion of correctly predicted nonepitopes with respect to the number of all nonepitopes. At the thresholds of 0.35 and 0.90, both protective epitopes in HA1 were predicted among 16 and 5 predictions, respectively, while no epitope was predicted at the threshold of 1.30 (Table 35.2).

Currently, the achieved accuracy of continuous epitope prediction is ∼60–66%, applying either combinations of amino acid scales or sophisticated machine-learning techniques. The higher accuracy could possibly be achieved by improving the quality of existing B-cell epitope data sets, which contain many erroneously delineated epitopes (Greenbaum et al., 2007). Thus, the location of continuous epitopes, which were taken from the AntiJen database and mapped to 42 proteins with known 3D structures, correlated poorly with solvent accessibilities, secondary structure elements, predicted locations of

disordered regions, and post-translational modifications present on the protein surfaces (Blythe, 2006). Such a finding confirms earlier suggestions (Laver et al., 1990) that experimental methods of epitope delineation often lead to incorrect epitope identification.

Structure-Based Methods

Six different prediction methods based on three-dimensional structures of antigens and implemented as web servers are discussed here, including the only two existing methods developed specifically for epitope prediction, CEP (Kulkarni-Kale, Bhosle, and Kolaskar, 2005) and DiscoTope (Haste Andersen, Nielsen, and Lund, 2006); two protein–protein docking methods, DOT (Mandell et al., 2001) and PatchDock (Schneidman-Duhovny et al., 2003); and two structure-based methods for protein–protein binding site prediction, PPI-PRED (Bradford, 2005) and ProMate (Neuvirth, Raz, and Schreiber, 2004). The results of prediction are demonstrated in the example of four structural epitopes in HA1 (Figure 35.2, Table 35.3).

CEP web server (Kulkarni-Kale, Bhosle, and Kolaskar, 2005) predicts both continuous and discontinuous epitopes in a given three-dimensional structure. The algorithm performs calculation of solvent accessibility of residues using the Voronoi polyhedron and delineates continuous epitopes based on the presence of at least three contiguous accessible residues. Discontinuous epitopes are predicted by collapsing predicted continuous epitopes for which Cα atoms are within a distance of 6 Å. The authors claimed (Kulkarni-Kale, Bhosle, and Kolaskar, 2005) that CEP correctly predicted structural epitopes in 60% structures of 21 analyzed antibody–protein complexes. In HA1, CEP predicted (with no scores assigned) 17 nonoverlapping continuous epitopes of different lengths varying 4–28 residues and covering 55% of HA1 protein sequence. Both protective HA1 epitopes were predicted (Table 35.2).

The number of discontinuous epitopes predicted by CEP in a protein varied and was related to the protein size: for more than half of 48 proteins analyzed in the work (Ponomarenko and Bourne, 2007), CEP gave more than 5 predictions and for a quarter of proteins, more than 10 predictions. Since CEP predictions were not ranked, the best (by *P-value*) prediction was reported for each epitope in Table 35.3. In HA1, CEP predicted 14 epitopes, along with the tested structural epitopes (Table 35.3).

DiscoTope (Haste Andersen, Nielsen, and Lund, 2006) is a method focusing explicitly on discontinuous epitope prediction. The method has been tested and trained on a data set of 76 X-ray structures of antibody–protein complexes divided into 25 groups by protein sequence homology, which were further split into 5 data sets used for cross-validated training and evaluation of the method. DiscoTope method is based on a linear combination (with equal weights) of the normalized values (obtained by subtracting the mean and dividing with the standard deviation) of the following measures for each residue: Parker's hydrophilicity value (Parker, Guo, and Hodges, 1986), amino acid statistics, number of contacts, and area of relative solvent accessibility. To measure amino acid statistics, the authors used log-odds ratios of each of 20 amino acids calculated based on their representation in all 9mer peptides in the analyzed antigens. A_{ROC} values averaged over all analyzed antigens measured the methods performance. DiscoTope gave A_{ROC} value of 0.71 and performed significantly better than the methods based on either single measure or any combination of two measures.

DiscoTope calculates a score value for each protein residue and assigns a residue being a part of a predicted epitope depending on the score threshold; one epitope is

predicted per protein. The specificity and sensitivity of the method were measured for each protein counting epitope and nonepitope residues using formulas provided in Table 35.3 footnote and then averaged over all proteins. Thus, at the thresholds of −3.1 the sensitivity and specificity were 15.5% and 95%, at the threshold of −7.7, 47.3%, and 75%, respectively (Haste Andersen, Nielsen, and Lund, 2006). At the threshold of −7.7, DiscoTope predicted in HA1 the only epitope, HC63, which was not used for the method development while, surprisingly, it did not predict the BH151 and HC45 epitopes, which were in the training data set (Table 35.3).

Protein–protein docking algorithms generate models of antibody–protein complexes for given structures of an antibody and a protein and assign a score for each model; the best model that closely represents the original complex will have the highest score. When antibody and antigen structures are used from the same original complex, the docking is called ''bound'', otherwise it is ''unbound.'' Unbound docking usually performs poorly on antibody–protein complexes (Duhovny, Nussinov, and Wolfson, 2002).

The DOT algorithm (Mandell et al., 2001) performs rigid-body docking based on the electrostatic and van der Waals energy calculation of interacting molecules and fast Fourier Transform correlation approach, which is used to produce a convolution of calculated energies. The method is implemented at the ClusPro web server (Comeau et al., 2004). PatchDock (Schneidman-Duhovny et al., 2003) is also a rigid-body protein–protein docking algorithm based on local shape feature matching. The algorithm filters out the models in which the paratope does not intersect the CDR regions, which are detected by aligning the sequence of a given antibody to a consensus sequence from a library of antibodies (Duhovny, Nussinov, and Wolfson, 2002). In HA1, using ''bound'' docking, DOT and PatchDock performed similarly: epitopes are often not predicted by the first models; by taking into account the top 10 models, both methods predicted only two epitopes, HC45 and HC19 with, however, low sensitivity and positive-predictive values (ppvs) (Table 35.3).

Protein–Protein Binding Site Prediction Methods. The motivation to review and test protein–protein binding site prediction methods on structural epitopes lies in the observation that antibody–protein interactions are similar to other transient protein–protein interactions by their physical–chemical nature: they can be categorized as intermediate by affinity, transient nonobligate interactions (Janin and Wodak, 1978; Jones and Thornton, 1996; Lo Conte, Chothia, and Janin, 1999). In the past, general methods for the prediction of intermediate transient nonobligate protein–protein interactions have been applied to structural epitopes (Jones and Thornton, 1996; Lo Conte, Chothia, and Janin, 1999). Also, epitopes and other protein–protein interfaces share many properties. Thus, Blythe (Blythe, 2006) compared 57 protein–protein binding interfaces of 44 proteins (Neuvirth, Raz, and Schreiber, 2004) with epitopes and paratopes inferred from X-ray structures of 37 complexes and calculated the following interface properties: amino acid composition, hydrophobicity by the Eisenberg's scale (Eisenberg, Weiss, and Terwilliger, 1984), amino acid contribution to form intermolecular hydrogen bonds, residue evolutionary conservancy, and geometrical parameters, such as planarity and complementarity of interfaces. Epitopes and nonobligate heterodimer interfaces were similar considering all aforementioned properties except residue conservancy; epitope residues were more variable than heterodimer interfaces (Blythe, 2006). Independent analysis of 55 representative structural epitopes showed that by residue evolutionary conservation scores calculated by ConSurf (Landau et al., 2005), epitope residues were

significantly less evolutionarily conserved than average protein surfaces (Ponomarenko and Bourne, 2007).

ProMate (Neuvirth, Raz, and Schreiber, 2004) is a method based on structural and sequence properties of protein interfaces from a manually curated data set of 67 proteins involved in heterodimeric transient interactions. ProMate is built on the following properties of interfaces: (i) preference of hydrophobic and polar residues; (ii) preference of β-structure, long loops and coil regions and disfavoring α-helixes; (iii) higher degree of evolutionary conservancy of residues at the interface compared to other surface residues; (iv) lower temperature factor than for protein surfaces in general; and (v) preference of interfaces to be solvated by more bound water molecules than protein surfaces in general. The authors reported a success rate of 70% (Neuvirth, Raz, and Schreiber, 2004). In HA1, no epitope was predicted by ProMate (Table 35.3).

PPI-PRED (Bradford, 2005) is a protein–protein binding site prediction method that is also based on characteristic properties of protein interfaces from manually curated data set of 180 proteins from 149 obligate and transient complexes. Antibody–protein complexes were excluded from both ProMate and PPI-PRED data sets, which, surprisingly, shared only 15 proteins. PPI-PRED is based on a support vector machine (SVM) training on seven properties of protein interfaces: residue interface propensity, amino acid hydrophobicity scores, residue evolutionary conservancy, solvent accessibility, electrostatic potential, interface surface shape index, and curvedness. The reported success rate of the method was 76% (Bradford, 2005). In HA1, PPI-PRED successfully predicted all structural epitopes (Table 35.3).

The considered six structure-based methods have been evaluated on the benchmark data set of 59 epitopes inferred from X-ray structures of 48 one-chain (monomer) proteins in complexes with variable fragments of two-chain antibodies (Ponomarenko and Bourne, 2007). The overall performance of the methods did not exceed 40% ppv (precision) at 46% sensitivity (recall) and A_{ROC} value of 0.70 (Ponomarenko and Bourne, 2007). When the top 10 models were considered, the docking methods performed the best. Despite the fact that structural epitopes and protein–protein nonobligate transient heterodimer interfaces share many properties (Blythe, 2006), protein–protein binding site prediction methods, ProMate and PPI-PRED, poorly predicted structural epitopes (Ponomarenko and Bourne, 2007).

Another evaluation of PPI-PRED and four docking algorithms (DOT, PatchDock, ZDOCK, and webGRAMM) have been done on 37 structural epitopes from the Blythe's data set (Blythe, 2006). Using the Matthew's Correlation Coefficient (MCC), Blythe measured the correlation between predicted and structural epitopes and paratopes. For the first models, all evaluated methods demonstrated near-random correlations. Likewise, when the top 10 models for each complex were considered, low and negative MCC values prevailed over positive values for all algorithms except PatchDock, which gave average MCC values of 0.540 for epitopes and 0.694 for paratopes. The last method succeeded because of application of the CDR filter, which other algorithms do not use; if they did, the results would be comparable. Thus, using predefined CDRs for antibodies, DOT method showed MCC values of 0.526 for epitopes and 0.783 for paratopes compared with 0.293 and 0.488, respectively, when no information on CDRs was used (Blythe, 2006). Overall, docking algorithms gave a ∼50% likelihood of resembling the native antibody–protein complex structure when the top 10 models were considered (Blythe, 2006).

There are several potential strategies that can be explored to improve current epitope prediction methods. First, prediction methods would be through training of the

methods on data sets containing only well-defined epitopes inferred from X-ray structures of antibody–protein complexes. Second, the performance of docking algorithms could be improved by developing them specifically for antibody–antigen complexes. Third, features that discriminate epitopes from nonepitopes, for example, the evolutionary conservation score can be used as additional features for discrimination. Finally, the observation that the average frequencies of amino acids in epitopes and randomly selected protein surface patches are not similar can be used as leverage for discrimination. Although, Blythe analyzed single epitopes in 37 protein antigens (Blythe, 2006) and found no differences in the frequency of amino acid types in epitopes, the study did not account for randomly selected patches that may contain other epitopes. Instead, analysis of multiple epitopes was performed for 52 representative antigens with consideration of whether the surface residue is a part of any structural epitope inferred from all known X-ray structures of antibodies complexed with the same antigen (187 X-ray structures were analyzed; the data set is provided in Ponomarenko and Bourne, 2007). This study shows that hydrophobic residues were infrequent among epitope residues in comparison to surface residues in general, while charged residues were abundant in epitopes (95% certainty); also, epitope residues are frequently located in β-turns and bends, but not in coil regions (68% certainty; comparison with antigen surface residues). Thus, as more X-ray structures on antibody–protein complexes become available, more discriminatory features will be revealed through a more precise analysis of epitopes and, hence, better epitope predictors may become available.

APPLICATIONS

Prediction of epitopes has two major applications: immunodiagnostics and vaccines. Diagnostic peptides representing antigenic epitopes of a molecule are utilized for the detection of antibodies that cross-react with this molecule. These peptides can also be used for production of antipeptide cross-reactive antibodies that, in turn, can serve as diagnostic tools. Peptides representing cross-reactive immunogenic epitopes are attractive candidates for prophylactic and therapeutic vaccines.

Diagnostic Peptides Representing Antigenic Epitopes

Peptide-based tests have many advantages over immunoassays based on whole molecules because such tests are safe, standardized, specific, reproducible, and suitable for large-scale analysis *in vitro*. Synthetic peptides that mimic epitopes have been used in diagnostic systems for various human diseases, including cancers, allergies, autoimmune diseases, viral, bacterial, and parasitic infections (Leinikki et al., 1993; Meloen et al., 2003; Gomara and Haro, 2007).

The safety of peptides compared to pathogens' molecules is the main advantage of using peptides in the diagnostics of infectious agents. Thus, a peptide-based enzyme-linked immunosorbent assay (ELISA) has been applied for early diagnostics of severe acute respiratory syndrome (SARS) (Hsueh et al., 2004). Demonstrating 100% specificity, this assay was based on synthetic peptides representing continuous epitopes from the spike, membrane, and nucleocapsid proteins of SARS-associated coronavirus. Synthetic peptides have also opened new perspectives for diagnosis of parasitic infections, such as malaria, Chagas' disease, leishmaniasis, and many others (Noya et al., 2003). Diagnostic peptides

can also be isolated from phage-displayed libraries by affinity selection with serum antibodies from patients, for example, with Lyme disease (Hamby et al., 2005). The sequence similarity search between such isolated peptides and the pathogen proteins uncovered a new epitope in the VlsE antigen of *Borrelia burgdorferi*. The identification of this epitope demonstrated the potential of peptide library screening, even though its diagnostic sensitivity was relatively low compared to the experimentally identified epitopes (Hamby et al., 2005).

Synthetic peptides are also attractive diagnostic tools for allergies and autoimmune diseases. Epitopes of autoantigens may detect autoantibodies against intracellular autoantigens present during autoimmune diseases; autoantibodies help establish disease prognosis (Fournel and Muller, 2003; Routsias, Vlachoyiannopoulos, and Tzioufasm, 2006). One of the proposed therapeutics for autoimmune diseases is based on the design of molecules that are complementary to autoantigen epitopes (Routsias, Vlachoyiannopoulos, and Tzioufasm, 2006). Synthetic peptides are also considered useful *in vitro* diagnostic tools for allergies since they allow a diagnosis without the risks associated with tests based on *in vivo* allergen challenges (Eigenmann, 2004).

Synthetic Vaccines

A major reason for analyzing and predicting epitopes is because they may lead to the development of peptide-based synthetic vaccines. Many studies have been carried out to develop synthetic prophylactic vaccines against numerous viral, bacterial, and parasitic infections as well therapeutic vaccines for chronic infections and noninfectious diseases, including autoimmune diseases, various neurological disorders, allergies, and cancers (Arnon, 1987; Nicholson, 1994; Arnon and Ben-Yedidia, 2003; Hans et al., 2006). Thousands of peptides have been preclinically examined; over 100 of them have progressed to phase I clinical trials and about 30 to phase II. However, not a single peptide vaccine has passed phase III and became available to the public (Hans et al., 2006). Some notable successes have been obtained in the development of vaccines based on T-cell epitopes (De Groot and Martin, 2003; Flower et al., 2003; Larche and Wraith, 2005; Jiang et al., 2006), which are not discussed in the current review. The software and database developed to facilitate vaccine design are reviewed in (Davies and Flower, 2007; Taylor and Flower, 2007).

In order to be considered a vaccine candidate, a peptide must first demonstrate antigenicity. Furthermore, its possible immunogenicity must be experimentally tested. However, adequate immunogenicity is a substantially rarer phenomenon than antigenicity for short peptides. To induce an immune response, peptides shorter than 10–15 amino acids usually require to be covalently bound to a carrier protein or a liposome. Also, to achieve sufficient immunogenicity, the synthetic vaccine should include T-cell epitopes. Furthermore, short peptides tend to rapidly degrade. This problem can be solved through stabilization of peptide conformations by cyclization or other chemical procedures (Sundaram et al., 2004; Dakappagari et al., 2005; Hans et al., 2006). Peptide stability can also be improved by replacing L-peptides by ''retroinverso'' (RI) peptides composed of D-amino acids assembled in reverse order (Muller et al., 1998). For example, the RI peptide analogue of gonadotropin-releasing hormone (GnRH) was shown to be a potential contraceptive vaccine (Fromme et al., 2003).

The majority of peptides examined as vaccine candidates correspond to continuous epitopes. However, most epitopes in proteins are discontinuous, and methods to predict such

epitopes identify only a certain number of residues that are part of the discontinuous structure without indicating how these residues could be assembled by synthesis to reconstitute an active epitope. Every attempt at reconstituting a discontinuous epitope by synthesis is therefore a unique endeavor, with little success so far (Enshell-Seijffers et al., 2003; Oomen et al., 2003; Villen et al., 2004; Dakappagari et al., 2005; Timmerman et al., 2005), and it will remain to be seen whether the prediction of discontinuous epitopes will be useful for developing synthetic vaccines. Several attempts to develop synthetic vaccines against viral infections are briefly described in this chapter. For the peptide-based approaches to the development of cancer vaccines see a review (Sundaram, Dakappagari, and Kaumaya, 2002).

Foot-and-mouth virus (FDMV) infection of cattle and pigs occurs throughout the world and causes huge economic losses. Most studies to replace conventional FDMV vaccines by peptide-based vaccines have concentrated on two continuous epitopes of the viral VP1 protein located on a loop (residues 140–160; site A) and the C-terminal region (residues 201–213; site C). Conjugated to carriers, peptides corresponding to these epitopes elicited neutralizing antibodies in guinea pigs (Bittle et al., 1982; DiMarchi et al., 1986), while immunization of cattle with these peptide epitopes gave viral protection lower than 50% (Taboga et al., 1997). The discontinuous epitope corresponding to the site D formed by residues from the VP1, VP2, and VP3 proteins has also been actively studied; however, the synthetic construction that mimicked this epitope did not show the protection level required for an effective vaccine (Villen et al., 2004). The high rate of mutations present in FMDV constitutes a major problem for vaccine development.

A major worldwide health threat is influenza, which in the United States alone kills over 40,000 people yearly. The virus escapes the immune system by undergoing continuous antigenic variation, especially in the hemagglutinin (HA) surface glycoprotein. Two continuous epitopes of HA1 located in residues 91–108 and 139–147 have been studied as vaccine candidates. Conjugated to the tetanus toxoid, the 91–108 epitope elicited an immune response in rabbits and mice that partially protected those animals against influenza (Muller, Shapira, and Arnon, 1982). The addition of two conserved T-cell epitopes of the influenza nucleoprotein to the experimental vaccine increased the protection level (Levi et al., 1995; Levi and Arnon, 1996). Studies with human/mouse radiation chimeras showed that the recombinant vaccine based on the mixture of these peptides had considerable potential for use in humans (Ben-Yedidia and Arnon, 2005).

Canine parvovirus causing enteritis and myocarditis in dogs and minks is the first veterinary virus for which an effective synthetic peptide vaccine has been developed. Several peptides from the N-terminal region of the viral VP2 protein (residues 1–15, 7– 1, and 3–19) coupled to a carrier induced an immune response in dogs and minks and protected them against infection (Langeveld et al., 1994; Casal et al., 1995; Langeveld et al., 1995). The success achieved with these peptides is probably due to the high flexibility of the N-terminal region of VP2 that facilitates the cross-reactivity between peptides and virus particles.

In spite of global research efforts, no vaccine against HIV has yet been developed (Koff, Kahn, and Gust, 2007). Many attempts have been made to design HIV peptide vaccines using structural data on HIV proteins in complexes with neutralizing monoclonal antibodies (Burton et al., 2004). One of them involves the continuous epitope ELDKWAS corresponding to the membrane-proximal external region of the gp41 protein. This epitope is widely regarded as a promising vaccine candidate because it is recognized by the anti-HIV broadly cross-reactive neutralizing monoclonal antibody 2F5 and is located in the

conserved region of the gp41, which is necessary for the fusion of the virus with the host cell membrane. When incorporated into various synthetic constructs, the ELDKWAS peptide improved its binding to the 2F5 antibody, although it did not induce neutralizing antibodies (Ho et al., 2005). In efforts to assess which structural elements close to the epitope influenced the neutralizing properties of the antibody, the crystal structures of 2F5 complexed with 7mer, 11mer, and 17mer peptides of the gp41 incorporating the ELDKWAS sequence have been solved (Ofek et al., 2004). In all complexes, the conformations of the peptides were significantly different from the conformation of the epitopic region in the gp41. Moreover, the epitopic region in the gp41 demonstrates a high degree of conformational flexibility, and its conformation depends on the fusogenic state of the protein. Therefore, it is not clear which conformation should be stabilized in the synthetic constructs intended for vaccination (Van Regenmortel, 2007). Such transitional epitopes with variable conformations represent a major challenge in the development of synthetic vaccines (Zolla-Pazner, 2004).

This lack of success in developing synthetic peptide-based vaccines suggests that some of the assumptions underlying this type of research are probably unsound and that novel approaches should be investigated if synthetic vaccines are to become a reality.

CONCLUSION

This chapter summarized current attempts at prediction of epitopes using bioinformatics tools. Although combining sequence-based prediction methods with 3D structural information was found to lead to slightly better predictions, the improvement was still rather modest.

There are three main reasons for the low success rate of epitope predictions. First, it is difficult to determine to what extent a protein reacting with an antibody has retained its native conformation, meaning that even structural information is an unreliable guide for epitope prediction. Furthermore, proteins are dynamic structures able to undergo significant conformational rearrangements when binding to antibodies.

Second, peptides, called continuous epitopes, are usually poor mimics of discontinuous epitopes—the actual atomic configurations that antibodies recognize at the surface of proteins. Furthermore, the existing databases of continuous epitopes are known to be of dubious value for assessing the reliability of epitope prediction.

Third, discontinuous epitopes, which are the most common ones found in proteins, can be defined only in terms of residues in contact with the antibody, and they cannot be isolated as independent entities that possess binding activity outside of the protein context in which they are embedded. The so-called prediction of discontinuous epitopes usually consists in predicting that certain surface residues are likely to be a part of an epitope, but it does not entail predicting the minimum set of combined residues in a defined configuration that is likely to possess immunological activity. Furthermore, attempts to reconstitute active discontinuous epitopes by chemical assembly of a limited number of their constituent residues have so far met with little success.

As a result, predicting discontinuous epitopes has few applications at present. It can only be hoped that the development of new bioinformatics tools will improve our ability to predict which synthetic peptides are likely to be useful constructs for replacing intact proteins as antigens and immunogens.

ABBREVIATIONS

CDR	Complementary Determined Region of the Antibody
Fab	antigen-binding fragment of antibody that includes one complete light chain paired with one heavy chain fragment containing the variable domain and the first constant domain
A_{ROC}	area under the receiver operating characteristic curve
TP, FN, TN, FP	true positives, false negatives, true negatives, and false positives, respectively.

ACKNOWLEDGMENT

JP work was supported by the National Institutes of Health Contract HHSN26620040006C.

Q1

REFERENCES

Alix AJ (1999): Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine* 18(3–4):311–314.

Amit AG, Mariuzza RA, Phillips SE, Poljak RJ (1986): Three-dimensional structure of an antigen–antibody complex at 2.8 Å resolution. *Science* 233(4765):747–753.

Arnon R (1987): *Synthetic Vaccines*. Boca Raton, FL: CRC Press.

Arnon R, Ben-Yedidia T (2003): Old and new vaccine approaches. *Int Immunopharmacol* 3(8): 1195–1204.

Bajaj K, Chakrabarti P, Varadarajan R (2005): Mutagenesis-based definitions and probes of residue burial in proteins. *Proc Natl Acad Sci USA* 102(45):16221–16226.

Batori V, Friis EP, Nielsen H, Roggen EL (2006): An *in silico* method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context. *J Mol Recognit* 19(1):21–29.

Ben-Yedidia T, Arnon R (2005): Towards an epitope-based human vaccine for influenza. *Hum Vaccin* 1(3):95–101.

Benjamin DC, Perdue SS (1996): Site-directed mutagenesis in epitope mapping. *Methods* 9(3): 508–515.

Bentley GA, Boulot G, Chitarra V (1994): Cross-reactivity in antibody–antigen interactions. *Res Immunol* 145(1):45–48.

Berger C, Weber-Bornhauser S, Eggenberger J, Hanes J, Pluckthun A, Bosshard HR (1999): Antigen recognition by conformational selection. *FEBS Lett* 450(1–2):149–153.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000): The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.

Berzofsky JA (1985): Intrinsic and extrinsic factors in protein antigenic structure. *Science* 229(4717): 932–940.

Bittle JL, Houghten RA, Alexander H, Shinnick TM, Sutcliffe JG, Lerner RA, Rowlands DJ, Brown F (1982): Protection against foot-and-mouth disease by immunization with a chemically synthesized peptide predicted from the viral nucleotide sequence. *Nature* 298(5869):30–33.

Blythe MJ, Flower DR (2005): Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci* 14(1):246–248.

Blythe MJ (2006): Computational Characterisation of B Cell Epitopes. PhD thesis. School of Animal and Microbial Sciences, The Edward Jenner Institute for Vaccine Research, The University of Reading. p. 243.

Bosshard HR (2001): Molecular recognition by induced fit: how fit is the concept?*News Physiol Sci* 16:171–173.

Braden BC, Souchon H, Eisele JL, Bentley GA, Bhat TN, Navaza J, Poljak RJ (1994): Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J Mol Biol* 243(4):767–781.

Braden BC, Goldman ER, Mariuzza RA, Poljak RJ (1998): Anatomy of an antibody molecule: structure, kinetics, thermodynamics and mutational studies of the antilysozyme antibody D1.3. *Immunol Rev* 163:45–57.

Braden BC, Poljak RJ (2000): Structure and energetics of anti-lysozyme antibodies, *Protein–Protein Recognition*. In: Kleanthous C, editor. New York: Oxford University Press, pp. 126–161.

Bradford JR, David WD, Westhead R (2005): Improved prediction of protein–protein binding sites using a support vector machines approach. *Bioinformatics* 21(8):1487–1494.

Bublil EM, Freund NT, Mayrose I, Penn O, Roitburd-Berman A, Rubinstein ND, Pupko T, Gershoni JM (2007): Stepwise prediction of conformational discontinuous B-cell epitopes using the mapitope algorithm. *Proteins* 68(1):294–304.

Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A (2007): Ab and T cell epitopes of influenza A virus, knowledge and opportunities. *Proc Natl Acad Sci USA* 104(1):246–251.

Burton DR, Desrosiers RC, Doms RW, Koff WC, Kwong PD, Moore JP, Nabel GJ, Sodroski J, Wilson IA, Wyatt RT (2004): HIV vaccine design and the neutralizing antibody problem. *Nat Immunol* 5(3): 233–236.

Carneiro J, Stewart J (1994): Rethinking ''shape space'': evidence from simulated docking suggests that steric shape complementarity is not limiting for antibody–antigen recognition and idiotypic interactions. *J Theor Biol* 169(4):391–402.

Casal JI, Langeveld JP, Cortes E, Schaaper WW, van Dijk E, Vela C, Kamstrup S, Meloen RH (1995): Peptide vaccine against canine parvovirus: identification of two neutralization subsites in the N terminus of VP2 and optimization of the amino acid sequence. *J Virol* 69(11):7274–7277.

Castrignano T, De Meo PD, Carrabino D, Orsini M, Floris M, Tramontano A (2007): The MEPS server for identifying protein conformational epitopes. *BMC Bioinformatics* 8:(Suppl 1): S6.

Chavali GB, Papageorgiou AC, Olson KA, Fett JW, Hu G, Shapiro R, Acharya KR (2003): The crystal structure of human angiogenin in complex with an antitumor neutralizing antibody. *Structure* 11(7):875–885.

Chou PY, Fasman GD (1974): Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry* 13(2):211–222.

Chou PY, Fasman GD (1978): Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* 47:45–148.

Collis AV, Brouwer AP, Martin AC (2003): Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. *J Mol Biol* 325(2):337–354.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004): ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 20(1):45–50.

Cunningham BC, Wells JA (1993): Comparison of a structural and a functional epitope. *J Mol Biol* 234(3):554–563.

Dakappagari NK, Lute KD, Rawale S, Steele JT, Allen SD, Phillips G, Reilly RT, Kaumaya PT (2005): Conformational HER-2/neu B-cell epitope peptide vaccine designed to incorporate two native disulfide bonds enhances tumor cell binding and antitumor activities. *J Biol Chem* 280(1):54–63.

Davies MN, Flower DR (2007): Harnessing bioinformatics to discover new vaccines. *Drug Discov Today* 12(9–10):389–395.

De Groot AS, Martin W (2003): From immunome to vaccine: epitope mapping and vaccine design tools. *Novartis Found Symp* 254:57–72; discussion 72–76, 98–101, 250–252.

Decanniere K, Transue TR, Desmyter A, Maes D, Muyldermans S, Wyns L (2001): Degenerate interfaces in antigen–antibody complexes. *J Mol Biol* 313(3):473–478.

DiMarchi R, Brooke G, Gale C, Cracknell V, Doel T, Mowat N (1986): Protection of cattle against foot-and-mouth disease by a synthetic peptide. *Science* 232(4750):639–641.

Duhovny D, Nussinov R, Wolfson HJ (2002): Efficient unbound docking of rigid molecules. 2nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452: pp. 185–200.

Eigenmann PA (2004): Do we have suitable *in-vitro* diagnostic tests for the diagnosis of food allergy? *Curr Opin Allergy Clin Immunol* 4(3):211–213.

Eisenberg D, Weiss RM, Terwilliger TC (1984): The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA* 81(1):140–144.

Emini EA, Hughes JV, Perlow DS, Boger J (1985): Induction of hepatitis A virus: neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55(3):836–839.

Enshell-Seijffers D, Denisov D, Groisman B, Smelyanski L, Meyuhas R, Gross G, Denisova G, Gershoni JM (2003): The mapping and reconstitution of a conformational discontinuous B-cell epitope of HIV-1. *J Mol Biol* 334(1):87–101.

Flower DR, McSparron H, Blythe MJ, Zygouri C, Taylor D, Guan P, Wan S, Coveney PV, Walshe V, Borrow P, Doytchinova IA (2003): Computational vaccinology: quantitative approaches. *Novartis Found Symp* 254:102–120; discussion 120–225, 216–222, 250–252.

Folgori A, Tafi R, Meola A, Felici F, Galfre G, Cortese R, Monaci P, Nicosia A (1994): A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *Embo J* 13(9):2236–2243.

Fournel S, Muller S (2003): Synthetic peptides in the diagnosis of systemic autoimmune diseases. *Curr Protein Pept Sci* 4(4):261–274.

Fromme B, Eftekhari P, Van Regenmortel M, Hoebeke J, Katz A, Millar R (2003): A novel retro-inverso gonadotropin-releasing hormone (GnRH) immunogen elicits antibodies that neutralize the activity of native GnRH. *Endocrinology* 144(7):3262–3269.

Garnier J, Osguthorpe DJ, Robson B (1978): Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 120(1):97–120.

Geysen HM, Meloen RH, Barteling SJ (1984): Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc Natl Acad Sci USA* 81(13): 3998–4002.

Geysen HM, Rodda SJ, Mason TJ, Tribbick G, Schoofs PG (1987): Strategies for epitope analysis using peptide synthesis. *J Immunol Methods* 102(2):259–274.

Goede A, Jaeger IS, Preissner R (2005): SUPERFICIAL: surface mapping of proteins via structure-based peptide library design. *BMC Bioinformatics* 6:223.

Gomara MJ, Haro I (2007): Synthetic peptides for the immunodiagnosis of human diseases. *Curr Med Chem* 14(5):531–546.

Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar AS, Lund O, Morrison S, Mumey B, Ofran Y, Pellequer JL, Pinilla C, Ponomarenko JV, Raghava GP, van Regenmortel MH, Roggen EL, Sette A, Schlessinger A, Sollner J, Zand M, Peters B (2007): Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit* 20(2):75–82.

Hager-Braun C, Tomer KB (2005): Determination of protein-derived epitopes by mass spectrometry. *Expert Rev Proteomics* 2(5):745–756.

Halperin I, Ma B, Wolfson H, Nussinov R (2002): Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 47(4):409–443.

Hamby CV, Llibre M, Utpat S, Wormser GP (2005): Use of peptide library screening to detect a previously unknown linear diagnostic epitope: proof of principle by use of lyme disease sera. *Clin Diagn Lab Immunol* 12(7):801–807.

Hans D, Young PR, Fairlie DP (2006): Current status of short synthetic peptides as vaccines. *Med Chem* 2(6):627–646.

Haste Andersen P, Nielsen M, Lund O (2006): Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 15(11):2558–2567.

Ho J, Uger RA, Zwick MB, Luscher MA, Barber BH, MacDonald KS (2005): Conformational constraints imposed on a pan-neutralizing HIV-1 antibody epitope result in increased antigenicity but not neutralizing response. *Vaccine* 23(13):1559–1573.

Hopp TP, Woods KR (1981): Prediction of Protein Antigenic Determinants from Amino Acid Sequences. *Proc Natl Acad Sci USA* 78:3824–3828.

Hopp TP, Woods KR (1983): A computer program for predicting protein antigenic determinants. *Mol Immunol* 20(4):483–489.

Hopp TP (1993): Retrospective: 12 years of antigenic determinant predictions, and more. *Pept Res* 6(4):183–190.

Hsueh PR, Kao CL, Lee CN, Chen LK, Ho MS, Sia C, Fang XD, Lynn S, Chang TY, Liu SK, Walfield AM, Wang CY (2004): SARS antibody test for serosurveillance. *Emerg Infect Dis* 10(9):1558–1562.

Huang J, Gutteridge A, Honda W, Kanehisa M (2006): MIMOX: a web tool for phage display based epitope mapping. *BMC Bioinformatics* 7:451.

Ivanciuc O, Schein CH, Braun W (2003): SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 31(1):359–362.

James LC, Roversi P, Tawfik DS (2003): Antibody multispecificity mediated by conformational diversity. *Science* 299(5611):1362–1367.

Jameson BA, Wolf H (1988): The antigenic index: a novel algorithm for predicting antigenic determinants. *Comput Appl Biosci* 4(1):181–186.

Janin J, Wodak S (1978): Conformation of amino acid side-chains in proteins. *J Mol Biol* 125(3):357–386.

Jiang S, Song R, Popov S, Mirshahidi S, Ruprecht RM (2006): Overlapping synthetic peptides as vaccines. *Vaccine* 24(37–39):6356–6365.

Johnson MA, Pinto BM (2004): NMR spectroscopic and molecular modeling studies of protein–carbohydrate and protein–peptide interactions. *Carbohydr Res* 339(5):907–928.

Jones S, Thornton JM (1996): Principles of protein–protein interactions. *Proc Natl Acad Sci USA* 93(1):13–20.

Jones S, Thornton JM (1997): Prediction of protein–protein interaction sites using patch analysis. *J Mol Biol* 272(1):133–143.

Jones S, Thornton J (2000): Analysis and classification of protein–protein interactions from a structural perspective. In: Kleanthous C, editor. *Protein–Protein Recognition*. New York: Oxford University Press, pp. 33–59.

Karplus PA, Schulz GE (1985): Prediction of chain flexibility in proteins. A tool for the selection of peptide antigens. *Naturwissenschaften* 72:S.212.

Koff WC, Kahn P, Gust ID (2007): *AIDS Vaccine Development: Challenges and Opportunities.* Caister Academic Press.

Kolaskar AS, Tongaonkar PC (1990): A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276(1–2):172–174.

Kolaskar AS, Kulkarni-Kale U (1999): Prediction of three-dimensional structure and mapping of conformational epitopes of envelope glycoprotein of Japanese encephalitis virus. *Virology* 261(1):31–42.

Kulkarni-Kale U, Bhosle S, Kolaskar AS (2005): CEP: a conformational epitope prediction server. *Nucleic Acids Res* 33 (Web server issue):W168–W171.

Kyte J, Doolittle RF (1982): A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157(1):105–132.

Lachumanan R, Devi S, Cheong YM, Rodda SJ, Pang T (1993): Epitope mapping of the Sta58 major outer membrane protein of *Rickettsia tsutsugamushi*. *Infect Immun* 61(10): 4527–4531.

Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N (2005): ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33 (Web server issue):W299–302.

Langeveld JP, Casal JI, Osterhaus AD, Cortes E, de Swart R, Vela C, Dalsgaard K, Puijk WC, Schaaper WM, Meloen RH (1994): First peptide vaccine providing protection against viral infection in the target animal: studies of canine parvovirus in dogs. *J Virol* 68(7):4506–4513.

Langeveld JP, Kamstrup S, Uttenthal A, Strandbygaard B, Vela C, Dalsgaard K, Beekman NJ, Meloen RH, Casal JI (1995): Full protection in mink against mink enteritis virus with new generation canine parvovirus vaccines based on synthetic peptide or recombinant protein. *Vaccine* 13(11): 1033–1037.

Larche M, Wraith DC (2005): Peptide-based therapeutic vaccines for allergic and autoimmune diseases. *Nat Med* 11 (Suppl 4):S69–S76.

Larsen JE, Lund O, Nielsen M (2006): Improved method for predicting linear B-cell epitopes. *Immunome Res* 2:2.

Laver WG, Air GM, Webster RG, Smith-Gill SJ (1990): Epitopes on protein antigens: misconceptions and realities. *Cell* 61(4):553–556.

Leinikki P, Lehtinen M, Hyoty H, Parkkonen P, Kantanen ML, Hakulinen J (1993): Synthetic peptides as diagnostic tools in virology. *Adv Virus Res* 42:149–186.

Lerner RA (1984): Antibodies of predetermined specificity in biology and medicine. *Adv Immunol* 36:1–44.

Levi R, Aboud-Pirak E, Leclerc C, Lowell GH, Arnon R (1995): Intranasal immunization of mice against influenza with synthetic peptides anchored to proteosomes. *Vaccine* 13(14):1353–1359.

Levi R, Arnon R (1996): Synthetic recombinant influenza vaccine induces efficient long-term immunity and cross-strain protection. *Vaccine* 14(1):85–92.

Levitt M (1978): Conformational preferences of amino acids in globular proteins. *Biochemistry* 17(20): 4277–4285.

Lo Conte L, Chothia C, Janin J (1999): The atomic structure of protein–protein recognition sites. *J Mol Biol* 285(5):2177–2198.

Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF (2001): Protein docking using continuum electrostatics and geometric fit. *Protein Eng* 14(2):105–113.

Mayrose I, Shlomi T, Rubinstein ND, Gershoni JM, Ruppin E, Sharan R, Pupko T (2007): Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. *Nucleic Acids Res* 35(1):69–78.

Meloen RH, Puijk WC, Langeveld JP, Langedijk JP, Timmerman P (2003): Design of synthetic peptides for diagnostics. *Curr Protein Pept Sci* 4(4):253–260.

Moreau V, Granier C, Villard S, Laune D, Molina F (2006): Discontinuous epitope prediction based on mimotope analysis. *Bioinformatics* 22(9):1088–1095.

Morris GE (1996): *Epitope Mapping Protocols.* Humana Press.

Muller GM, Shapira M, Arnon R (1982): Anti-influenza response achieved by immunization with a synthetic conjugate. *Proc Natl Acad Sci USA* 79(2):569–573.

Muller S, Benkirane N, Guichard G, Van Regenmortel MH, Brown F (1998): The potential of retro-inverso peptides as synthetic vaccines. *Expert Opin Investig Drugs* 7(9): 1429–1438.

Mumey BM, Bailey BW, Kirkpatrick B, Jesaitis AJ, Angel T, Dratz EA (2003): A new method for mapping discontinuous antibody epitopes to reveal structural features of proteins. *J Comput Biol* 10(3–4):555–567.

Murali R, Sharkey DJ, Daiss JL, Murthy HM (1998): Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: the Fab is directed against an intermediate in the helix-coil dynamics of the enzyme. *Proc Natl Acad Sci USA* 95(21):12562–12567.

Naruse H, Ogasawara K, Kaneda R, Hatakeyama S, Itoh T, Kida H, Miyazaki T, Good RA, Onoe K (1994): A potential peptide vaccine against two different strains of influenza virus isolated at intervals of about 10 years. *Proc Natl Acad Sci USA* 91(20):9588–9592.

Neuvirth H, Raz R, Schreiber G (2004): ProMate: a structure based prediction program to identify the location of protein–protein binding sites. *J Mol Biol* 338(1):181–199.

Nicholson BH (1994): *Synthetic Vaccines*. Oxford: Blackwell Scientific Publ,.

Novotny J, Handschumacher M, Haber E, Bruccoleri RE, Carlson WB, Fanning DW, Smith JA, Rose GD (1986): Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proc Natl Acad Sci USA* 83(2):226–230.

Noya O, Patarroyo ME, Guzman F, Alarcon de Noya B (2003): Immunodiagnosis of parasitic diseases with synthetic peptides. *Curr Protein Pept Sci* 4(4):299–308.

Odorico M, Pellequer JL (2003): BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. *J Mol Recognit* 16(1):20–22.

Ofek G, Tang M, Sambor A, Katinger H, Mascola JR, Wyatt R, Kwong PD (2004): Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J Virol* 78(19):10724–10737.

Oomen CJ, Hoogerhout P, Bonvin AM, Kuipers B, Brugghe H, Timmermans H, Haseley SR, van Alphen L, Gros P (2003): Immunogenicity of peptide-vaccine candidates predicted by molecular dynamics simulations. *J Mol Biol* 328(5):1083–1089.

Parker JM, Guo D, Hodges RS (1986): New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25(19): 5425–5432.

Pattnaik P (2005): Surface plasmon resonance: applications in understanding receptor–ligand interaction. *Appl Biochem Biotechnol* 126(2):79–92.

Pellequer JL, Westhof E, Van Regenmortel MH (1991): Predicting location of continuous epitopes in proteins from their primary structures. *Methods Enzymol* 203:176–201.

Pellequer JL, Westhof E (1993): PREDITOP: a program for antigenicity prediction. *J Mol Graph* 11(3): 204–210 191-2.

Pellequer JL, Westhof E, Van Regenmortel MH (1993): Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunol Lett* 36(1):83–99.

Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, Nemazee D, Ponomarenko JV, Sathiamurthy M, Schoenberger S, Stewart S, Surko P, Way S, Wilson S, Sette A (2005): The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* 3(3):e91.

Ponomarenko JV, Bourne PE (2007): Antibody–protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol* 7. (64).

Radford AJ, Wood PR, Billman-Jacobe H, Geysen HM, Mason TJ, Tribbick G (1990): Epitope mapping of the *Mycobacterium bovis* secretory protein MPB70 using overlapping peptide analysis. *J Gen Microbiol* 136(2):265–272.

Rini JM, Schulze-Gahmen U, Wilson IA (1992): Structural evidence for induced fit as a mechanism for antibody–antigen recognition. *Science* 255(5047):959–965.

Routsias JG, Vlachoyiannopoulos PG, Tzioufas AG (2006): Autoantibodies to intracellular autoantigens and their B-cell epitopes: molecular probes to study the autoimmune response. *Crit Rev Clin Lab Sci* 43(3):203–248.

Rowley MJ, O';Connor K, Wijeyewickrema L (2004): Phage display for epitope determination: a paradigm for identifying receptor–ligand interactions. *Biotechnol Annu Rev* 10:151–188.

Saha S, Bhasin M, Raghava GP (2005): Bcipep: a database of B-cell epitopes. *BMC Genomics* 6(1):79.

Saha S, Raghava GP (2006): Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65(1):40–48.

Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, Barzilai A, Dror O, Haspel N, Nussinov R, Wolfson HJ (2003): Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* 52(1):107–112.

Schreiber A, Humbert M, Benz A, Dietrich U (2005): 3D-Epitope-Explorer (3DEX): localization of conformational epitopes within three-dimensional structures of proteins. *J Comput Chem* 26(9): 879–887.

Sollner J (2006): Selection and combination of machine learning classifiers for prediction of linear B-cell epitopes on proteins. *J Mol Recognit* 19(3):209–214.

Sollner J, Mayer B (2006): Machine learning approaches for prediction of linear B-cell epitopes on proteins. *J Mol Recognit* 19(3):200–208.

Sundaram R, Dakappagari NK, Kaumaya PT (2002): Synthetic peptides as cancer vaccines. *Biopolymers* 66(3):200–216.

Sundaram R, Lynch MP, Rawale SV, Sun Y, Kazanji M, Kaumaya PT (2004): *De novo* design of peptide immunogens that mimic the coiled coil region of human T-cell leukemia virus type-1 glycoprotein 21 transmembrane subunit for induction of native protein reactive neutralizing antibodies. *J Biol Chem* 279(23):24141–24151.

Sundberg EJ, Mariuzza RA (2004): Antibody structure and recognition of antigen. In: Honjo AFT, Neuberger M, editor. *Molecular Biology of B Cells*. Elsevier Science (USA). pp. 491–509.

Taboga O, Tami C, Carrillo E, Nunez JI, Rodriguez A, Saiz JC, Blanco E, Valero ML, Roig X, Camarero JA, Andreu D, Mateu MG, Giralt E, Domingo E, Sobrino F, Palma EL (1997): A large-scale evaluation of peptide vaccines against foot-and-mouth disease: lack of solid protection in cattle and isolation of escape mutants. *J Virol* 71(4):2606–2614.

Tainer JA, Getzoff ED, Alexander H, Houghten RA, Olson AJ, Lerner RA, Hendrickson WA (1984): The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature* 312(5990):127–134.

Taylor PD, Flower DR (2007): Immunoinformatics and computational vaccinology: a brief introduction, In: Flower DR, Timmis J, editors. *Silico Immunology*. Springer, pp. 23–46.

Thornton JM, Edwards MS, Taylor WR, Barlow DJ (1986): Location of 'continuous' antigenic determinants in the protruding regions of proteins. *Embo J* 5(2):409–413.

Timmerman P, Beld J, Puijk WC, Meloen RH (2005): Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces. *Chembiochem* 6(5):821–824.

Toseland CP, Clayton DJ, McSparron H, Hemsley SL, Blythe MJ, Paine K, Doytchinova IA, Guan P, Hattotuwagama CK, Flower DR (2005): AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res* 1(1):4.

Tribbick G (2002): Multipin peptide libraries for antibody and receptor epitope screening and characterization. *J Immunol Methods* 267(1):27–35.

Tulip WR, Harley VR, Webster RG, Novotny J (1994): N9 neuraminidase complexes with antibodies NC41 and NC10: empirical free energy calculations capture specificity trends observed with mutant binding data. *Biochemistry* 33(26):7986–7997.

Van Regenmortel MHV, Daney de Marcillac G (1988): An assessment of prediction methods for locating continuous epitopes in proteins. *Immunol Lett* 17(2):95–107.

Van Regenmortel MHV (1989): Structural and functional approaches to the study of protein antigenicity. *Immunol Today* 10(8):266–272.

Van Regenmortel MHV (1996): Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. *Methods* 9(3):465–472.

Van Regenmortel MHV (2006): Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines. *J Mol Recognit* 19(3):183–187.

Van Regenmortel MHV (2007): The rational design of biological complexity: a deceptive metaphor. *Proteomics* 7(6):965–975.

Villen J, de Oliveira E, Nunez JI, Molina N, Sobrino F F Andreu D (2004): Towards a multi-site synthetic vaccine to foot-and-mouth disease: addition of discontinuous site peptide mimic increases the neutralization response in immunized animals. *Vaccine* 22(27–28):3523–3529.

Walter G (1986): Production and use of antibodies against synthetic peptides. *J Immunol Methods* 88(2):149–161.

Welling GW, Weijer WJ, van der Zee R, Welling-Wester S (1985): Prediction of sequential antigenic regions in proteins. *FEBS Lett* 188(2):215–218.

Westhof E, Altschuh D, Moras D, Bloomer AC, Mondragon A, Klug A, Van Regenmortel MH (1984): Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature* 311(5982):123–126.

Zolla-Pazner S (2004): Identifying epitopes of HIV-1 that induce protective antibodies. *Nat Rev Immunol* 4(3):199–210.

Author Query

1. Please include the following references in the reference list or else delete their citation from the text: Kendrew et al., 1958; Anderer and Schlumberger, 1965; Poljak et al., 1973; Hopp, 1981; Stanfield et al., 1990; Schonbach et al., 2000.
2. Please check whether the meaning of the sentence "A variety of . . . regions of a certain protein." is retained after the edits.

Serodiagnosis of Chronic Chagas Infection by Using EIE-Recombinant-Chagas-Biomanguinhos Kit

Yara M Gomes/+, Valéria RA Pereira, Mineo Nakazawa, Daniela S Rosa, Maria das Neves DS Barros*, Antonio GP Ferreira, Edimilson D Silva**, Sueli F Yamada Ogatta***, Marco Aurélio Krieger****, Samuel Goldenberg****

Departamento de Imunologia, Centro de Pesquisas Aggeu Magalhães-Fiocruz, Av. Moraes Rego s/nº, Cidade Universitária, 50670-420 Recife, PE, Brasil *Ambulatório de Doença de Chagas, Hospital Universitário Oswaldo Cruz-UPE, Recife, PE, Brasil **Laboratório de Reativos do Instituto de Tecnologia em Imunobiológicos, Bio-Manguinhos-Fiocruz, Rio de Janeiro, RJ, Brasil ***Departamento de Microbiologia, Universidade Estadual de Londrina, Londrina, PR, Brasil ****Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz-Fiocruz, Rio de Janeiro, RJ, Brasil

A kit based on an enzyme immunoassay, EIE-Recombinant-Chagas-Biomanguinhos, developed by the Oswaldo Cruz Foundation, was evaluated for the serodiagnosis of chronic Chagas disease. Evaluation was performed with 368 serum samples collected from individuals living in an endemic area for Chagas disease: 131 patients in the chronic phase with confirmed clinical, epidemiological, and serological diagnosis (indirect immunofluorescence, indirect hemagglutination or enzyme-linked immunosorbent assay) and 237 nonchagasic seronegative individuals were considered negative control. The EIE-Recombinant-Chagas-Biomanguinhos kit showed high sensitivity, 100% (CI 95%: 96.4-100%) and high specificity, 100% (CI 95%: 98-100%). The data obtained were in full agreement with clinical and conventional serology data. In addition, no cross-reaction was observed with sera from patients with cutaneous (n=14) and visceral (n=3) leishmaniasis. However, when these sera were tested by conventional serological assays for Chagas disease, cross-reactions were detected in 14.3% and 33.3% of the patients with cutaneous and visceral leishmaniasis, respectively. No cross-reactions were observed when sera from nonchagasic seronegative patients bearing other infectious disease (syphilis, n=8; HTLV, n=8; HCV, n=7 and HBV, n=12) were tested. In addition, sera of patients with inconclusive results for Chagas disease by conventional serology showed results in agreement with clinical evaluation, when tested by the kit. These results are relevant and indicate that the refered kit provides a safe immunodiagnosis of Chagas disease and could be used in blood bank screening.

Key words: serodiagnosis - recombinant antigens - *Trypanosoma cruzi* - Chagas disease

Chagas disease is still a major health problem in Latin America where 16-18 million individuals are infected with the causative agent *Trypanosoma cruzi* and at least 90 million people are estimated to be at risk of infection (WHO 1996). Under natural conditions, infected reduviid bugs transmit the *T. cruzi* to humans when broken skin or mucous membranes contact metacyclics trypomastigotes from insect excreta. However, *T. cruzi* may bypass the vector bugs and be transmitted to man by a number of alternative mechanisms: blood transfusion, congenital transmission, accidental laboratory contamination, organ transplantation from infected donors and transmission by oral route (Umezawa et al. 1996, Gomes 1997). Blood transfusion is the second most common means of infection and the human migration from endemic areas to urban centers is proving a rising risk of transfusional Chagas disease in all Latin America and in non endemic countries (Schmuñis 1991).

Diagnosis of chronic Chagas disease is based on the detection of parasite by indirect parasitological methods (xenodiagnosis and hemoculture) or more usually on the detection of IgG antibodies against *T. cruzi* in the sera of patients by immunological methods (complement fixation-CF, indirect immunofluorescence-IIF, direct agglutination-DA, indirect hemagglutination-IHA and enzyme-linked immunosorbent assay-ELISA). The methods based on detection of the parasite, although highly specific, are of limited sensitivity, because parasites are

This work was partially supported by a PADCT-Finep to MAK. SG is recipient of a research fellowship from CNPq.
+Corresponding author. Fax: +55-81-453.2449. E-mail: yara@cpqam.fiocruz.br
Received 16 June 2000
Accepted 20 December 2000

detected in only 20-50% of individuals known to be infected, resulting in many false negative results (Gomes 1997). On the other hand, the methods based on the detection of an immune response to the parasite in the mamalian host lack specificity since they use crude or partially purified parasite extracts. Cross-reaction to *T. cruzi* has been observed with related protozoan diseases, particularly leishmaniasis. The problems with conventional assays (CF, IHA, DA, IIF and ELISA) may be overcome by using recombinant polypeptides containing specific *T. cruzi* epitopes that elicit an immune response in the majority of chagasic patients.

Several *T. cruzi* genes have been cloned and some of recombinant antigens have been assayed for their use in diagnosis (Affranchino et al. 1989, Levin et al. 1989, Almeida et al. 1990, Paranhos et al. 1990, Cotrin et al. 1990, Zingales et al. 1990, Goldenberg et al. 1991, Umezawa et al. 1999). Two recombinant antigens, CRA and FRA, expressed in the bacterium *Escherichia coli* were analyzed by Krieger et al. (1992) in a diagnostic test for Chagas disease. The data indicated that recombinant antigens displayed better results when used in combination than separately. These authors developed a direct ELISA which involves the use of peroxidase-labeled antigens to detect the immune-complexes. The results indicate that the recombinant (CRA+FRA) ELISA was better than the conventional ELISA in the diagnosis of Chagas disease, providing 100% specificity and sensitivity in all sera tested (Almeida et al. 1990, Krieger et al. 1992). These antigens were characterized and shown to display a repetitive epitope structure (Lafaille et al. 1989, Krieger et al. 1990). FRA (flagellar repetitive antigen) is located in the flagellum of the parasite and displays a 68-amino acid repeat, while CRA (cytoplasmic repetitive antigen) is distributed throughout the cytoplasm and has a 14-amino acid repeat (Lafaille et al. 1989).

Recently, a kit for diagnosis of chronic Chagas disease, using CRA+FRA antigens was developed by Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil. The kit is based on enzyme immunoassay, the direct ELISA.

In the present work we report the evaluation of the EIE-Recombinant-Chagas-Biomanguinhos kit for the diagnosis of *T. cruzi* infection using characterized serum samples from individuals living in Chagas disease endemic areas and from individuals with other infectious diseases.

MATERIALS AND METHODS

Human sera - Serum samples were collected from 368 patients between the ages of 5 and 76 years old from Hospital Universitário Oswaldo Cruz (Recife) living in Chagas disease endemic areas in the State of Pernambuco, Brazil: 131 patients in the chronic phase of the Chagas disease with confirmed clinical, epidemiological, and serological diagnosis and 237 nonchagasic individuals with negative serology were considered negative controls. Serum samples were previously classified as negative when two serological tests (ELISA, IIF or HAI) gave nonreactive results against *T. cruzi* antigens and as positive when two tests were reactive. Serum samples of 14 patients with cutaneous leishmaniasis (CL) and 3 with visceral leishmaniasis (VL) were also tested. Diagnosis of CL was based upon the collective analysis of a set of elements: presence of typical lesions, compatible epidemiological history and direct parasite detection. VL was diagnosed by clinical, epidemiological history, positive serological assay (IIF) and by detection of parasites in bone marrow aspirate. These sera were also tested by conventional serological assay to Chagas disease (ELISA and HAI) to evaluate cross-reaction.

Sera from patients with other infectious disease (8 with syphilis, 8 with HTLV, 7 with HCV, and 12 with HBV) as well as sera of patients with inconclusive results for Chagas disease (Table II) from blood center Fundação Hemope/Hemocentro, Pernambuco, Brazil were included in this study. All sera were analyzed by EIE-Recombinant-Chagas-Biomanguinhos kit. Blood samples from the individuals were taken by venopuncture and the sera obtained were stored at -20ºC until use.

Enzyme immunoassay - The enzyme immunoassay, with the EIE-Recombinant-Chagas-Biomanguinhos, was performed according to manufacturer. Briefly, microplates sensitized with the recombinant antigens were incubated with undiluted patient sera (50 µl) at 37ºC for 30 min. After washes to remove the unbound antibodies, the plates were incubated for 30 min at 37ºC with 50 µl of peroxidase conjugated antigens. After repeated cycles of washes the immune complexes were revealed by the addition of hydrogen peroxide and 3, 3', 5, 5'- tetramethylbenzidine. The reaction was stopped with 2 M H_2SO_4, and the optical density (OD) at 450 nm was determined in a ELISA reader (Bio-Rad 3550). The cutoff (CO) values as well as the gray zone were calculated for each plate according to manufacturer. Sera with OD values equal or greater than CO value were considered reactive, and consequently considered positive for antibodies to *T. cruzi*. Sera with OD values below CO were considered non-reactive and negative for antibodies to *T. cruzi*.

Data analysis - The figure of samples recorded as OD_{450nm} was distributed by using computer graphics software. The values of sensitivity and specificity were calculated according to Camargo (1992). The confidence interval (CI) was calculated at the level of 95%.

RESULTS

ELISA results are shown in the Figure. The CO and the gray zone values for each plate are shown in Table I. The high sensitivity of 100% (CI 95%: 96.4-100%) and specificity of 100% (CI 95%: 98-100%) show the excellent performance of the EIE-Recombinant-Chagas-Biomanguinhos kit. All 131 cases of confirmed Chagas disease, which had been diagnosed by clinical and conventional serology were positive and all 237 sera from nonchagasic individuals were negative.

Analysis of the sera negative for Chagas disease but positive for CL and VL shows that the responses do not give rise to false positive results (Figure). When these sera were tested by conventional serological assays for Chagas disease, cross-reactions were observed in 14.3% (2/14) of the patients with CL and 33.3% (1/3) patients with VL. No cross-reactions were observed when sera from patients with syphilis, HTLV, HBV and HCV were tested by the recombinant kit. The sera with inconclusive results for Chagas disease by using conventional serological assays were in agreement with the clinical evaluation (two chagasic patients and four nonchagasic individuals) when tested by recombinant kit (Table II).

TABLE I

Cutoff and gray zone values obtained for each plate in present report

Plates	Cutoff values (OD)	Gray zone values (OD)
1	0.233	0.233-0.279
2	0.244	0.244-0.293
3	0.198	0.198-0.237
4	0.192	0.192-0.231
5	0.178	0.178-0.213
6	0.183	0.183-0.220

TABLE II

Comparative evaluation of the sera with the inconclusive results by conventional serological tests and EIE-Recombinant-Chagas-Biomanguinhos kit

Sera	Conventional serological tests		EIE-Recombinant-kit
	ELISA OD/CO	IIF	OD/CO
1[a]	0.289/ 0.180	NR	0.087/0.198
2[a]	0.245/0.180	NR	0.087/0.198
3[a]	0.286/0.180	NR	0.088/0.198
4[a]	0.254/0.180	NR	0.096/0.198
6[b]	1.188/0.180	NR	1.242/0.198
7[b]	0.215/0.180	R	1.150/0.198

a: nonchagasic sera; *b*: chagasic sera; NR: non reactive sera; R: reactive sera; OD: optical densities; CO: cutoff.

DISCUSSION

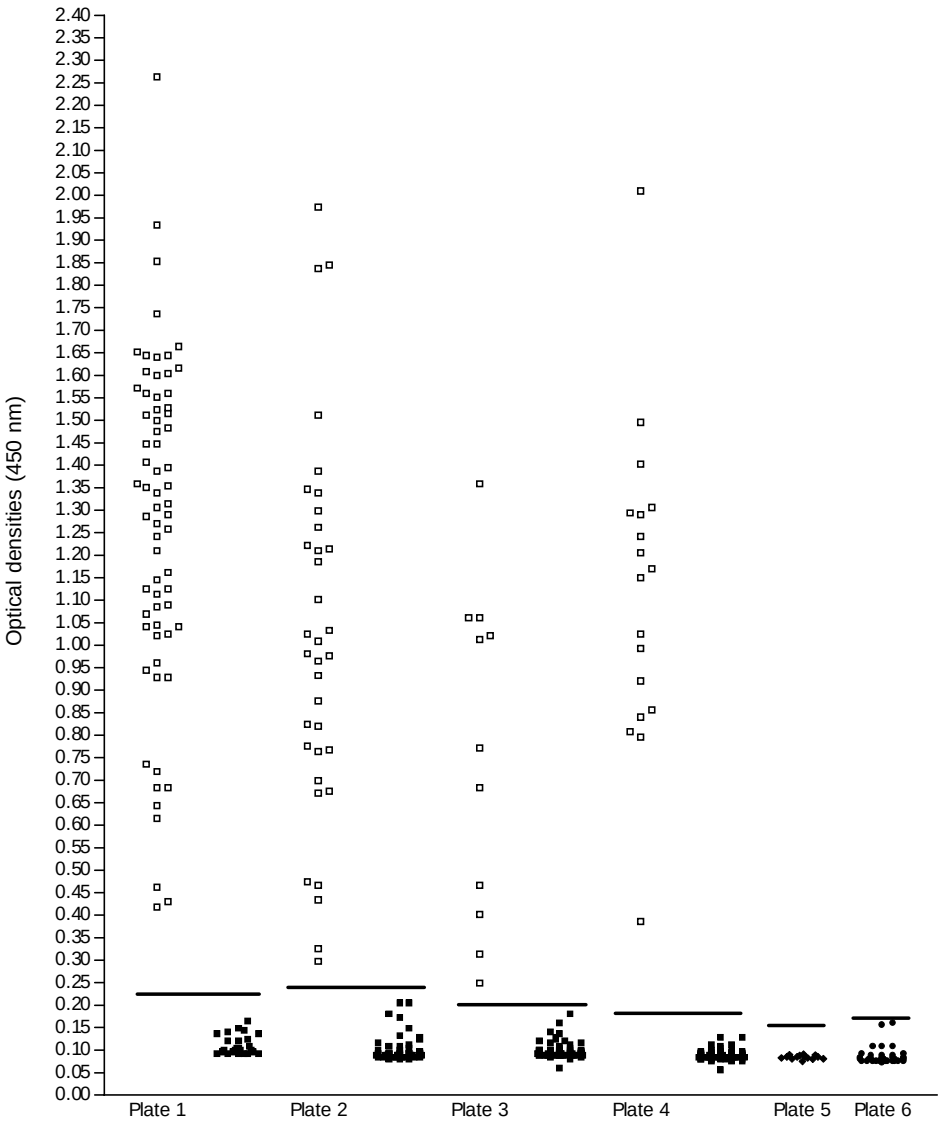
The serological conventional assays (CF, IIF, IHA and ELISA) for diagnosis of Chagas disease are used for individual diagnosis and for screening of donated blood, as well as in epidemiological studies. However, a persistent problem with these conventional assays has been the occurrence of false-positive results. Due to this problem, the World Health Organization recommends that serum specimens should be tested in two conventional assays before being accepted as positive (Gomes 1997). This approach carries with it an enormous logistic and economic burden, especially for a blood bank. Even if some blood banks use three different serological tests to reach a diagnosis, the amount of blood discarded that could be transfused, if there was a safe test, is significant. As an example, in the major Brazilian blood center (Fundação Pró-Sangue Hemocentro de São Paulo) some 10,000 blood units are discarded per year due their reactivity for Chagas disease in conventional serologic tests, at a cost of approximately U$ 60.00 per bag (Carvalho et al. 1993). In addition, the diagnostic tests that give some false-positive results should be avoided because such tests can create social problems for false-positive chagasic patients. They can also produce erroneous data in epidemiological studies (Krieger et al. 1992).

In order to overcome these problems, a new diagnostic kit EIE-Recombinant-Chagas-Biomanguinhos was developed by Fiocruz to detect antibody to *T. cruzi* in sera and plasma. The test is a direct ELISA that use the CRA+FRA recombinant antigens.

In the present report we have evaluated the EIE-Recombinant-Chagas-Biomanguinhos kit using serum samples from individuals living in Chagas disease endemic areas of State of Pernambuco, Brazil. The test showed to be highly sensitive and specific, detecting 100% of the chagasic (131/131) and nonchagasic individuals from endemic areas (237/237), respectively. Sensitivity and specificity were 100% (CI 95%: 96.4-100%) and 100% (CI 95%: 98-100%), respectively. In addition, no cross-reaction was observed with sera from patients with CL and VL. Even if low reactivity sera were not included in this study, we may predict a high sensitivity according to the results obtained with these confirmed chagasic patients. Several conventional serological tests as well as ELISAs commercially available in Brazil showed cross-reactivity to sera from patients with other diseases. These false-positive results are frequent in the case of patients with leishmaniasis (Carvalho et al. 1993). The absence of cross-reactions in EIE-Recombi-

nant-Chagas-Biomanguinhos shows its high specificity for Chagas disease.

According to Camargo (1992), a problem that occurs during screening of chagasic sera in blood banks is the variable percentage of samples showing a reactivity in the gray zone around the cutoff values. In order to verify this problem we tested by recombinant kit, sera from blood center Fundação Hemope/Hemocentro, Pernambuco, Brazil with inconclusive results for Chagas disease obtained by conventional serological assays. Reactivity in the gray zone was not observed and the results were in full agreement with clinical evaluation. The kit EIE-Recombinant-Chagas-Biomanguinhos for the diagnosis of chronic Chagas disease has several advantages over other available methods: (i) the use of specific *T. cruzi* recombinant antigens avoids false-positive reaction; (ii) the direct ELISA increases the sensitivity of the method allowing the evaluation of low titer sera, and corroborates its specificity; (iii) the use of undiluted serum samples reduces the possibility of error due to manipulation; (iv) the procedure is quick (taking 2 h to perform), easily performed and reliable. Taken as a whole, these facts indicate that the EIE-Recombinant-Chagas-Biomanguinhos is



Distribution of optical densities values from chagasic and nonchagasic individuals using the EIE-Recombinant-Chagas-Biomanguinhos kit. The horizontal line inside the drops for each plate represents the cutoff values. Chagasic individuals (□); nonchagasic individuals (■); individuals with visceral and cutaneous leishmaniasis (◆); syphilis, HTLV, HCV and HBV infections (●)

suitable for the diagnosis of Chagas disease and could be used in blood bank screening. We are currently evaluating this kit for screening donated blood of the Fundação Hemope.

ACKNOWLEDGEMENTS

To Edileuza Brito for providing sera of leishmaniasis and Wayner Souza for performing the statistical analysis. To Genilda Medeiros and Ana Cristina B Souza of the blood center Fundação Hemope/Hemocentro, Pernambuco for performing several serological assays for Chagas disease and providing sera with other infectious diseases.

REFERENCES

Affranchino JL, Ibanez CF, Luquetti AO, Rassi A, Ryes MB, Macina RA, Aslund L, Petterson U, Frasch ACC 1989. Identification of a *Trypanosoma cruzi* antigen that is shed during the acute phase of Chagas' disease. *Mol Biochem Parasitol 34*: 221-228.

Almeida E, Krieger MA, Carvalho MR, Oelemann W, Goldenberg S 1990. Use of recombinant antigens for the diagnosis of Chagas' disease and blood bank screening. *Mem Inst Oswaldo Cruz 85*: 513-517.

Camargo ME 1992. An appraisal of Chagas' disease serodiagnosis. In S Wendel, Z Brener, ME Camargo, A Rassi (eds), *Chagas' Disease (American Trypanosomiasis): its Impact on Transfusion and Clinical Medicine*, ISBT Brazil'92, São Paulo, p. 165-168.

Carvalho MR, Krieger MA, Almeida E, Oelemann W, Shikanai-Yassuda MA, Ferreira AW, Pereira JB, Sáez-Alquézar A, Dorlhiac-Llacer PE, Chamone DF, Goldenberg S 1993. Chagas' disease diagnosis: evaluation of several tests in blood bank screening. *Transfusion 33*: 830-834.

Cotrim PC, Paranhos GS, Mortara RA, Wanderley J, Rassi A, Camargo ME, Franco da Silveira J 1990. Expression in *Escherichia coli* of a dominant immunogen of *Trypanosoma cruzi* recognized by human chagasic sera. *J Clin Microbiol 28*: 519-524.

Goldenberg S, Krieger MA, Lafaille JJ, Almeida E, Oelemann W 1991. Use of *Trypanosoma cruzi* antigens in the immunological diagnosis of Chagas' disease. *Mem Inst Butantan 53*: 71-76.

Gomes YM 1997. PCR and sero-diagnosis in chronic Chagas'disease: biotechnological advances. *Appl Biochem Biotechnol 66*: 107-119.

Krieger MA, Salles JM, Almeida E, Linss J, Bonaldo MC, Goldenberg S 1990. Expression and polymorphism of a *Trypanosoma cruzi* gene encoding a cytoplasmic repetitive antigen. *Exp Parasitol 70*: 247-254.

Krieger MA, Almeida E, Oelemann W, Lafaille JJ, Pereira JB, Carvalho MR, Goldenberg S 1992. Use of recombinant antigens for the accurate immunodiagnosis of Chagas' disease. *Am J Trop Med Hyg 46*: 427-434.

Lafaille JJ, Linss J, Krieger MA, Souto-Padron T, de Souza W, Goldenberg S 1989. Structure and expression of two *Trypanosoma cruzi* genes encoding antigenic proteins bearing repetitive epitopes. *Mol Biochem Parasitol 35*: 127-136.

Levin M, Mesri E, Benarous R, Levitus G, Schijman A, Leyati P, Chiale P, Ruiz AM, Khan A, Rosenbaum M, Torres HN, Segura EL 1989. Identification of major *Trypanosoma cruzi* antigenic determinants in chronic Chagas' disease. *Am J Trop Med Hyg 41*: 530-538.

Paranhos GS, Cotrin PC, Mortara RA, Rassi A, Corral R, Freilij HL, Grinstein S, Wanderley J, Camargo ME, Franco da Silveira J 1990. *Trypanosoma cruzi*: cloning and expression of an antigen recognized by acute and chronic human chagasic sera. *Exp Parasitol 71*: 284-293.

Schmuñis GA 1991. *Trypanosoma cruzi*, the etiologic agent of Chagas' disease: status in the blood supply in endemic and nonendemic countries. *Transfusion 31*: 547-557.

Umezawa ES, Nascimento MS, Kersper Jr N, Coura JR, Borges-Pereira J, Junqueira ACV, Camargo ME 1996. Immunoblot assay using excreted-secreted antigens of *Trypanosoma cruzi* in serodiagnosis of congenital, acute and chronic Chagas'disease. *J Clin Microbiol 34*: 2143-2147.

Umezawa ES, Bastos S, Camargo ME, Yamauchi LM, Santos MR, Gonzalez A, Zingales B, Levin MJ, Sousa O, Rangel-Aldao R, da Silveira JF 1999. Evaluation of recombinant antigens for serodiagnosis of Chagas' disease in South and Central America. *J Clin Microbiol 37*: 1554-1560.

WHO-World Health Organization 1996. *Control of Tropical Disease. Chagas' Disease. A Disease whose Days are Numbered*, WHO, Geneve, 16 pp.

Zingales B, Gruber A, Ramalho CB, Umezawa ES, Colli W 1990. Use of recombinant proteins of *Trypanosoma cruzi* in the serological diagnosis of Chagas' disease. *Mem Inst Osvaldo Cruz 85*: 519-522.