

SCIENTIFIC REPORTS



OPEN

Evolving serodiagnostics by rationally designed peptide arrays: the *Burkholderia* paradigm in Cystic Fibrosis

Claudio Peri¹, Alessandro Gori¹, Paola Gagni¹, Laura Sola¹, Daniela Girelli², Samantha Sottotetti², Lisa Cariani², Marcella Chiari¹, Marina Cretich¹ & Giorgio Colombo¹

Efficient diagnosis of emerging and novel bacterial infections is fundamental to guide decisions on therapeutic treatments. Here, we engineered a novel rational strategy to design peptide microarray platforms, which combines structural and genomic analyses to predict the binding interfaces between diverse protein antigens and antibodies against *Burkholderia cepacia complex* infections present in the sera of Cystic Fibrosis (CF) patients. The predicted binding interfaces on the antigens are synthesized in the form of isolated peptides and chemically optimized for controlled orientation on the surface. Our platform displays multiple *Burkholderia*-related epitopes and is shown to diagnose infected individuals even in presence of superinfections caused by other prevalent CF pathogens, with limited cost and time requirements. Moreover, our data point out that the specific patterns determined by combined probe responses might provide a characterization of *Burkholderia* infections even at the subtype level (genomovars). The method is general and immediately applicable to other bacteria.

Bacterial infections and epidemics present a continuous threat to mankind through their impact on morbidity and mortality combined to the steady increase of drug-resistance, all factors that make efficient medical intervention strategies and infrastructures urgent necessities. The integration of new research approaches with prevention, care, treatment, and surveillance, can aptly be combined to define the best therapeutic options for affected populations. In this context, the development of quick and effective diagnostic methods is a key factor in patient management, especially for novel, aggressive, or drug-resistant infections that must be tackled in short timeframes. The quest for improved screening and diagnostic approaches includes the development of advanced platforms with the ability to detect and discriminate among the different types of antibodies (IgM, IgG, IgA) from biological fluids, providing fundamental information on infection type and status. State-of-the-art methods entail enzyme immune assays (EIA), commonly employed in the form of ELISA tests, which use antigens (e.g. proteins and lipopolysaccharides) adsorbed on a rigid support as baits to capture antibodies. ELISA is reasonably quick but limited in the number of probes and the use of complete antigens (such as recombinant proteins or complex saccharides) sets severe limitations in terms of cost and versatility^{1,2}.

The transition to microarray technologies would allow to increase the number of probes and make the tests high-throughput, with the ability to screen simultaneously a large number of molecular probes from the same or multiple pathogens^{3,4}. Despite the technological advancements represented by protein microarrays, their cost still undermines widespread application in diagnostics as the production of individual recombinant proteins remains a limiting factor.

In this framework, peptide microarrays represent a viable solution to overcome such limitations. Recent initiatives explored the potential of peptide microarrays using libraries of linear peptides spotted on a single chip and interrogated for antigenicity⁵⁻⁷. These methods, aimed mainly at epitope mapping, maintain the advantage of a high-throughput test while improving in terms of simplicity and manageability with respect to the use of full-length antigens. In principle, such methodology could be further improved by exploiting highly specific

¹Istituto di Chimica del Riconoscimento Molecolare, ICRM, CNR. Via Mario Bianco 9, 20131, Milano (Italy). ²Cystic Fibrosis Microbiology Laboratory, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, via San Barnaba 8, 20122, Milano (Italy). Correspondence and requests for materials should be addressed to Marina C. (email: m.cretich@icrm.cnr.it) or G.C. (email: g.colombo@icrm.cnr.it)

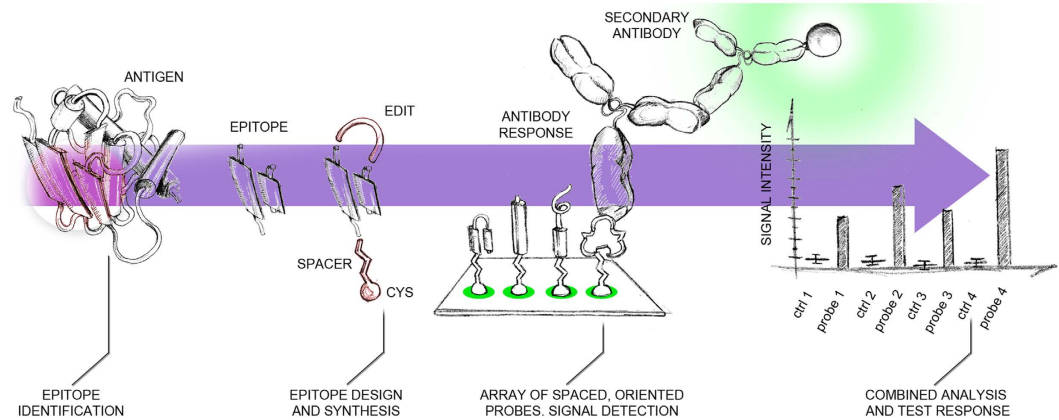


Figure 1. Pictorial scheme of the method's workflow. Starting from an antigen's 3D structure, epitope identification is carried out by MLCE computational predictions. The epitopes are modified at the structural level in order to be synthesized as peptides and spotted as spaced and oriented baits. The platform is interrogated with patients' antisera and detection occurs by means of secondary antibodies. The signal intensity is subsequently acquired and analyzed for a test response.

baits, designed as peptide-based mimics that recapitulate the fundamental molecular determinants of antigen recognition^{8–11}. In this context, the rational identification of substructures on selected protein antigens (epitopes) can be translated into the synthesis of easy-to-manage, small sets of *ad hoc* designed peptidic baits displayed on microarrays. The main advantage in this kind of approach stems from the ability to facilitate the analysis and interpretation of results compared to large peptide libraries while exploiting the combination of responses originated from probe redundancy. The use of diverse sets of molecular probes can in fact improve on the reliability of the tests reducing the risk of noise due to unexpected cross-reactivity. Furthermore, the analysis of combined signal patterns opens new perspectives for serologic diagnosis: the presence and the combined response of antibodies directed against their specific synthetic epitopes could report on the status of the infection, identify patient subgroups, discriminate different pathogens (and their variants) and assist in medical decision making. Peptide-focused diagnostic design is a conceptually fascinating but still highly unexplored avenue.

Here, we present an original integration of the results of computational epitope design, peptide synthesis and optimal modification for probe display on microarrays with the aim to move the application of molecular diagnostics beyond its current limits: in the case of a highly invalidating rare disease such as Cystic Fibrosis (CF), we prove the possibility to generate highly efficient, selective peptide-based microarray diagnostic tools that can be predictive of the infection state as well as show potential for the definition of the genomic variant of the pathogen. Specifically, we describe a novel platform for the screening of *Burkholderia cepacia* complex (BCC) infections in subjects affected by Cystic Fibrosis (CF).

Chronic and recurrent bacterial infections often characterize lung conditions in CF. Even if advances in antibiotic therapy contribute enormously to increase survival in CF, the growth of bacterial biofilms in CF airways can make their eradication extremely difficult – even aggressive antibiotic treatments can be ineffective when facing multidrug-resistant organisms. Routine laboratory techniques can identify only a small fraction of the microbes present in the CF airway¹² and currently there is no readily available methodology to discriminate among these organisms to guide clinical therapeutic decisions¹³. In this scenario, the most prevalent pathogen *P. aeruginosa*, and others such as *Achromobacter xylosoxidans* and the *Burkholderia cepacia* complex often resist treatment, leading to respiratory failure, with lung transplantation being the only therapeutic option¹⁴. *Burkholderia cepacia* complex (BCC) bacteria have gained notoriety in CF due to the difficulties they pose in diagnosis and treatment. They are associated to a poor prognosis and can also easily spread amongst CF individuals. This group of closely related *Burkholderia* species can be further classified into 17 different genomovars¹⁵. Rapid and accurate diagnosis is the key to select the best therapeutic strategy and to control such infections. However, BCC diagnosis is currently based on bacterial culture isolation and sequencing, which is a cumbersome and time consuming process requiring specialized laboratories¹⁶ while ELISA tests for antibody detection are generally unstandardized and so far not widely used in clinics due to cross-reactivity issues.

The selective detection of BCC species in CF patients, even in the face of other common respiratory superinfections, represent thus an optimal benchmark for our rationally-designed peptide microarray platform to verify its ability at providing a test that is fast, effective, reliable and specific. A visual overview of the methodology employed is summarized in Fig. 1.

Very briefly, the platform proved to be highly predictive, with seven probes discriminating correctly the BCC positive samples with confidence exceeding 97%, and differences in signal distributions relative to the control characterized by good statistical significance ($p < 0.001$).

Since it starts from the knowledge and re-design of conserved epitopes, our method is fully general and applicable to any pathogen for which structural information on antigenic proteins is available and it may be immediately extended to the screening of multiple infections on a single chip.

ANTIGEN/PDB CODE	EPITOPE NAME	PROBE SEQUENCE
BPSL2765/4B5C	PAL1	Cys-PEG-(144)GKEKPVALGHDEASWAQNRRLADLVYQ
BPSL2765/4B5C	PAL2	Cys-PEG-(132)GVGDAQMEAV-PEG-(89)YLKSHPQRHIL
BPSL2765/4B5C	PAL3	Cys-PEG-PEG-(72)SYSVQDQYQALLQQAQYLYK
BPSL2765/4B5C	PAL3H	Cys-PEG-PEG-(72)SYSVQDQYQALLQQAQYLYK-[Lys-N3]QAE[Prg]
BPSS2141/3ZS6	OPPA1	Cys-PEG-PEG-(106)KAPDTgg(184)KTEVPVSY
BPSS2141/3ZS6	OPPA2	Cys-PEG-PEG-(357)GVKGVQRFPDWA
BPSS2141/3ZS6	OPPA3	Cys-PEG-PEG-(488)AEANQKLDDGARAALLTQAHDLA
BPSS2141/3ZS6	OPPA4	Cys-PEG-PEG-(121)WSNGQPVTAAADFVYAWQR
BPSS2141/3ZS6	OPPA5	Cys-PEG-PEG-(299)ELRPGLQLATYYYYLYK
BPSL1050/2MPE	2MPE1	Cys-(O2OC)2-(38)VGYGGHGHPTQVRIVAPHAHVRYAH
BPSL1050/2MPE	2MPE2	(84)DGAARFERYLAALPRKLAAWENARGVDFGSRTQADAL-(O2OC)2-Cys
BPSL0919/3F7T*	ISPH1	Cys-(O2OC)2-(304)ALEGIEENVSFPLPRGL

Table 1. Epitopes and sequences. At the heart of each probe is a synthetic peptide corresponding to an immunoreactive epitope from *B. pseudomallei* antigens. For each epitope is shown its corresponding antigen in form of its UniProt ID⁴⁰ and RCSB PDB⁴¹ code. The last column includes the amino acid sequence and the chemical modifications carried out for synthesis. All sequences are based on linear epitopes except for epitope PAL2 and OPPA1, which are based on conformational antibody binding sites. *BPSL0919 does not feature a structural record in the PDB, thus the antigen was modelled based on homology with PDB ID 3F7T, corresponding to *E. coli* IspH.

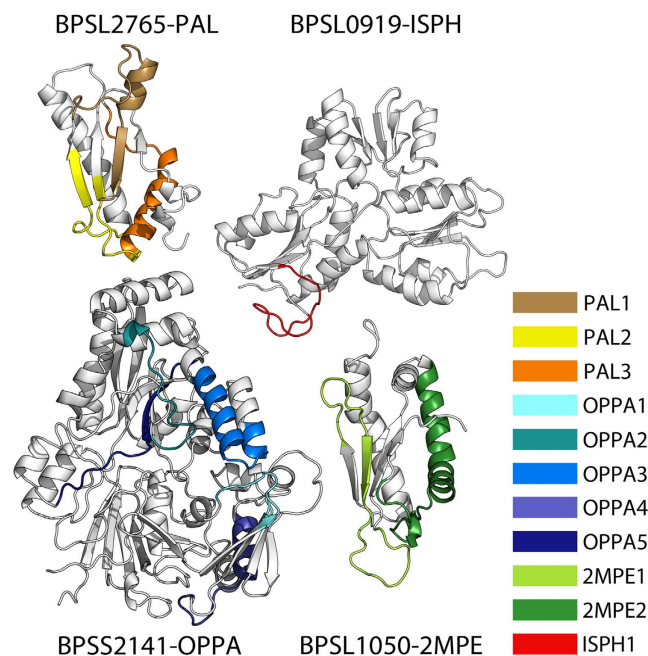


Figure 2. Visual representation of the epitopes in their native structural context. The figure shows all four antigens from *B. pseudomallei* chosen for this study and their corresponding epitopes highlighted in different colors. The PDB codes used for each antigen are 4B5C, 3ZS6 and 2MPE for proteins BPSL2765, BPSS2141 and BPSL1050 respectively. The structure of BPSL0919 was reconstructed by homology modeling from template 3F7T.

Results

Selection of the probes. Each of the selected peptidic probes, corresponding to either a linear or conformational immunoreactive epitope, was rationally designed on the basis of the three-dimensional structure of cognate antigenic protein, as described in previous papers^{8–11}. The list of antigens and corresponding epitopes is described in Table 1 and shown as Fig. 2.

Specifically, to establish the suitability of designed epitopes for BCC applications in CF, we performed a sequence conservation analysis of the selected epitopes against six genomovars of the *cepacia* complex (*B. cepacia*, *B. multivorans*, *B. cenocepacia*, *B. vietnamiensis*, *B. dolosa*, *B. ambifaria*) using the EnsemblGenomes resources (<http://ensemblgenomes.org>)¹⁷. The analysis also included other relevant *Burkholderia* species (*B. mallei*, pathogenic, and *B. thailandensis*, usually non-pathogenic) and common, but unrelated, pathogenic microbes affecting

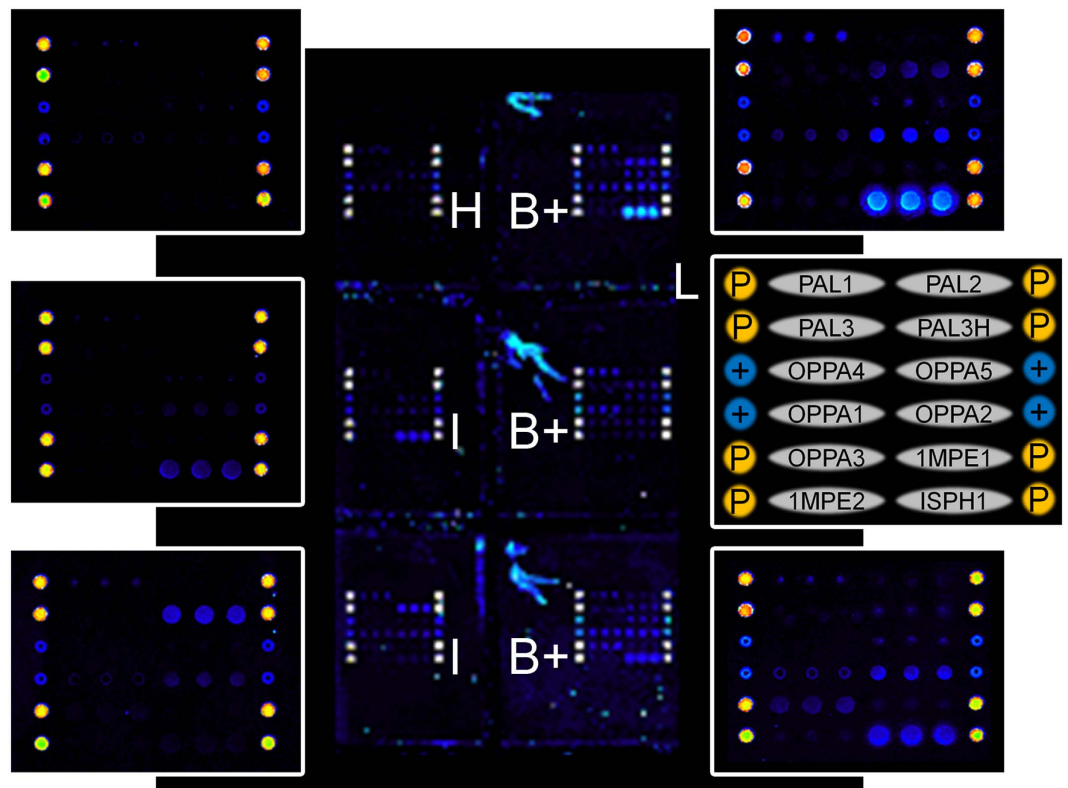


Figure 3. Representative fluorescence results of the peptide microarray. Each subarray is composed of the 12 peptides spotted in triplicates, 8 positional control spots (P) and 4 positive control spots (+), as it is shown in the legend (L). Each slide is spotted with 16 subarrays and incubated with 16 samples using a multiwell silicon chamber. At the centre, a portion of the slide displaying 6 adjacent subarrays that were incubated with a healthy control (H), two BCC negative controls (I) and three BCC positive patients (B+). Five representative subarrays are magnified.

CF patients (*P. aeruginosa*, *A. xylosoxidans*, *S. aureus*, *S. maltophilia*) and a fungus, *C. albicans* (Supplementary Figure S1). The results confirmed that the selected epitopes are well conserved and mostly identical among the different *Burkholderia* species, while poorly conserved if not totally absent in all other pathogens, with the sole exception of epitope ISPH1 from protein BPSL0919 (Supplementary Figure S1). The latter antigen appears to be homologue to other proteins from the same family across the *Burkholderia* genus, *P. aeruginosa* and *A. xylosoxidans*.

In previous reports, all epitopes except for ISPH1, which was chosen here as a control on sequence conservation, were individually reported to be immunoreactive and showed interesting potential in discriminating between the sera of donors who recovered from a *B. pseudomallei* acute infection and healthy negative controls (seronegative and seropositive individuals) in ELISA tests^{8–11}.

Microarray tests and diagnostic performance. A peptide microarray, composed of the epitopes reported in Table 1, was tested for its ability to diagnose *Burkholderia cepacia* complex (BCC) infection detecting peptide-specific IgG in the serum of infected CF patients. Since CF patients are often affected by multiple respiratory tract infections/colonization we also checked the cross-reactivity of the peptide microarray with a set of CF patients sera that were diagnosed as BCC negative while positive for other common CF pathogens such as *P. aeruginosa*, *A. xylosoxidans* and *S. maltophilia* (see details in the Supplementary Information, Table S1).

Epitopes reported in Table 1, terminated by a spacer and a cysteine residue, were arrayed on a silicon slide coated with a maleimido-functionalised polymer to promote the oriented immobilisation of peptides *via* specific thiol addition to maleimides on a modified copoly(DMA-NAS-MAPS)¹⁸ coated surface. The arrays were probed with 14 serum samples from BCC positive CF patients, diagnosed by microbiological culture and MALDI-TOF spectrometry and further characterized by sequencing for genomovar identification; 11 sera from CF individuals that were negative for BCC (although positive for other respiratory tract pathogens) and 14 healthy donors (see details on samples in the Methods section and in the Supplementary Table S1). The sera from CF patients used for this study are relative to all samples suitable for testing collected in a 1 year timespan at Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan. The details for microarray production are reported in the Methods section.

The peptide-specific IgG content in each serum was evaluated by fluorescence detection using an anti-human IgG labelled with the dye Cy3. A representative microarray is reported in Fig. 3.

Probe	AUC: Area Under the ROC Curve (B+): BCC positive patients (H): healthy controls	Unpaired t Test ns: not significant Significant: $p < 0.05$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$. (B+): BCC positive patients (H): healthy controls (I): BCC negative	
	AUC B+ vs H	B+ vs H	B+ vs I
PAL1	0.9505	**	ns
PAL2	1	**	*
PAL3	1	****	***
PAL3H	1	***	**
OPPA1	0.6786	ns	ns
OPPA2	1	***	**
OPPA3	0.8352	*	**
OPPA4	1	****	****
OPPA5	1	****	**
2MPE1	0.6071	ns	**
2MPE2	1	****	****
ISPH1	0.6684	ns	ns

Table 2. Evaluation of the diagnostic performance of the peptide array. Each probe is followed by a column describing the AUC measure of performance for the classification of donors as BCC positive or negative subjects. 7 probes out of 12 reach a full discrimination potential (AUC = 1). The Unpaired t Test was used as a measure of statistical significance. The test was carried out on the performance of discrimination against a pool of healthy controls (B+ vs H) and a pool of infected FC patients negative to BCC screening (B+ vs I).

Healthy controls (labelled H) typically provided negligible fluorescence signals (top left array), CF patients positive for BCC (right arrays, labelled B+) showed patterns of fluorescence signals (corresponding to sero-reactive peptides) that were significantly different from patterns detected in BCC negative patients (left arrays, labelled I). The ability of each epitope to distinguish between controls and patients was evaluated performing the unpaired t Test (significant if all p values < 0.05) on the peptide-specific fluorescence signals detected in healthy individuals, BCC positive patients (14 samples) and the control group of CF patients affected by pathogens other than BCC (11 samples). Accuracy of discriminating between BCC infection vs healthy controls was evaluated quantitatively for each probe by a Receiver Operating Characteristic (ROC) curve analysis.

A summary of the diagnostic performance for each peptide of the array is reported in Tables 2 and S2 in the Supplementary Information, while detailed histogram representations of the t Test analysis are shown in Supplementary Figures S2 and S3.

Nine probes were found to be significantly serodiagnostic, allowing to distinguish BCC positive individuals from healthy controls (all p-values < 0.05), seven of them providing 100% sensitivity and specificity (Area under the ROC curve = 1).

Extending the comparison to CF infected controls, eight probes are diagnostic for BCC and lack of cross-reactivity with the other common CF pathogens (Area under the ROC curve = 1 for 3 probes and > 0.92 for 5 probes, see details in Supplementary Information, Table S2) demonstrating the high specificity of the epitopes for BCC infection detection. The peptide chemoselective immobilization strategy here adopted resulted to be critical for the analytical outcomes and the test diagnostic accuracy.

Performance analysis. To provide a semi-quantitative evaluation of the performance of the microarray, a comprehensive cluster analysis of all data was carried out, collecting and comparing for each probe the fluorescence of BCC positive patients (B+) versus randomized controls. The analysis is presented in Fig. 4. Controls were randomized by assigning each spot (probe fluorescence value) to a donor at random for each control group (Healthy (H) or Infected (I)). This was made necessary by the fact that we did not know the identities of the donors due to privacy legal requirements and to the difficulties in collecting large sets of samples for CF infections. The fluorescence data were standardized for each probe, in order to minimize the statistical influence associated with extreme data points in the distribution (see Methods).

To sort the data in terms of similarity between patient profiles and probe performance, we clustered the standardized intensities grouping probes vs. patients using a hierarchical approach based on Manhattan distance and average linkage (see Methods). We expressed the relations between patients and probes, as well as the performance of the method with a clustered heatmap of signal intensity (Fig. 4). Concerning the probes (X-axis), it is evident that peptides ISPH1, OPPA1 and 2MPE1, and partially OPPA3 have poor or absent discrimination potential for B+ patients compared to controls as shown by the presence of positive values scattered across all serum samples. They are thus gathered in a separate cluster.

All other probes are closely related in their discrimination potential. In particular, probes OPPA4 and 2MPE2 optimally discriminate B+ cases vs. controls. PAL3H and PAL2 are considered as outgroups indicating a similar discrimination pattern and positive performance. All other probes from PAL and OPPA antigens are joined in the same cluster with similar response.

Considering the donors (Y-axis), the cluster analysis produces three main groups. The first one gathers the controls, without noticeable difference between healthy and infected donors, while the second one is composed of

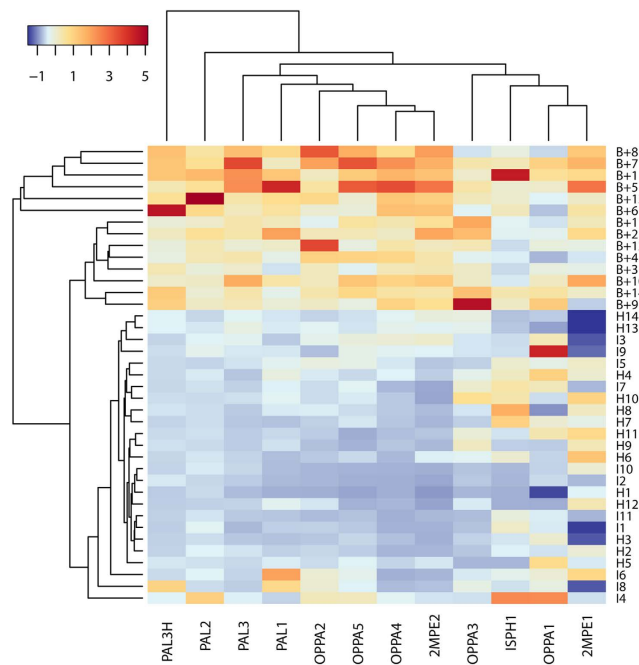


Figure 4. Heatmap of fluorescence intensity built after data clustering. All the data collected from the microarray platform test are reported as subdivided by probes (X axis) and donors (Y axis). Controls (H and I classes) were randomized before analysis and assigned an arbitrary donor number, while signal intensities for B+ patients are relative to specific donors. The color code follows an adimensional scale (due to the standardization of signals). Negative values are displayed in blue gradient and positive values are displayed in yellow/red gradient. The distances and relations between clusters of probes and patients are expressed with dendrograms on their corresponding axis.

patients B + 8, B + 7, B + 1, B + 5, B + 13, B + 6, having a very pronounced response as a common characteristic. The third group is composed of all other B+ patients, positively discriminated by the probes against the controls.

A second test restricted to the B+ population was designed to verify whether the probes are sensitive enough to discriminate between bacterial subtypes, separating the different BCC genomovars.

The results of this test are shown in Fig. 5. We kept the B+ patients and removed the unresponding probes as they eventually would not be part of a diagnostic platform. The discrimination of signals into genomovars is based on much more subtle patterns than the clustering of positives vs negatives. Our aim here is to test whether a trend can be defined from the microarray data. As a relevant caveat, it is important to note that a much wider dataset would be required for a full statistical analysis aimed at setting a reference for each specific genomic variant. Herein, our analysis is aimed at returning significant observable trends.

Using the same setup as for the previous analysis, the clustering process was successful in combining into separate groups the fluorescence from genomovar IV and genomovar I. There is only one patient infected with a genomovar V BCC, so that, expectedly, the clustering process is ineffective for this subtype. Genomovars II and III feature highest fluorescence intensity in their response and, according to the dendrogram, appear to be more related to each other than the rest. We tried different setups and strategies for the cluster analysis (Euclidean distance, different clustering linkage), to test the trend shown in Fig. 5 for robustness. Depending on the strategy, the cluster analysis is able to distinguish different genomovars with consistent results. For instance, with a single variation in the clustering procedure (agglomeration of clusters by WPGMC algorithm), it was furthermore possible to correctly group and sort patients for genomovars I, II and IV (data not shown). This preliminary analysis on a limited number of patients supports the feasibility of extending the statistical analysis to a wider population of patients, in order to test the possibility of improving the sensitivity of the test to even predict the bacterial subtype.

Discussion

The genomic revolution has had a deep impact on life sciences in general and on immunology in particular. The availability of entire bacterial genomes and the development of high-throughput computational and experimental analysis tools have made it possible to rapidly and efficiently identify pathogenic protein Antigens (Ags) that may represent reactive candidates for the development of vaccines or diagnostic probes.

In parallel, progress in protein structural knowledge has favored the growth of structure-based Ag design, pushing the expansion of Structural Vaccinology, a new driving force that is being exploited for vaccine and immunodiagnostic development.

In this work, we have developed and tested a novel approach for the generation of new, rapid, efficient and cost-effective immunodiagnostic tools that integrate structural computational biology with advanced microarray technologies to diagnose BCC infections from the analysis of serum samples from Cystic Fibrosis patients.

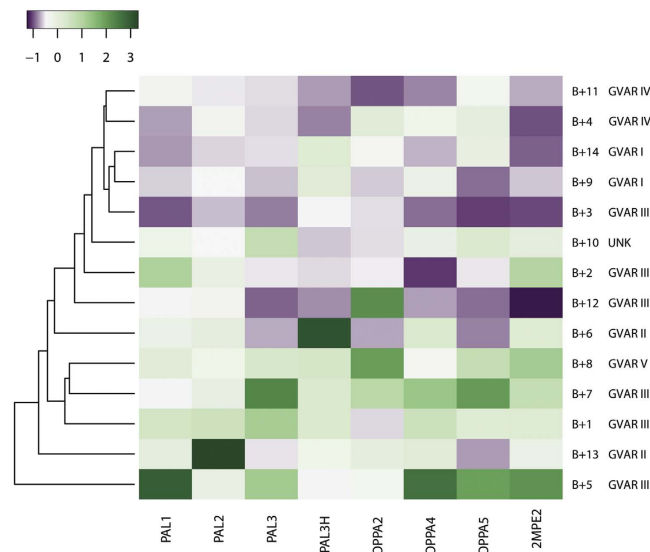


Figure 5. Cluster analysis of responding probes on B+ patients. In this heatmap the dataset was restricted solely to the B+ patients (Y axis) and the probes which demonstrated discriminating potential for BCC infections (X axis). The data was standardized and the color code follows an adimensional scale emphasizing the differences. The distances and relations between clusters of patients are expressed with dendrograms. Each donor is labelled with the corresponding genomovar (GVAR) of BCC infection. UNK is used to label the patient infected with BCC of unknown type.

Importantly, our method and tools showed a high degree of reliability and sensitivity, being able to discriminate *Burkholderia* infections even in the presence of several related respiratory superinfections, which is indeed a common, and largely unsolved, medical case in CF patients.

We designed our platform as an intrinsically multiplex approach, where different probes (corresponding to different epitopes from different conserved *Burkholderia* antigens) are simultaneously interrogated. In this prototype, we started from 12 different epitopes (probes) selected from 4 antigens that are well conserved among BCC members, at the same time being poorly conserved in the other CF co-infecting microorganisms considered. The epitopes were designed starting from the knowledge of the recently solved 3D structures of antigenic proteins from *B. pseudomallei*, which showed seroreactivity in preliminary ELISA tests^{8–10} (Fig. 2).

Summarizing, when integrated into our microarray platform, seven probes from antigens BPSL2765, BPSL2141 and BPSL1050 displayed a full discrimination potential for infected sera (AUC = 1, Table 2), and an eighth one (PAL1) showed interesting potential, supporting the potential of our diagnostic system for development and testing on a larger scale using larger sets of clinical samples.

In the case of seven probes, we obtained AUC values of 1 due to the fact that none of the healthy controls matched the fluorescence signal of affected individuals, indicating a high discrimination potential. A high score outcome is to be expected in small sample sets, such as in this case (39 subjects). However, the number of probes with a good discrimination potential is encouraging for future tests and clinical applications of this platform. Each AUC value with 95% CI is reported in the Supplementary Information (Table S2). It is worth noting here that such limitations are intimately connected to the fact that CF is a rare disease, and *Burkholderia* infections represent a subgroup of the CF affected population. Indeed, extensions of the studies to include more patient samples from different centers would be needed to fully establish the clinical relevance of our approach. Such steps are currently being organized.

The positive control on epitope conservation (epitope ISPH1 from antigen BPSL0919) returned as expected a cross-reactive response with controls, and only two other epitopes scored poorly with significant cross reactivity (Table 2, Fig. 4). The redundancy given by the use of many probes guarantees a high specificity for the test, producing a specific pattern given by the combined probe intensities (Fig. 4), which represents a “fingerprint” of the response.

Moreover, our platform demonstrated high diagnostic specificity and sensitivity even in the face of bacterial co-infections, for which we observe no cross-reactivity, providing high content results using just 1 microliter of serum, in a total analysis time of 3 hours.

While other methods based on PCR and cell culture techniques are available for CF opportunistic infections, our work is unique in that it directly advances the integration of genomic analysis, atomic-resolution computational epitope design, peptide synthesis and microarray technology, into a new and efficient tool for serological antibody screening in BCC infections that is reasonably cheap and usable in any laboratory. Moreover, further clustering analysis of the results defines trends that support the possibility to characterize BCC infection even at the level of bacterial genotypes (genomovar).

In fact, the various fluorescence signals from different probes define a distinct “fingerprint” for each patient. With a multiplex approach based on baits designed from a common cognate pathogen (*B. pseudomallei*), our aim

was to exploit the adaptability and specificity of the polyclonal response against their epitope¹⁹ to modulate a differential Ag-antibody response among the different bacterial variants. Thus, one additional use of this fingerprint is to detect and differentiate this specific response. Based on our results, we successfully sorted patients affected by genomovars I and IV with respect to the rest of the samples with a hierarchical cluster analysis based on average fingerprint similarity (Fig. 5).

We acknowledge that the sample set used in this study is inevitably too limited to allow a significant and quantitative correlation between a patient's fingerprint and a specific genomovar prediction. The limited sampling is strictly tied to the rare prevalence of FC in European population (1 in 13000 people or less)²⁰, meaning that further trials and testing of our technology will need cooperation with a panel of FC institutions. For this reason, at current state the clustering of genomovars reveals no more than an interesting trend, providing qualitative support to the feasibility of extending our serological procedure to larger clinical sera collections.

Many investigations support the use of serological diagnostics in CF to alert clinicians to seek for other evidence of infection in the event of negative cultures²¹, for improved understanding of the colonisation/infection process²², and indications on the necessity for antibiotic treatment to prevent chronic infection²³. However, only a few examples of serological diagnostic approaches applied to CF patients infected by *BCC* and other associated pathogens are present in the literature, due to the paucity of available recombinant antigens and poorly standardised methods. Moreover, the inherent complexity of the multiplex CF respiratory infections require several ELISA tests to be run for each plausible pathogen, making the process to monitor antibody response for the entire spectrum of CF's associated infections expensive and cumbersome. Conversely, the integration of additional diagnostic probes may transform this peptide array into an even more complete and versatile platform, while keeping (or even improving) advantages like high-throughput testing and cost effectiveness. One further advantage of the peptide array approach is that the detailed interrogation of immunity provided by the combined use of multiple, well characterized synthetic epitopes reduces the susceptibility of serology to cross-reactivity and false negatives offering the potential for superior test accuracy compared to full-antigen based assays which employ pathogen extracts, lipopolysaccharides or proteins used as single antibody baits.

In addition, peptides can be synthesized introducing non-natural functionalities such as linkers, or covalent bonding. Such strategies have a dual purpose: they assist the production of peptides mimicking more closely the native spatial conformation of epitopes, and they aid a chemoselective and oriented microarray immobilization on purposely designed microarray surfaces, favoring in turn a better and specific antibody recognition.

These technical hurdles and the improvements here proposed are not specific of *BCC* infection-screening alone. For instance, there are similar issues with the diagnosis of melioidosis and glanders, caused by *B. pseudomallei* and *B. mallei* respectively. In perspective, considering the conservation of epitope sequences across the *Burkholderia* genus, we anticipate a possible use of the peptide array presented here for the screening of multiple *Burkholderia* infections.

In the context of the diagnosis of melioidosis, which is characterized by very dangerous and acute episodes, one cannot just rely on the screening of IgG molecules. In our case, we focused the present investigation on the detection of IgGs only since the sera we analyzed were from patients known to be affected by chronic lung infection. In addition, the relevance of IgA vs IgG serum levels in CF patients affected by *BCC*, in both intermittently and chronically colonised patients, is still unclear²⁴. However, we expect the peptide array platform to be perfectly suited to monitor IgM and IgA response as well, to monitor the first phases of an infection and to establish its onset.

For the aforementioned reasons, we consider our method ready to be carefully tested on a larger clinical scale. In parallel, we expect the method validated here for the specific case of *BCC* in CF infections to be considered as general, with the potential to be extended to other diseases and meet other diagnostic needs.

Materials and Methods

Reagents. Reagents for peptide synthesis were from Iris Biotech (Marktredwitz, Germany). Other chemicals were from Sigma-Aldrich (St. Louis, MO, USA) if not stated otherwise. Goat anti-human IgG labeled by Cy3 were obtained from Jackson ImmunoResearch (West Grove, PA, USA). Silicon slides were from SVM (Sunnyvale, CA).

Epitope prediction and design. We first verified that selected antigens BPSS2141, BPSL2765 and BPSL1050 are well conserved among *BCC* members, and at the same time they are poorly conserved in the other microorganisms considered, which represent common co-infecting pathogens in CF patients (see Supplementary Figure S1). For this task, we used the web-based, freely available resources at EnsemblGenomes¹⁷, searching and aligning our query sequences (antigens and epitopes) using BLASTp.

The epitopes were selected and designed using the MLCE prediction method (Matrix of Local Coupling Energies, MLCE^{25,26}, available also as a public web-based service <http://bioinf.uab.es/BEPPE>)^{8,9,11,27,28}. MLCE is based on the idea that structural dynamics determines why specific parts of a protein antigen may bind to antibodies and/or tolerate mutations that allow the pathogen to escape protective immune responses, all typical traits of B-cell epitopes. These salient aspects are captured in MLCE by the concept that epitope regions correspond to non-optimized intramolecular interaction-networks in the Ag structure.

The structure of antigen BPSL0919 was reconstructed by homology with protein IspH1 from *E. coli*. (PDB ID: 3F7T, 58% S-W identity and 83% S-W similarity²⁹ with BPSL0919 protein sequence. The initial model was built using Modeller 9³⁰, and further refined via MD simulation. The simulation and analysis were performed using the GROMACS 4.54 software package³¹, GROMOS 53A6 force field³² and the SPC water model³³. The antigen was simulated in NPT conditions for 50 ns at 300 K. Epitope predictions were carried out using the MLCE prediction

method^{25,26} on the representative structure of the most populated structural cluster obtained using the method developed by Daura *et al.*³⁴.

Peptide synthesis. All linear peptides were synthesized by stepwise microwave-assisted Fmoc-SPSS on a Biotage ALSTRA Initiator+ peptide synthesizer according to well-established protocols^{8–10,11,35}. Briefly, peptides were assembled on a 2-CTC resin. Chain elongation was performed by iterative cycles of amino acids coupling (using Oxyma/DIC as activators) and Fmoc-deprotection using a 20% piperidine solution in DMF. Upon complete chain assembly, peptides were cleaved from the resin using a 2.5% TIS, 2.5% thioanisole, 2.5% water, 92.5% TFA mixture. Crude peptides were then purified by preparative RP-HPLC. MS analysis was performed separately on purified material.

In the case of conformational epitopes (PAL2, OPPA1), we used additional suitable spacers (either PEG or additional Gly residues) to combine different parts of the epitope sequences such that the relative distance between the terminals could reproduce the average one observed in the cognate antigen (Peri *et al.*²⁸).

All epitopes were conjugated to PEG units as spacers to separate the probes from the microarray surface.

Peptide microarrays. Silicon slides were coated by copoly(DMA-NAS-MAPS) according to the protocol described in³⁶. Briefly, silicon slides were immersed in a polymer solution (1% w/v in 0.9 M (NH₄)₂SO₄) for 30 min, then rinsed with water, dried under nitrogen and cured for 15 min under vacuum at 80 °C. To functionalize slides with maleimido groups, slides were immersed in a solution containing 6 mM of N-(2-Aminoethyl) maleimide trifluoroacetate salt and 7 mM N,N-Diisopropylethylamine in dimethyl formamide (DMF) for 5 hours at 40 °C. Slides were then thoroughly washed in DMF and water and dried under a stream of nitrogen.

Peptides were first dissolved with 50% acetonitrile solution (v/v, in Milli-Q water) to 10 mg/mL stock solution and then diluted to 2 mg/mL into the printing buffer (25 mM Na/Acetate pH 4.8, 15 mM trehalose). Microarrays were prepared using a non-contact sciFLEXARRAYER S3 (Sciencion Co., Berlin, Germany) spotter. All the peptide samples were printed in triplicates, each sub-array has positional controls (Cy3-streptavidin) and positive controls (human IgG at the concentration of 0.2 mg/mL). Sixteen arrays were spotted on each silicon slide corresponding to the sixteen compartments created by NEXTERION[®] IC-16 sixteen well Incubation Chamber from Schott (Jena, Germany). Printed slides were placed in a humid chamber and incubated overnight at room temperature. Then they were blocked with 2% w/v BSA in PBS for 1 h, washed with water and dried under a stream of nitrogen.

Serum was diluted 1:100 in LowCross-Buffer[®] from Candor (Wangen, Germany) and 40 µL were added into each microarray well, incubated for 60 min on a shaker (150 rpm, 22 °C). BCC positive, BCC negative and healthy sera were incubated in parallel up to 16 samples per slide. Negative controls included blank arrays incubated only with the secondary antibody and with incubation buffer.

The microarray slide was then rinsed for 3 times with washing buffer (0.05 M Tris/HCl pH 9, 0.25 M NaCl, 0.05% v/v Tween 20) and PBS and incubated with 40 µL of 1 µg/mL Cy-3 labelled goat anti-hIgG for another 60 min followed by the same washing steps as described above.

Fluorescence was detected by a ProScanArray scanner (PerkinElmer, Boston, MA) using 70% photomultiplier (PMT) gain and laser power.

Fluorescence intensities were analyzed using the QuantArray software from PerkinElmer. The fluorescence intensities were corrected for spot-specific background, values for replicate spots were averaged. t Student tests over the groups of samples and ROC curve analysis were performed using Prism 6 from GraphPad.

Cluster analysis. The standardization of data was carried out on the raw fluorescence intensities collected from the microarray tests using median and MAD (Median Absolute Deviation) of the distribution $f(Z)$ from each probe.

The standardized value for each data point associated to the same probe is calculated as:

$$\tilde{Z}_i = \frac{Z_i - \text{median}(fZ)}{\text{MAD}}; \quad (1)$$

The cluster analysis was carried out on the standardized dataset using a bottom-up hierarchical approach based on Manhattan³⁷ metric to define the distance $d(x_i, y_j)$ between points $x_i \in X$ and $y_j \in Y$, and average linkage to define the distance $D(X, Y)$ between clusters:

$$D(X, Y) = \frac{1}{N_X \times N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} d(x_i, y_j); \quad (2)$$

Statistical analysis were performed using R³⁸.

Samples. Serum samples drawn from healthy controls and CF patients were from the biobank of Fondazione IRCCS Ospedale Maggiore Policlinico, Milano. All samples were collected, analyzed and stored according to the rules of the local ethical committee. Table S2 in the Supplementary informations summarizes the details on the serum samples used in this work. The microbiological evaluation was performed on sputum samples from patients followed at CF centre of Milan, routinely processed and investigated, according to CF microbiological standard procedure, every 3 months over a one year period. Samples, obtained from the low respiratory tract, were firstly mixed with dithiothreitol at a 1:1 dilution to reduce viscosity, and then incubated at 37 °C for 30 minutes. Treated samples were plated on several media and plates were incubated at 37 °C for 72 hours. In particular, *B. cepacia* complex strains grow on *Burkholderia Cepacia* Selective Agar (BCSA) and on MacConkey

agar. All presumptive *B. cepacia* complex colonies were firstly identified by Maldi-Tof spectrometry and successively spread over blood agar for 24 h and stored at -80°C in order to permit further examinations to have genomovar identification. The *recA* gene of the *B. cepacia* complex strains was examined, PCR primers used were: BCR1: TGACCGCCGAGAAGAGCAA and BCR2: CTCTTCTTCGTCATCGCCTC.

Analysis of the *recA* gene of the *B. cepacia* complex provides a rapid and robust nucleotide sequence-based approach to identify and classify this taxonomically complex group of opportunistic pathogens³⁹. The sequences were submitted to GenBank as an aligned set using BLAST, in order to get the identification.

A document informing that samples could be used for research purposes has been signed by all subjects whose serum was used in the experiments. Experiments have been carried out following current guidelines and regulations.

References

1. Limmathurotsakul, D. *et al.* Enzyme-Linked Immunosorbent Assay for the Diagnosis of Melioidosis: Better Than We Thought. *Clin. Infect. Dis.* **52**, 1024–1028 (2011).
2. Hara, Y., Chin, C.-Y., Mohamed, R., Puthuchery, S. D. & Nathan, S. Multiple-antigen ELISA for melioidosis—a novel approach to the improved serodiagnosis of melioidosis. *BMC Infect. Dis.* **13**, 165 (2013).
3. Felgner, P. L. *et al.* A Burkholderia pseudomallei protein microarray reveals serodiagnostic and cross-reactive antigens. *Proc. Natl. Acad. Sci. USA* **106**, 13499–13504 (2009).
4. Liang, L. *et al.* Immune profiling with a Salmonella Typhi antigen microarray identifies new diagnostic biomarkers of human typhoid. *Sci. Rep.* **3**, 1043 (2013).
5. Buus, S. *et al.* High-resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays. *Mol. Cell. Proteomics* **11**, 1790–1800 (2012).
6. Carmona, S. J. *et al.* Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants. *Mol. Cell. Proteomics* **14**, 1871–1884 (2015).
7. Lu, Y. *et al.* Chimeric peptide constructs comprising linear B-cell epitopes: application to the serodiagnosis of infectious diseases. *Sci. Rep.* **5**, 13364 (2015).
8. Lassaux, P. *et al.* A Structure-Based Strategy for Epitope Discovery in Burkholderia pseudomallei OppA Antigen. *Structure* **21**, 167–175 (2013).
9. Gourlay, L. J. *et al.* Exploiting the Burkholderia pseudomallei acute phase antigen BPSL2765 for structure-based epitope discovery/design in structural vaccinology. *Chem. Biol.* **20**, 1147–1156 (2013).
10. Gaudesi, D. *et al.* Structure-based design of a B cell antigen from *B. pseudomallei*. *ACS Chem. Biol.* **10**, 803–812 (2015).
11. Gori, A. *et al.* Flexible vs Rigid Epitope Conformations for Diagnostic- and Vaccine-Oriented Applications: Novel Insights from the Burkholderia pseudomallei BPSL2765 Pal3 Epitope. *ACS Infect. Dis.* **2**, 221–230 (2016).
12. Chmiel, J. F. *et al.* Antibiotic management of lung infections in cystic fibrosis: I. The microbiome, methicillin-resistant Staphylococcus aureus, gram-negative bacteria, and multiple infections. *Ann. Am. Thorac. Soc.* **11**, 1120–1129 (2014).
13. McCaughey, G., Gilpin, D., Elborn, J. & Tunney, M. M. The future of antimicrobial therapy in the era of antibiotic resistance in cystic fibrosis pulmonary infection. *Expert Rev. Respir. Med.* **7**, 385–396 (2013).
14. Ciofu, O., Tolker-Nielsen, T., Jensen, P. Ø., Wang, H. & Høiby, N. Antimicrobial resistance, respiratory tract infections and role of biofilms in lung infections in cystic fibrosis patients. *Adv. Drug Deliv. Rev.* **85**, 7–23 (2015).
15. Abbott, I. J. & Peleg, A. Y. Stenotrophomonas, Achromobacter, and Nonmelioid Burkholderia Species: Antimicrobial Resistance and Therapeutic Strategies. *Semin. Respir. Crit. Care Med.* **36**, 99–110 (2015).
16. Lau, S. K. P. *et al.* Laboratory diagnosis of melioidosis: past, present and future. *Exp. Biol. Med. (Maywood)*. **240**, 742–751 (2015).
17. Kersey, P. J. *et al.* Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.* **44**, 574–580 (2015).
18. Pirri, G., Damin, F., Chiari, M., Bontempi, E. & Depero, L. E. Characterization of A Polymeric Adsorbed Coating for DNA Microarray Glass Slides. *Anal. Chem.* **76**, 1352–1358 (2004).
19. Regenmortel, M. H. V. Van. Synthetic Peptide Vaccines and the Search for Neutralization B Cell. *Open Vaccine J.* **2**, 33–44 (2009).
20. Farrell, P. M. The prevalence of cystic fibrosis in the European Union. *J. Cyst. Fibros.* **7**, 450–453 (2008).
21. Sagel, S., Gibson, R. & Emerson, J. Impact of Pseudomonas and Staphylococcus Infection on Inflammation and Clinical Status in Young Children with Cystic Fibrosis. *J. Pediatr.* **154**, 183–188 (2009).
22. Hayes, D., Farrell, P. M., Li, Z. & West, S. E. Pseudomonas aeruginosa serological analysis in young children with cystic fibrosis diagnosed through newborn screening. *Pediatr. Pulmonol.* **45**, 55–61 (2010).
23. Tramper-Stranders, G. *et al.* Initial Pseudomonas aeruginosa infection in patients with cystic fibrosis: Characteristics of eradicated and persistent isolates. *Clin. Microbiol. Infect.* **18**, 567–574 (2012).
24. Adeboyeke, D. U., Sanderson, J. L., Taylor, P. & Hodson, M. E. Longitudinal study of immunoglobulin A and immunoglobulin G antibodies to Burkholderia cepacia in adult Cystic Fibrosis (CF) patients. *Thorax* **55**, A11–A11 (2000).
25. Scarabelli, G., Morra, G. & Colombo, G. Predicting Interaction Sites from the Energetics of Isolated Proteins: A New Approach to Epitope Mapping. *Biophys. J.* **98**, 1966–1975 (2010).
26. Peri, C. *et al.* In *Pept. Antibodies Methods Protoc.* (ed. Gunnar, H.) **1348**, 13–22 (Springer, 2015).
27. Soriani, M. *et al.* Exploiting Antigenic Diversity for Vaccine Design: The chlamydia Artj paradigm. *J. Biol. Chem.* **285**, 30126–30138 (2010).
28. Peri, C. *et al.* Rational epitope design for protein targeting. *ACS Chem. Biol.* **8**, 397–404 (2013).
29. Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
30. Webb, B. & Sali, A. In *Curr. Protoc. Bioinforma* (John Wiley & Sons, Inc., doi: 10.1002/0471250953.bi0506s47 2014).
31. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinforma.* **29**, 845–854 (2013).
32. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004).
33. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
34. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie Int. Ed.* **38**, 236–240 (1999).
35. Nithichanon, A. *et al.* Sequence- and Structure-Based Immunoreactive Epitope Discovery for Burkholderia pseudomallei Flagellin. *PLoS Negl. Trop. Dis.* **9**, e0003917 (2015).
36. Cretich, M. *et al.* High sensitivity protein assays on microarray silicon slides. *Anal. Chem.* **81**, 5197–5203 (2009).
37. Black, P. E. *Manhattan distance*. Vreda Pieterse Paul E. Black, eds. (2006). Available at: <http://www.nist.gov/dads/HTML/manhattanDistance.html>. (Accessed: 1 June 2016).
38. R Core Team. *R: a language and environment for statistical computing*. *R Found. Stat. Comput. Vienna, Austria* (2015). Available at: <http://www.r-project.org/> (Accessed: 1 June 2016).

39. Mahenthiralingam, E. *et al.* DNA-Based Diagnostic Approaches for Identification of Burkholderia cepacia Complex, Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia stabilis, and Burkholderia cepacia Genomovars I and III. *J. Clin. Microbiol.* **38**, 3165–3173 (2000).
40. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
41. Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

Acknowledgements

This project was supported by Fondazione CARIPLO (Progetto Vaccini, contract number 2009–3577) and by Bando Regione Lombardia - Fondazione Cariplo 2013, progetto PROVA, ID 42666248 (2013-0349).

Author Contributions

C.P., Marina C. and G.C. wrote the main manuscript. C.P. analyzed the sequences and the epitopes, designed the peptides, performed statistical analysis and prepared the article figures. M.C. supervised the microarray experiments and analyzed microarray data. P.G. performed the microarray experiments. A.G. designed and synthesized the peptides. L.S. synthesized polymers for microarray coatings. D.G. and S.S. collected and characterized the serum samples under the supervision of L.C. and Marina C. developed the microarray surface chemistry used in this study and G.C. conceived and supervised the project.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Peri, C. *et al.* Evolving serodiagnostics by rationally designed peptide arrays: the *Burkholderia* paradigm in Cystic Fibrosis. *Sci. Rep.* **6**, 32873; doi: 10.1038/srep32873 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016