B-cell epitope mapping of TprC and TprD variants of *Treponema pallidum* subspecies

Barbara Molini[a+], Mark C. Fernandez[a+], Charmie Godornes[a], Anastassia Vorobieva[b]

Sheila A. Lukehart[a,c], and Lorenzo Giacani[a,c #]

[a]Department of Medicine, University of Washington, Seattle WA, USA

[c]Center for Structural Biology, Vlaam Instituut voor Biotechnologie, Vrije Universiteit Brussels, Brussels, Belgium

[c]Department of Global Health, University of Washington, Seattle WA, USA

[+]Contributed equally to this work

[#]Corresponding author: giacal@u.washington.edu

Running title: TprC/D B-cell epitope mapping

Keywords: *Treponema pallidum*, syphilis, Tpr proteins, B-cell epitope mapping, vaccine development

ABSTRACT

Several recent studies have focused on the identification, functional analysis, and structural characterization of outer membrane proteins (OMPs) of *Treponema pallidum* (*Tp*). The *Tp* species encompasses the highly related *pallidum*, *pertenue,* and *endemicum* subspecies of this pathogen, known to be the causative agents of syphilis, yaws, and bejel, respectively. These studies highlighted the importance of identifying surface-exposed OMP regions and the identification of B-cell epitopes that could be protective and used in vaccine development efforts. We previously reported that the TprC and TprD OMPs of *Tp* are predicted to contain external loops scattered throughout the entire length of the proteins, several of which show a low degree of sequence variability among strains and subspecies. In this study, these models were corroborated using AlphaFold2, a state-of-the-art protein structure modeling software. Here, we identified B-cell epitopes across the full-length TprC and TprD variants using the Geysan pepscan mapping approach with antisera from rabbits infected with syphilis, yaws, and bejel strains and from animals immunized with refolded recombinant TprC proteins from three syphilis strains. Our results show that the humoral response is primarily directed to sequences predicted to be on surface-exposed loops of TprC and TprD proteins, and that the magnitude of the humoral response to individual epitopes differs among animals infected with various syphilis strains and *Tp* subspecies. Rather than exhibiting strain-specificity, antisera showed various degrees of cross-reactivity with variant sequences from other strains. The data support the further exploration of TprC and TprD as vaccine candidates.

INTRODUCTION

The human treponematoses (syphilis, yaws, and bejel) are caused by a group of highly related pathogens classified as subspecies of the spirochete bacterium *Treponema pallidum* (*Tp*). Classically, the *pallidum* subspecies is said to causes syphilis, while the *pertenue* and *endemicum* subspecies are regarded as the causes of yaws and bejel, respectively (1), although the modes of transmission and the clinical manifestations are similar among subspecies. These diseases are still a concern for public and global health, as they continue to result in substantial morbidity and mortality worldwide. According to the World Health Organization, the global prevalence of syphilis is ~20 million cases, with an incidence of ~6.3 million new cases every year (2). Although most of these infections occur in low- and middle-income countries, syphilis has resurged in industrialized nations of Asia, Europe, and North America. In the United States, the incidence of infectious syphilis has risen steadily over the last two decades (3-7) reaching approximately 39,000 cases in 2019, a 6.5-fold increase compared to the ~6,000 cases reported in 2000. If left untreated, syphilis can progress to affect the cardiovascular and central nervous systems of patients, potentially leading to manifestations such as aortic aneurysm, stroke, hearing or visual loss, dementia, paralysis, and death (8). Additionally, vertical transmission of syphilis is estimated to account for ~1/3 of stillbirths in sub-Saharan Africa and a high proportion of perinatal morbidity and mortality globally (9, 10). Past public health initiatives to eliminate syphilis and congenital syphilis promoted by the CDC and WHO (11, 12) have significantly aided in reducing syphilis incidence and in generating awareness of this disease, but have not achieved their intended elimination goals. Compared to syphilis, less accurate

epidemiological data are available on yaws and bejel. Although it was recently estimated that ~65,000 cases of yaws occurred annually in 13 endemic countries, this is likely an underestimate of the global burden of the disease, given that in at least 19 potentially endemic countries the incidence of yaws is unknown (13). While the ongoing elimination campaign in Asia and Africa using mass administration of azithromycin has demonstrated promising results (14), such efforts could be undermined by the spreading of macrolide resistant *Tp* subsp. *pertenue,* as recently demonstrated in Papua New Guinea (15). Foci of bejel have been reported in the last two decades, mostly in the Near East and Sahelian Africa (16-19), and bejel strains have recently reported to be transmitted sexually (20).

The chance of success of current and future control campaigns for all treponematoses would significantly increase if effective vaccines were available (21, 22). The most rational approach to vaccine development for these infections requires a clear understanding of the type of immune response that is protective and the identification of suitable candidate antigens to be tested in a pre-clinical animal model (21, 22). Furthermore, because there is very limited or no cross-immunity between subspecies of *Tp* and only sporadic cross-immunity between syphilis strains (23, 24), the identification of antigenic differences in potential vaccine candidates among subspecies and strains is of pivotal importance, as such differences could be key to devising a broadly protective vaccine (22). There is consensus that vaccine candidates are most likely to be found among these spirochetes' surface-exposed antigens, such as (but not limited to) integral outer membrane proteins (OMPs). As in all dual-membrane bacteria, *Tp* integral OMPs will necessarily contain a membrane-embedded β-barrel domain composed of antiparallel β-strands joined together by loops that alternatively protrude toward the

89 extracellular environment or the periplasm (25). Because *Tp* clearance from early lesions is
90 dependent on opsonophagocytosis of *Tp* cells by activated macrophages (26, 27), the
91 identification of surface-exposed epitopes that can be targeted by immunization to induce
92 opsonic antibodies and promote macrophage activation is key to vaccine development. Such
93 tasks, however, have been historically challenging due to the inability to steadily propagate the
94 *Tp* subspecies *in vitro*, which was only recently achieved (28), and also because of the
95 uncommon fragility and limited protein content of these spirochetes' OM (29, 30). These
96 limitations have been partially overcome by the ability to predict *in silico* OMP-encoding genes
97 and the structure of their encoded proteins, enabling investigation using structural and
98 functional experimental approaches (31, 32).
99 Among *Tp* putative OMPs identified to date, there are several members of the *T.*
100 *pallidum* repeat (Tpr) family of paralogous proteins, including TprC and TprD (encoded by the
101 *tp0117* and *tp0131* genes in the reference Nichols strain, respectively) (33); these are reported
102 to have OM localization and porin activity (34, 35). In this study, we examine the protein
103 sequence variation in TprC and TprD among *T. pallidum* strains and subspecies, and predict,
104 then confirm, the locations of B cell epitopes using antisera from infected and immunized
105 rabbits. Variant specificity and cross-immunity are analyzed so that epitopes with broad
106 coverage among strains and subspecies can be identified for future evaluation as vaccine
107 antigens.
108
109 **RESULTS**
110 **Sequence analysis of TprC and TprD variants.**

Although the TprC and TprD proteins are identical in the Nichols, Chicago, and Bal73-1 strains, allelic variants of TprC and TprD exist among syphilis strains and the three *Tp* subspecies (35, 36). Among the treponemal strains used in this study (Fig.1), four alleles were found at the *tprD* locus, which include the reference *tprD* allele (found in the syphilis Nichols, Chicago, and Bal73-1 strains), and the *tprD*$_2$ allele (found in the syphilis strains MexicoA, Sea81-4, Bal3, and UW249) which encodes the TprD$_2$ protein (35). Also the subsp. *pertenue* SamoaD strain and subsp. *endemicum* IraqB strains harbor a *tprD*$_2$ allele in the *tprD* locus, but their TprD$_2$ amino acid sequences differ from the subsp. *pallidum* TprD$_2$ sequence due to five amino acid substitutions scattered throughout the length of the protein (Fig.1) (35). TprD$_2$ has four unique regions that differentiate it from the reference TprD sequence. These include a large central region of 110 amino acids and three smaller regions toward the COOH-terminal end of the protein (Fig.1) (35). As previously reported, the *tprC* locus of MexicoA, Sea81-4, and Bal3 encodes a TprC variant with a limited number of amino acid (aa) changes (15 aa for MexicoA, 9 aa for Sea81-4, and 9 aa for Bal3) compared to the reference TprC found in Nichols, Chicago, and Bal73-1 strains (Fig.1) (35). The TprC protein of the *pertenue* and *endemicum* strains studied here also shows limited amino acid changes compared to the reference TprC (31 aa for SamoaD and 26 aa for IraqB; Fig.1), albeit higher compared to the subsp. *pallidum* strain (35). We previously reported that TprC and TprD/D$_2$ sequence variation does not occur randomly, but rather is localized in discrete variable regions (DVRs; Fig.1) (35). When TprC and TprD variants are compared (with the exclusion of TprD$_2$), seven DVRs are found throughout the protein sequence, while 8 DVRs can be identified within the TprD$_2$ sequences (Fig. 1). To obtain predictions of TprC and TprD$_2$ structures from their amino acid sequences (Fig. 1) and map the

DVRs on these models, we used the recently developed AlphaFold2 software (https://AlphaFold.ebi.ac.uk/) (37). These new models revealed remarkably similar structures for TprC and TprD/D_2 and identified these proteins as relatively large β-barrel integral OMPs of 20 β-strands connected by ten external loops (ExLs, protruding toward the extracellular milieu), and nine periplasmic loops (Fig.2A). The local model quality, indicated by the pLDDT gradient was high in the transmembrane and periplasmic loop regions, and slightly lower in the predicted ExLs, suggesting conformational flexibility (Fig. 2B). Except for two substitutions (aa 407 and 410 mapping to a periplasmic β-turn; Fig.1), all DVRs localized within a subset of the surface-exposed ExLs (Fig.2C). More specifically, DVRs were located in ExL1, ExL5-6 and ExL8-10 of the TprC and TprD/D_2 models; while ExL2-4 harbored conserved loops. ExL7 is also conserved between $TprD_2$ sequences from various isolates, although its shows only 60% of sequence identity to the ExL7 of other TprC and TprD variants (Fig.1). *In silico* prediction of B-cell epitopes using BepiPred2.0 (http://www.cbs.dtu.dk/services/BepiPred/), IEDB (https://www.iedb.org/), and BCpreds (http://ailab-projects1.ist.psu.edu:8080/bcpred /data.html) (File S1) showed that the putative TprC and TprD ExLs were also enriched in immunogenic epitopes. Therefore, it is possible that the antigenic variability in the ExLs regions has functional significance in immunity to the *T. pallidum* subspecies. To validate the B-cell epitope prediction and evaluate the cross-reactivity of these epitopes across species and strains, we performed experimental B-cell epitope mapping of the TprC, and TprD/D_2 proteins with a Geysan pepscan approach based on overlapping synthetic peptides (38) using sera from animals infected with *Tp* subsp. *pallidum*, *Tp* subsp. *pertenue*, and *Tp* subsp. *endemicum* strains. Furthermore, we compared antibody

reactivity in sera from infected rabbits with that of sera from rabbits immunized with a subset of full-length refolded recombinant TprC proteins.

Humoral responses to homologous TprC and TprD/D_2 peptides in experimentally infected rabbits.

Serum samples from groups of three laboratory rabbits infected intratesticularly (IT) with one of seven syphilis strains (Nichols, Chicago, Bal73-1, MexicoA, Sea81-4, Bal3, and UW249), one yaws strain (SamoaD), and one bejel strain (IraqB) were obtained at day 30, 60, and 90 post-infection. Pooled sera from animals in each infection group/time point were tested in ELISA to assess reactivity to homologous overlapping synthetic peptides (20-mers overlapping by 10 amino acids) representing the TprC and TprD/D_2 variants previously identified in each strain. The full list of synthetic peptides used in this study, with amino acids encompassing predicted ExLs highlighted in red with yellow text, and percentage amino acid homology among peptides across strains is shown in Table 1. Peptide nomenclature is explained in Table 1 footnote. Cumulative absorbance data from the three timepoints (sum of the mean absorbance values for day 30, 60, 90 values for each infected rabbit group) are reported in Fig.3A-C. Epitope mapping studies of the NH_2-terminal portion of the protein resulted in the identification of six highly reactive peptide regions (Fig.3A) representing sequences shared by all TprC and TprD genes in the studied subspecies *pallidum* strains: C1-C3, C6, C13-C14, C18, C20, and C25-C29. Based on AlphaFold2 structural predictions, 9 of these 13 peptides had at least 7 amino acids mapping to the predicted external loops of the protein, while only four reside in predicted transmembrane transmembrane scaffolding and periplasmic

loop regions (C1, C6, C20 and C25; Fig.1A, and Table 2). It is noteworthy that all three B cell epitope prediction programs uniformly predicted all six of the experimentally determined epitope-containing regions of the NH_2-terminal portion of the subspecies *pallidum* TprC and D proteins (File S1).

Several epitopes were also identified in the COOH-terminal region of these proteins, and corresponded to peptides the same regions in *pallidum* and non-*pallidum* subspecies: C46 and C47 homologs from Nichols (Fig.3A, Table 2), SamoaD (S-C46, S-C47; Fig.3B and Table 2) and IraqB (I-C46, I-C47; Fig.3B and Table 2), C51 homologs from Nichols (N-C51), SamoaD and iraqB (S/I-C51) (Fig.3A-B and Table 2); and C53-C55 homologs from Nichols, Bal3/Sea81-4 (Fig.3A and Table 2), and IraqB (Fig.3B and Table 2). Similarly, the C43, and D45-D47 (ExL8) peptides, mapping to the $TprD_2$ COOH-terminus were found to contain B-cell epitope(s) (Fig.3C and Table 2). Additional $TprD_2$ peptides found to be reactive were D33-D35 (ExL6), I-C39, D40-41 (ExL7), C49, and D51 (ExL9). In our 3D models of these proteins, all the reactive peptides in the COOH-terminus fall within predicted ExLs (Tables 1-2 and Fig.1), except for C43, most of I-C39 (75%), C49, and C53, which are predicted transmembrane transmembrane scaffolding sequences. Of these "scaffold epitopes", only one (C43) was predicted by a B-cell prediction program (File S1).

The percentage of immune sera that showed reactivity to many of the peptides was variable. For example, peptides C3, and C13 were recognized by rabbits infected with 28% of the *Tp* subsp. *pallidum* strains; peptides C1, and C27 were recognized by rabbits infected with 42% of the strains; peptides C14 and C28 were recognized by rabbits infected with 57% of the strains; C6 was recognized by rabbits infected with 71% of the strains; and peptides C2 and C18 were recognized by rabbits infected with 85% of the *Tp* subsp. *pallidum* strains (Fig.1). Overall,

198 based upon the AlphaFold2 models, these results show that humoral reactivity elicited to these
199 Tpr antigens during experimental infection is directed primarily to predicted surface-exposed
200 regions of the TprC/D and TprD$_2$ proteins.
201
202 **Reactivity to non-homologous TprC and TprD$_2$ peptides**
203 Epitope mapping using short peptides based on the TprC and TprD/D$_2$ sequences from
204 multiple *Tp* strains and sera from infected animals also allowed us to investigate cross-reactivity
205 to non-homologous peptides to determine the fine specificity of the antibody response to these
206 antigens. Such analyses focused on peptides mapping to the proteins` COOH-terminal regions,
207 due to the higher sequence variability in this region, compared to the more conserved NH$_2$-
208 terminal region (Fig.1). Major variable regions include peptides C46 - C47 (mapping to the
209 predicted ExL8), C51 (ExL9), and C55 (ExL10) (Fig.1 and Tables 1-2). Four distinct variants of
210 each of the C46, C47, and C55 peptides, and three variants of C51, representing all sequences
211 found in the strains studied here, were tested against all nine pools of immune sera obtained at
212 day-90 post-experimental infection.
213 As shown in Fig.4A-D, very few sera were reactive only to their homologous peptide. For
214 example, the Bal73-1and SamoaD antisera were primarily reactive only to their own C46
215 sequences (Fig.4A), although the Bal73-1 antiserum showed a very modest reactivity to the
216 IraqB peptide variant (Fig.4A). In contrast, none of the sera tested against the C47 variants
217 exhibited reactivity against the homologous peptide, but the Chicago, Bal73-1, IraqB and Sea81-
218 4 sera were reactive against some heterologous variants (Fig.4B). Only Chicago and Bal73-1 sera
219 showed complete strain-specificity for the C51 peptides (Fig.4C), while none of the sera reactive

to C55 showed complete strain-specificity (Fig.4D). When tested against TprD2 peptides, most antisera did not show any reactivity. There were, however, two exceptions, as the Chicago sera cumulatively showed reactivity to the D34 and D47 peptides, with OD values of 3.6 and 6.6, respectively. However, only the D47 peptide was consistently recognized at all time points, while D34 was recognized only at day 60. Overall, these data indicate a relatively high level of cross immunity and perhaps suggest that immunization with a given sequence might generate cross-reactive antibodies able to overcome the obstacle of sequence diversity among TprC epitopes, a feature that is desirable in vaccine development as they may be broadly opsonic or neutralizing.

Humoral response to TprC peptides following rabbit immunization with full-length refolded antigens

Refolded antigens, analyzed using circular dichroism (CD), were found to have a β-barrel component of about 48% for all three antigen variants. Random coil was also found to be 48% of the protein structure, while only 4% was identified as alpha helices. Epitopes recognized following immunization with any of three recombinant full-length TprC variants from *Tp* subsp. *pallidum* strains (Nichols/Chicago/Bal73-1, Sea81-4/Bal3, and MexicoA) were also identified to evaluate differences with infection-induced immunity. Results showed that sera from animals immunized with the Nichols TprC sequence were highly reactive to peptides C1-C3, C6 and C47, and moderately reactive to peptides C5, C9, C16-18, C28, C32, C34-C35, C53 and C55 (Fig.5A). Of these 16 peptides, six mapped almost exclusively to putative surface-exposed loop regions (C28, C32, C34, C35, C47, and C55), five (C1, C5-C6, C16, and C53) mapped to predicted

transmembrane transmembrane scaffolding sequences, while five peptides (C2-C3, C9 and C17-C18) contained both surface-exposed loops and scaffold regions. Sequences of these peptides and location in the predicted protein models are reported in Table 3. When tested against non-homologous peptides (Fig.5B), the Nichols TprC-immunized sera strongly recognized the SamoaD/IraqB C2-C3 variants, and all three heterologous C47 variants (SamoaD, Iraq B, and Sea81-4), while modest reactivity was seen towards the MexicoA/UW249 C55 peptide variant, the SamoaD/IraqB C34, and both C26 variants from SamoaD and IraqB (Fig.5B). Immunization with the Bal3 variant of TprC elicited high reactivity to peptides C1-3, C6, and C13, and moderate reactivity to peptides C7, C16-18, C20, C43, C47, and C49 (Fig.5C, Table 3). Of these thirteen peptides, six (C2-C3, C13, C17-C18, and C47) mapped predominantly to ExLs, and seven (C1, C6, C7, C16, C20, C43, and C49) predominantly to the protein transmembrane transmembrane scaffolding (Table 3). Cross-reactivity to non-homologous peptides was seen predominantly to the SamoaD/IraqB C2 and C3, IraqB C22, and all variants of C47 and C51 (Fig.5D). Antisera from rabbits immunized with the MexicoA TprC variants primarily recognized homologous peptides C1-3 and, secondarily, C5, C6, and C28 (Fig.5E). Of these six, one peptide mapped to the predicted ExL6 (C28), three mapped only to the transmembrane transmembrane scaffolding (C1, C5-C6) and two mapped to a peptide predicted to contain portions of both (C2-C3) (Table 3). Cross-reactivity to the non-homologous SamoaD/IraqB C2 and C3 was also detected (Fig.5F). Overall, these data show that, as seen in infection-induced immunity, the humoral response following immunization with full-length TprC variants is mainly elicited by predicted surface-exposed sequences, rather than sequences mapping to the β-

263 barrel transmembrane scaffolding, and that cross-reactivity to non-homologous peptides is
264 common.

265 A side-by-side comparison of the infection- vs. immunization-induced humoral response
266 to peptides is shown in Fig.6. For this comparison, the mean value of the cumulative reactivity
267 seen in sera at day 30, 60, and 90 sera post-experimental infection is shown for each peptide.
268 Sera from immunized animals were obtained three weeks after the last immunization. All sera
269 were tested at the same dilution. In general, immunization-induced reactivity to most peptides
270 appeared to be higher than that elicited by experimental infection; specific examples include
271 C1-C3, C5, C9, C16-C17, C28, C32, C34-C35, N-C47, and C53 peptides (Fig.6A). For Nichols-clade
272 *T. pallidum* strains (Fig 6A), which contain identical *tprC* and *tprD* loci, this was most noticeable
273 for epitopes located in the NH_2- and COOH-terminal regions of the protein. In contrast,
274 infection-induced antibody responses to epitopes in the central part of the protein were
275 comparable to or higher than those induced by immunization. For TprD2-containing subsp.
276 *pallidum* strains (Fig.6B and 6C), immunization-induced responses were limited to the NH_2-
277 terminal portion of the protein (including ExL1-3) and virtually no immunization-induced
278 antibodies were detected for epitopes in the central and COOH-terminal regions, although
279 these were recognized by infection-induced responses. Overall, these data support that, in
280 most cases, immunization elicits a higher reactivity to TprC B-cell epitopes compared to
281 experimental infection, particularly for those epitopes located in the NH_2-terminal portion of
282 the protein. These data support the preferential use of the amino portion of TprC, which
283 contains multiple conserved ExLs, for vaccine studies.
284

DISCUSSION

The continuing prevalence of syphilis, in the face of highly effective therapy and active control programs, highlights the need for a protective vaccine. The development of such a vaccine calls for a deeper understanding of the mechanisms of protective immunity and the antigens and adjuvants that induce protection. Our laboratories have been examining these issues for many years (22, 31, 39-47). Much of that work has focused on the Tpr antigens of *T. pallidum*. In this current study, B-cell epitope mapping studies of the TprC/D and $TprD_2$ proteins of *Tp* reveal that antibodies arising during experimental infection recognize sequences predicted, using state-of-the-art modeling systems, to fall largely in the proteins' surface-exposed loops. Because opsonic antibodies are required for efficient ingestion and killing of *T. pallidum* by macrophages, surface-exposed epitopes are attractive targets as vaccine candidate antigens.

A broadly protective vaccine would need to be effective against most strains of *T. pallidum*, optimally including the agents of syphilis as well as the endemic treponematoses yaws and bejel. Because some of the external loops of Tpr C/D and $TprD_2$ demonstrate sequence heterogeneity among strains and subspecies of *T. pallidum*, we expected that these epitopes might be strain-specific, similar to the specificity demonstrated for the variable regions of TprK (41, 43, 48). For this reason, we included seven strains of *Tp* subsp. *pallidum* as well as strains from the subspecies *pertenue* and *endemicum* in our work. Unexpectedly, we saw considerable cross reactivity of antibodies toward the variant peptides (Fig.4). These findings support the use of TprC/D as at least one component of broadly effective candidate vaccine.

The AlphaFold2 structural predictions for TprC/D and $TprD_2$, as well as our CD analyses of purified refolded recombinant TprC variants, support our model (35) that these Tprs are membrane-localized 20-stranded β-barrel proteins containing numerous surface-exposed loops. Very similar models for TprC were previously obtained using I-TASSER (49) (https://zhanggroup.org/I-TASSER/) (35). Interestingly, when AlphaFold2 and I-TASSER results are compared, the only difference is that I-TASSER splits ExL6 (Fig.1) into two external loops separated by a β-hairpin, so that I-TASSER predictions harbor 11 external loops instead of 10. AlphaFold2, on the contrary, predicts a significantly larger ExL6, mapping approximately to the proteins' central domains. AlphaFold2 is the new standard for *ab-initio* structural prediction, and in the 2020 Critical Assessment of protein Structure Prediction (CASP) global challenge, it outperformed any other structure prediction algorithm, including I-TASSER (https://predictioncenter.org/casp14/zscores_final.cgi). Furthermore, in a recent preprint (50), AlphaFold2 was shown to work well on structural prediction for membrane proteins, although the exercise focused mostly on alpha-helical membrane proteins, and additional analyses are necessary to establish the same benchmark for β-barrel proteins.

In previous work by Anand *et al.* (34, 51) significantly different models for the TprC/D proteins were reported, compared to those provided here. These models, however, are not supported by AlphaFold2, which finds the structure of all Subfamily I and Subfamily II Tpr family members to be very similar to the structures for TprC/D and $TprD_2$ in Fig.2A. Although there is not unanimous agreement on the structure of these antigens within our scientific community, our epitope mapping data support our AlphaFold2 models, predicting a predominantly β-barrel structure for TprC and TprD/D2 (34, 51). Further studies and integration of all the structural,

329 functional, and immunological data are needed to establish a consensus on the structure of
330 these antigens until crystallographic (or equally reliable) data become available.
331 This study also provides evidence that infection with different strains might lead to
332 differences in the breadth and intensity of the humoral response against the same epitope, as
333 reported previously for responses to longer portions of the Tpr proteins (52). It is the case, for
334 example, of rabbits infected with the Sea81-4 strain of *Tp* that overall recognize more TprC/D
335 peptides compared to other *Tp* subsp. *pallidum* strains. The biological basis for these
336 differences is unclear at this time, in part due to the limitations of our understanding of *Tp*
337 biology and syphilis pathogenesis. As the technical gap in the approaches to study this difficult
338 pathogen narrows, and genomics, proteomics, and transcriptomics data populate public
339 repositories, more light will be shed on the causes of differential reactivity. Overall, however, it
340 is plausible to postulate that enhanced serological reactivity might be due to an overall
341 increased expression of the target antigen in each strain. This hypothesis is supported by
342 previous work where we showed the *tprC* mRNA level was higher in the Sea81-4 strain
343 compared to other *Tp* subsp. *pallidum* strains (Nichols, Chicago, Bal73-1) used in this study (53).
344 Our studies further demonstrated that epitopes in TprC/D and $TprD_2$ are nearly-
345 uniformly distributed across the length of the protein, even though the most reactive peptide
346 epitopes are in the NH_2- and COOH-terminal regions. Previously published (47) and ongoing
347 experiments have shown that both of these regions in the Nichols TprC protein contain
348 protective epitopes, as immunization with these antigen fragments significantly attenuated
349 lesion development upon infectious challenge (47), and polyclonal antisera elicited by
350 immunization with these portions facilitated treponemal ingestion by macrophages in

opsonophagocytosis assays compared to normal rabbit sera (47) . Further work, however, will be necessary to identify which specific surface-exposed sequences provide targets for opsonic antibodies, which may not coincide with sero-dominant epitopes, as the pathogen gains an obvious advantage by exposing to the immune system epitopes with little or no protective value.

Protective B-cell epitopes (contrary to T-cell epitopes) are often conformational, and even when a significant portion of an epitope appears to be a short linear peptide, as in our study, it does not necessarily mean that the peptide represents the full epitope or, if it does, that the sequence will not require a certain conformation to elicit optimal bioactivity. For this reason, in the immunization studies performed in this study, we used CD-confirmed refolded recombinant antigens. The immunization-induced antibodies generally identified the same epitopes seen in infection, supporting the role of refolding in mimicking native structure, but immunization also resulted in recognition of a broader range of epitopes than seen during infection, including transmembrane scaffolding regions. This is likely because the scaffold regions are not shielded by the outer membrane in an immunization setting and are thus more easily processed for recognition. Thus, the design of vaccine immunogens is critical. Possible approaches vary from placing epitopes within chimeric antigens that could work as scaffold or, alternatively, using portions of the protein containing protective epitopes as structural elements of the antigen, or even using single β-hairpins instead of the full-length antigens. The work reported here represents an important step in evaluating TprC/D and $TprD_2$ epitopes as part of the process that will lead to an effective vaccine for syphilis.

MATERIALS AND METHODS

Ethics Statement

New Zealand White rabbits were used for propagation of *T. pallidum* subspecies and strains and for experimental infections. Animal care was provided in accordance with the procedures described in the Guide for the Care and Use of Laboratory Animals (54) under protocols approved by the University of Washington Institutional Animal Care and Use Committee (IACUC, PI: Sheila Lukehart). The protocol number assigned by the IACUC committee that approved this study is 2090-08. No human samples were used in this study.

Strain propagation and experimental infection

Outbred adult male New Zealand White rabbits ranging from 3.0-4.0 Kg were obtained from R&R Rabbitry (Stanwood, WA). Prior to entry into the study, serum from each animal was tested with both a treponemal (FTA-ABS) and a non-treponemal (VDRL; BD, Franklin Lakes, NJ) test to rule out infection with the rabbit syphilis agent *Treponema paraluiscuniculi*. Only rabbits seronegative in both tests were used for either propagation or experimental infection for sample collection. *Tp* strains were propagated by intratesticular (IT) inoculation and harvested at peak orchitis as previously described (55). For experimental infections, groups of three rabbits were infected IT with a total of $5x10^7$ *Tp* cells per testis. In total, nine *Tp* isolates (one isolate per rabbit group) were used: seven *Tp* subsp. *pallidum* isolates (Nichols, Chicago, Bal73-1, Sea81-4, Bal3, MexicoA, and UW249), one *Tp* subsp. *endemicum* (IraqB) and one *Tp* subsp. *pertenue* (SamoaD) (Table 4). Briefly, on the day of infection bacteria were extracted from rabbit testes in sterile saline containing 10% normal rabbit serum (NRS), and testicular extract

was collected in sterile 15-ml tubes. Extracts were centrifuged twice at 1,000 rpm (180 x g) for 10 minutes in an Eppendorf 5810R centrifuge (Eppendorf, Hauppauge, NY) to remove gross rabbit cellular debris. Treponemes were enumerated under a dark-field microscope (DFM) and percentage of motile organisms was recorded. Extracts were then diluted in serum-saline to the desired concentration ($5x10^7$/ml). Following IT injection, treponemal motility was assessed again to ensure that the time elapsed before injection into the new host did not affect pathogen viability. After IT inoculation, establishment of infection was assessed by monitoring development of orchitis during the following three weeks as well as by performing FTA-ABS and VDRL tests on sera collected at day 30 post-inoculation. Immune sera were collected from the animals at day 30, 60, and 90 post-infection. Animals were then euthanized. Extracted sera were heat-inactivated at 56°C for 30 min and stored at -20°C until use for ELISAs.

Amplification and cloning of full-length *tprC* gene variants for expression of recombinant antigens.

Sequences for the *tprC* gene of *Tp* isolates (Nichols, Sea81-4, and MexicoA) were previously cloned (35). For expression, the *tprC* sequences were sub-cloned into the pET23b+ vector (Life Technologies) between BamHI and HindIII using the primers C-S (5'-cgggatccgatgg gcgtactcactccgca) and C-As (5'-gcaagcttccatgtcactttcattccac). For sub-cloning, the *tprC* ORF was amplified in a 100-µl final volume using 0.4 units of GoTaq polymerase (Promega) with approximately 10 ng of DNA template. $MgCl_2$ and dNTP final concentrations were 1.5 mM and 200 µM, respectively. Initial denaturation and final extension (72°C) were for 10 min each. Denaturation (94°C), annealing (60°C), and extension (72°C) were carried out for 1 min each for

a total of 35 cycles. Amplicons were purified, digested, and ligated into the pET23b+ vector. As a result of cloning into pET23b+, 28 additional amino acids were added to the TprC ORFs (14 NH_2-terminal and 14 COOH-terminal amino acids), including the COOH-terminal 6×His tag for affinity purification. Ligation products were used to transform OneShot TOP10 chemically competent *E. coli* cells (Life Technologies) according to the provided protocol. Transformations were plated on LB-Ampicillin (100 µg/ml) agar plates for selection. For each cloning reaction, individual colonies were screened for the presence of insert-containing plasmids using primers annealing upstream and downstream of the pET23b+ vector poly-linker (T7 promoter and terminator primers). Positive plasmids were extracted from overnight liquid cultures obtained from replica colonies by using the Plasmid Mini kit (Qiagen, Germantown, MD), and two to five clones for each strain were sequenced to ensure sequence fidelity to the previously cloned templates (35). For expression of recombinant antigens, a suitable clone for each *tprC* gene variant was used to transform *E. coli* Rosetta (DE3) competent cells (Life Technologies).

Expression, purification and refolding of recombinant proteins

E. coli cells were grown overnight in LB media supplemented with ampicillin (100 µg/ml). The following day, multiple flasks containing 200 ml of auto-inducing media (56), were inoculated with 20 ml of overnight culture in a 2-liter baffled flask and grown at room temperature for 72 h at 175 rpm in a shaking incubator. Expression of recombinant antigens in induced and un-induced controls was assessed by immunoblot using a monoclonal anti-poly-histidine antibody (Millipore-Sigma, diluted 1:2000) after SDS-PAGE. Prior to purification, presence of the recombinant protein in the soluble and insoluble cellular fractions was

evaluated by SDS-PAGE and immunoblot. Recombinant TprC purification was carried on under denaturing conditions. Briefly, *E. coli* cell pellets were resuspended in 5 ml/g of dry culture weight of binding buffer (50 mM NaH_2PO_4, 10 mM imidazole, pH 8.0) w/o denaturing agent, and the suspension was sonicated in ice with 100 pulses of 6 s each, with each pulse being separated by 10-s intervals. Insoluble components (containing the desired products) were precipitated by centrifugation and resuspended in 5 ml/ g of culture weight of binding buffer (50 mM NaH_2PO_4, 5 mM imidazole, pH 8.0) containing 6M Guanidine-HCl denaturing agent and sonicated again as above. Insoluble components were precipitated again by centrifugation and the supernates were saved. For affinity chromatography, 5.0 ml of nickel-agarose (Ni-NTA agarose, Qiagen) was packaged into a 1.5x14 cm column (Bio-Rad, Carlsbad, CA) and washed with 3 column volumes of molecular-grade H_2O and 6 column volumes of binding buffer + denaturing agent. Cell lysate was then loaded, and the flow was adjusted to 1 ml/min. Unbound proteins were washed using 10 bed volumes of binding buffer, followed by 6 column volumes of wash buffer (50 mM NaH_2PO_4, 20 mM imidazole, pH 8.0) containing denaturing agent. Washing continued until the A_{280} of the flow through was <0.01 AU. Recombinant TprC was eluted with 15 ml of elution buffer (50 mM NaH_2PO_4, 300 mM imidazole, pH 8.0) containing denaturing agent. Eluted fractions devoid of visible contaminants by SDS-PAGE and Coomassie staining were pooled, and protein concentration was assessed by micro-bicinchoninic (BCA) assay (Thermo-Fisher). Pooled fractions were then dialyzed in PBS using a 10 kDa MWCO Slide-A-Lyzer dialysis cassette (Thermo-Fisher) over 12 hours, ensuring PBS change every ~4 hours. Precipitated protein, resulting from elimination of Guanidine-HCl during dialysis was transferred into microcentrifuge tubes and spun down at full speed. After removing the supernate, the

pellet was resuspended in a volume of PBS containing 6M urea suitable to achieve a protein concentration of ~4 mg/ml, and protein concentration was then reassessed using the micro-BCA assay kit (Thermo-Fisher). Prior to immunizations, urea was eliminated using Profoldin (Hudson, MA) M7 renaturing columns for membrane proteins, which were used according to the manufacturer's protocol. M7 renaturing columns were found to provide the best yield when screened along with 19 other conditions offered by Profoldin. Lipid composition of the elute buffer included lysophosphatidylcholine (~5 mM), arginine (~150 mM), glycerol (~10%), dodecyl maltoside (0.7 mM), and Tris-HCl (0.1 mM), pH 7.5). Following buffer exchange, soluble protein concentration was evaluated using micro-BCA assay and analyzed by circular dichroism (CD) to evaluate percentage of β-sheet, alpha-helix, and random coil. CD spectra (190 to 260 nm) were acquired in triplicate at room temperature using 0.5 mg/ml of recombinant refolded TprC in a Jasco-1500 high-performance CD spectrometer. CD spectra were analyzed using the online platform Dichroweb (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) (57) and the spectra from buffer alone for background subtraction.

Rabbit immunization

Groups of three rabbits each were immunized with one of the purified, refolded recombinant TprC variants. Rabbits were injected with 125 µg of refolded protein every 3 weeks for a total of three immunizations. Prior to injection, antigen was mixed with an equal volume of in Titermax Gold Adjuvant (Millipore-Sigma), a water-in oil emulsion containing squalene, the block co-polymer CRL-8300, and a microparticle stabilizers to obtain a final volume of 1 ml. Immunogen-adjuvant preparation was performed according to the

manufacturer`s instruction, and immunizations were performed via four 250 µl injections (each containing 31.25 µg of protein) into 4 intramuscular sites. Three weeks after the last boost, immunized animals were deeply anaesthetized, bled through cardiac puncture, and then euthanized.

ELISA using synthetic peptides

Overlapping synthetic peptides (20-mers overlapping by 10 aa) were designed to represent the sequences of all TprC and TprD/D_2 loci present in each of the seven strains examined in this study starting after the predicted signal peptide (AA 1-22; Fig.1 and Fig.2). Only the C56 peptide and its variants (Table 1), which represent the proteins' COOH-terminus, were synthesized as 26-mers. A total of 120 peptides (Table 1) were produced by Genscript (Piscataway, NJ). Upon receipt, lyophilized peptides were rehydrated in sterile PBS to a stock solution of 200 µg/ml. Solubility of hydrophobic peptides was increased by adding up to 4% (v/v) DMSO per manufacturer's instruction when needed (peptides C1, C4-7, C10, C15-16, C20, C25, C38-C39, C43-44, C53; Table 1). Reconstituted peptides were stored at -20°C until use. For ELISA, peptides were further diluted to 10 µg/ml in PBS, and 50 µl of working dilution (500 ng total) were used to coat the wells of a 96-well Microwell Maxisorp flat-bottom plate (Thermo-Fisher, Waltham, MA) as previously described (42). Absorbance was measured at OD_{405} using a Molecular Devices SpectraMax Plus microplate reader (Molecular Devices, San Jose, CA). A micro-BCA protein assay (Thermo Fisher) was performed in plates coated with Ag and washed to demonstrate that all peptides bound to the well surfaces in the plates (data not shown). For each serum from each group, the value of each replicate experimental wells minus background

505 reactivity (i.e., three times the mean of the wells tested with pooled uninfected rabbit serum)
506 was calculated and plotted. If residual value for the No-antigen control wells was present after
507 subtraction, statistical significance was calculated with one-way ANOVA with the Bonferroni
508 correction of multiple comparisons or t-test, with significance set at $p<0.05$. Except for figures
509 showing cumulative absorbance, graphs represent the mean ± SEM for triplicate wells tested
510 with pooled sera from the 3 rabbits in each group after background subtraction.
511
512 **TprC/D and D2 structure modeling**
513 We used the ColabFold interface (58) to construct Multiple Sequence Alignments (MSA) for the
514 TprC and $TprD_2$ query sequences by searching UniRef30 (59), Mgnify (60) and ColabFold
515 sequence databases with MMSeq2 (61). The MSA was used as input for structure prediction
516 with AlphaFold2 (37) using the default settings (template=False, amber_relax=False, 3 recycles).
517 Visualization was performed using PyMol software (https://pymol.org) (62).
518
519 **TABLES**
520 **Table 1.** Peptides used in this study
521 **Table 2.** Sequence of reactive peptides following rabbit experimental infection
Table 3. Sequence of reactive peptides following rabbit immunization
Table 4. Treponemal strains used in this study
522
523 **FIGURE LEGENDS**
524 **Figure 1.** Alignment of amino acid sequences of the TprC and $TprD/D_2$ variants. *Tp.* subsp.

525 *pallidum* strains (Nichols, MexicoA, Sea81-4, Bal3, and UW249) are indicated in red font on the
526 left of the sequence. The *Tp* subsp. *pertenue* strain (SamoaD) is in green font, and the *Tp* subsp.
527 *endemicum* (IraqB) strain is in blue font. The Chicago and Bal73-1 TprC and TprD sequences (not
528 shown) are identical to the Nichols strain. The MexicoA, Sea8-14, Bal3, UW249, SamoaD, and
529 IraqB strains harbor a $TprD_2$ variant within the *tprD* locus. CSP: predicted cleavable signal
530 peptide; ExL: External Loop. Amino acids encompassing the ExLs predicted by AlphaFold2 are
531 highlighted in red with yellow text only in the top sequence. DVR: Discrete Variable Region.
532 DVRs are highlighted in black along the ruler. *Indicates a DVR found in $TprD_2$ but not TprC and
533 TprD variants.
534
535 **Figure 2.** Predicted structures of TprC and TprD/D_2 using AlphaFold2. **(A)** AlphaFold2 predicts
536 very similar 20-strands β-barrel structures for both TprC and TprD/D2 proteins. TprC is shown
537 on the figure and the 10 putative extracellular loops (ExLs) are color-coded. The model does not
538 include the cleavable signal peptide (CSP; aa 1-22 in Fig.1). Once the predicted structures were
539 superimposed to every available PDB structure by the DALI software (63) to identify structurally
540 similar porins. DALI analysis results and PDB matches are reported in File S1. The highest-
541 scoring structures did not the exact number of β-strands predicted by AlphaFold2 for TprC and
542 $TprD_2$ β-barrels, but slightly higher or slightly lower, but well within the models of integral
543 OMPs with no large periplasmic domains. These results suggest that these Tpr proteins belong
544 to a new family of porins not yet represented in the PDB. **(B)** The predicted structures of TprC
545 and TprD/D_2 are nearly identical in the transmembrane region (backbone root-mean-square
546 deviation, or RMSD, = 0.75) where the estimated per-residue model confidence is very high

(Predicted Local Distance Difference Test, pLDDT > 75). Full pLDDT analysis is reported in File S3 for TprC and $TprD_2$. More differences are seen in the ExL regions, which also show lower pLDDT, suggesting structural flexibility of these loops. **(C)** The DVRs (colored in red) identified by aligning TprC and $TprD_2$ sequences (Fig.1) of different strains localize in the predicted ExL regions.

Figure 3. Reactivity of sera from experimentally infected animals to homologous peptides representing the TprC, TprD and $TprD_2$ variants. **(A)** Reactivity to homologous peptides spanning TprC and TprD proteins of sera from rabbits infected with *Tp* subsp. *pallidum* (Nichols, Chicago, Bal73-1, MexicoA, Sea81-4, Bal3, and UW249B) collected at day 30, 60, and 90 post-infection. Nichols, Chicago, and Bal73-1 sequences are identical. **(B)** Reactivity to homologous peptides spanning TprC variants of immune sera from groups of rabbits infected with *Tp* subsp. *pertenue* (SamoaD) or *Tp* subsp. *endemicum* (IraqB) strains collected at day 30, 60, and 90 post-infection. **(C)** Reactivity to homologous peptides spanning TprD and $TprD_2$ variants of sera collected at day 30, 60, and 90 post-infection from all $TprD_2$-containing *Tp* subspecies and strains studied here. Cumulative Absorbance values are the sum of the mean OD values obtained from all animals in the infection group at all three time points. Boxed peptides contain at least seven amino acids (35% of the peptide length) belonging to a predicted ExL. Strain names on x axis are abbreviated as follows: N: Nichols; M: MexicoA; Sea: Sea81-4; B: Bal3; U: UW249; S: SamoaD; I: IraqB.

Figure 4. Reactivity of sera from experimentally infected animals to homologous and non-

homologous peptides C46, C47, C51, and C55. Humoral reactivity of day-90 sera from experimentally infected animals to homologous and non-homologous TprC peptides. **(A-D)** reactivity to C46, C47, C51, and C55 variants. Strain names on x axes are abbreviated as follows: N: Nichols; M: MexicoA; Sea: Sea81-4; B: Bal3; U: UW249; S: SamoaD; I: IraqB.

Figure 5. Humoral reactivity to TprC peptides following immunization with refolded recombinant full-length TprC antigens. Reactivity to TprC homologous (left panels) and non-homologous peptides (right panels) in sera from rabbits immunized with Nichols **(A, B)**, Bal3 **(C, D)**, and Mexico A **(E, F)** variants of TprC. Asterisk (*) indicates significant reactivity compared to no antigen control. Peptides encompassing sequences predicted to be within ExLs are boxed. Peptide sequence and homology among strains are reported in Table 1. Strain names on x axes are abbreviated as follows: N: Nichols; M: MexicoA; Sea: Sea81-4; B: Bal3; U: UW249; S: SamoaD; I: IraqB.

Figure 6. Comparison of reactivity of sera from infected animals vs. immunized animals. (A-C) Reactivity to peptides following immunization with TprC variants compared to experimental infection. Data shown are means +/- SEM of 3 rabbits per group: 3 weeks post final boost (immunized) and mean +/- SEM of values for days 30, 60, 90 post-infection (infected). Asterisk (*) indicates a significant difference in reactivity compared to the reactivity value following immunization. Peptides encompassing sequences predicted to be within ExLs are boxed. Strain names on x axes are abbreviated as follows: N: Nichols; M: MexicoA; Sea: Sea81-4; B: Bal3; U: UW249; S: SamoaD; I: IraqB.

591

592 **ACKNOWLEDGMENTS**

593 Research reported in this publication was supported by National Institute of Allergy &
594 Infectious Diseases of the National Institutes of Health under award number R01AI042143 grant
595 (to SAL). Tpr models using AlphaFold2 were generated thanks to support from Open
596 Philanthropy (to LG). This work was also partially supported also by the National Institute for
597 Allergy and Infectious Diseases of the National Institutes of Health grant number U19AI144133
598 (Project 2. Project 2 leader: LG; PI: Anna Wald, University of Washington). The content is solely
599 the responsibility of the authors and does not necessarily represent the official views of the
600 Funders. The funders had no role in study design, data collection and interpretation, or the
601 decision to submit the work for publication. The authors are grateful to Janelle Deane for aiding
602 with some of the experimental procedures.

603

604 **REFERENCES**

605 1. Giacani L, Lukehart SA. The endemic treponematoses. *Clin Microbiol Rev* (2014)
606 27(1):89-115. PMID: 24396138.
607 2. Rowley J, Vander Hoorn S, Korenromp E, Low N, Unemo M, Abu-Raddad LJ, et al.
608 Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates,
609 2016. *Bull World Health Organ* (2019) 97(8):548-62p. Epub 2019/08/07. doi:
610 10.2471/blt.18.228486. PMID: 31384073.
611 3. Savage EJ, Marsh K, Duffell S, Ison CA, Zaman A, Hughes G. Rapid increase in gonorrhoea
612 and syphilis diagnoses in England in 2011. *Euro Surveill* (2012) 17(29). PMID: 22835469.

613 4. Savage EJ, Hughes G, Ison C, Lowndes CM. Syphilis and gonorrhoea in men who have sex
614 with men: a European overview. *Euro Surveill* (2009) 14(47). PMID: 19941803.
615 5. Simms I, Fenton KA, Ashton M, Turner KM, Crawley-Boevey EE, Gorton R, et al. The re-
616 emergence of syphilis in the United Kingdom: the new epidemic phases. *Sex Transm Dis* (2005)
617 32(4):220-6. PMID: 15788919.
618 6. Tucker JD, Cohen MS. China's syphilis epidemic: epidemiology, proximate determinants
619 of spread, and control responses. *Curr Opin Infect Dis* (2011) 24(1):50-5. PMID: 21150594.
620 7. CDC. 2018 Sexually Transmitted Disease Surveillance. *Atlanta, GA: US Department of*
621 *Health and Human Services: Centers for Disease Control and Prevention* (2019).
622 8. LaFond RE, Lukehart SA. Biological basis for syphilis. *Clin Microbiol Rev* (2006) 19(1):29-
623 49. PMID: 16418521.
624 9. Goldenberg RL, Thompson C. The infectious origins of stillbirth. *Am J Obstet Gynecol*
625 (2003) 189(3):861-73. PMID: 14526331.
626 10. Moline HR, Smith JF, Jr. The continuing threat of syphilis in pregnancy. *Curr Opin Obstet*
627 *Gynecol* (2016) 28(2):101-4. Epub 2016/02/13. doi: 10.1097/gco.0000000000000258. PMID:
628 26871538.
629 11. CDC. The national plan to eliminate syphilis in the United States. Atlanta, GA: U.S.
630 Department of Health and Human Services: Centers for Disease Control and Prevention, (1999)
631 October 1999.
632 12. WHO. The global elimination of congenital syphilis: rationale and strategy for action.
633 (2007).

634 13. Mitjà O, Marks M, Konan DJ, Ayelo G, Gonzalez-Beiras C, Boua B, et al. Global
635 epidemiology of yaws: a systematic review. *Lancet Glob Health* (2015) 3(6):e324-31. Epub
636 2015/05/24. doi: 10.1016/s2214-109x(15)00011-x. PMID: 26001576.
637 14. Mitjà O, Houinei W, Moses P, Kapa A, Paru R, Hays R, et al. Mass treatment with single-
638 dose azithromycin for yaws. *N Engl J Med* (2015) 372(8):703-10. Epub 2015/02/19. doi:
639 10.1056/NEJMoa1408586. PMID: 25693010.
640 15. Mitjà O, Godornes C, Houinei W, Kapa A, Paru R, Abel H, et al. Re-emergence of yaws
641 after single mass azithromycin treatment followed by targeted treatment: a longitudinal study.
642 *Lancet* (2018) 391(10130):1599-607. Epub 2018/02/13. doi: 10.1016/s0140-6736(18)30204-6.
643 PMID: 29428183.
644 16. Pace JL, Csonka GW. Late endemic syphilis: case report of bejel with gummatous
645 laryngitis. *Genitourin Med* (1988) 64(3):202-4. PMID: 3410469.
646 17. Pace JL. Treponematoses in Arabia. *Saudi Med J* (1983) 4:211-20.
647 18. Julvez J, Michault A, Kerdelhue V. [Serologic studies of non-venereal treponematoses in
648 infants in Niamey, Niger]. *Med Trop (Mars)* (1998) 58(1):38-40. PMID: 9718553.
649 19. Galoo E, Schmoor P. [Identification of a focus of bejel in Mauritania]. *Med Trop (Mars)*
650 (1998) 58(3):311-2. PMID: 10088115.
651 20. Lieberman NAP, Lin MJ, Xie H, Shrestha L, Nguyen T, Huang ML, et al. *Treponema*
652 *pallidum* genome sequencing from six continents reveals variability in vaccine candidate genes
653 and dominance of Nichols clade strains in Madagascar. *PLoS Negl Trop Dis* (2021)
654 15(12):e0010063. Epub 2021/12/23. doi: 10.1371/journal.pntd.0010063. PMID: 34936652.

655 21. Lithgow KV, Cameron CE. Vaccine development for syphilis. *Expert Rev Vaccines* (2017)
656 16(1):37-44. Epub 2016/06/22. doi: 10.1080/14760584.2016.1203262. PMID: 27328030.
657 22. Cameron CE, Lukehart SA. Current status of syphilis vaccine development: need,
658 challenges, prospects. *Vaccine* (2014) 32(14):1602-9. PMID: 24135571.
659 23. Turner TB, Hollander DH. *Biology of the Treponematoses*. Geneva: World Health
660 Organization (1957).
661 24. Miller JN. Immunity in experimental syphilis. VI. Successful vaccination of rabbits with
662 *Treponema pallidum*, Nichols strain, attenuated by g- irradiation. *J Immunol* (1973)
663 110(5):1206-15. PMID: 4572631.
664 25. Koebnik R, Locher KP, Van Gelder P. Structure and function of bacterial outer membrane
665 proteins: barrels in a nutshell. *Mol Microbiol* (2000) 37(2):239-53. Epub 2000/08/10. doi:
666 10.1046/j.1365-2958.2000.01983.x. PMID: 10931321.
667 26. Baker-Zander SA, Lukehart SA. Macrophage-mediated killing of opsonized *Treponema*
668 *pallidum*. *J Infect Dis* (1992) 165(1):69-74. PMID**:** 1727898.
669 27. Lukehart SA, Miller JN. Demonstration of the in vitro phagocytosis of *Treponema*
670 *pallidum* by rabbit peritoneal macrophages. *J Immunol* (1978) 121(5):2014-24. PMID: 361893
671 28. Edmondson DG, Norris SJ. In Vitro Cultivation of the Syphilis Spirochete *Treponema*
672 *pallidum*. *Curr Protoc* (2021) 1(2):e44. Epub 2021/02/19. doi: 10.1002/cpz1.44. PMID:
673 33599121.
674 29. Walker EM, Borenstein LA, Blanco DR, Miller JN, Lovett MA. Analysis of outer membrane
675 ultrastructure of pathogenic *Treponema* and *Borrelia* species by freeze-fracture electron
676 microscopy. *J Bacteriol* (1991) 173(17):5585-8. PMID: 1885536.

677 30. Radolf JD, Norgard MV, Schulz WW. Outer membrane ultrastructure explains the limited
678 antigenicity of virulent *Treponema pallidum*. *Proc Natl Acad Sci USA* (1989) 86(6):2051-5. PMID:
679 2648388.
680 31. Centurion-Lara A, Castro C, Barrett L, Cameron C, Mostowfi M, Van Voorhis WC, et al.
681 *Treponema pallidum* major sheath protein homologue Tpr K is a target of opsonic antibody and
682 the protective immune response. *J Exp Med* (1999) 189(4):647-56. PMID: 9989979.
683 32. Cox DL, Luthra A, Dunham-Ems S, Desrosiers DC, Salazar JC, Caimano MJ, et al. Surface
684 immunolabeling and consensus computational framework to identify candidate rare outer
685 membrane proteins of *Treponema pallidum*. *Infect Immun* (2010) 78(12):5178-94. PMID:
686 20876295.
687 33. Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, et al. Complete
688 genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* (1998)
689 281(5375):375-88. PMID: 9665876.
690 34. Anand A, Luthra A, Dunham-Ems S, Caimano MJ, Karanian C, LeDoyt M, et al. TprC/D
691 (Tp0117/131), a trimeric, pore-forming rare outer membrane protein of *Treponema pallidum*,
692 has a bipartite domain structure. *J Bacteriol* (2012) 194(9):2321-33. PMID: 22389487.
693 35. Centurion-Lara A, Giacani L, Godornes C, Molini BJ, Brinck Reid T, Lukehart SA. Fine
694 Analysis of Genetic Diversity of the *tpr* Gene Family among Treponemal Species, Subspecies and
695 Strains. *PLoS Negl Trop Dis* (2013) 16(7):e2222. PMID: 23696912.
696 36. Centurion-Lara A, Sun ES, Barrett LK, Castro C, Lukehart SA, Van Voorhis WC. Multiple
697 alleles of *Treponema pallidum* repeat gene D in *Treponema pallidum* isolates. *J Bacteriol* (2000)
698 182(8):2332-5. PMID: 10735882.

699 37. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
700 protein structure prediction with AlphaFold2. *Nature* (2021) 596(7873):583-9. Epub
701 2021/07/16. doi: 10.1038/s41586-021-03819-2. PMID: 34265844.
702 38. Geysen HM, Meloen RH, Barteling SJ. Use of peptide synthesis to probe viral antigens
703 for epitopes to a resolution of a single amino acid. *Proc Natl Acad Sci U S A* (1984) 81(13):3998-
704 4002. Epub 1984/07/01. doi: 10.1073/pnas.81.13.3998. PMID: 6204335.
705 39. Haynes AM, Godornes C, Ke W, Giacani L. Evaluation of the Protective Ability of the
706 *Treponema pallidum* subsp. *pallidum* Tp0126 OmpW Homolog in the Rabbit Model of Syphilis.
707 *Infect Immun* (2019) 87(8). Epub 2019/06/12. doi: 10.1128/iai.00323-19. PMID: 31182617.
708 40. Giacani L, Lukehart S, Centurion Lara A. *Syphilis*. Barrett A, L. S, editors: Academic Press
709 (2009).
710 41. Morgan CA, Lukehart SA, Van Voorhis WC. Protection against syphilis correlates with
711 specificity of antibodies to the variable regions of *Treponema pallidum* repeat protein K. *Infect*
712 *Immun* (2003) 71(10):5605-12. PMID: 14500480.
713 42. Morgan CA, Molini BJ, Lukehart SA, Van Voorhis WC. Segregation of B and T cell
714 epitopes of *Treponema pallidum* repeat protein K to variable and conserved regions during
715 experimental syphilis infection. *J Immunol* (2002) 169(2):952-7. PMID: 12097401.
716 43. Morgan CA, Lukehart SA, Van Voorhis WC. Immunization with the N-terminal portion of
717 *Treponema pallidum* repeat protein K attenuates syphilitic lesion development in the rabbit
718 model. *Infect Immun* (2002) 70(12):6811-6. PMID: 12438357.

719 44. Cameron CE, Lukehart SA, Castro C, Molini B, Godornes C, Van Voorhis WC. Opsonic
720 potential, protective capacity, and sequence conservation of the *Treponema pallidum*
721 subspecies *pallidum* Tp92. *J Infect Dis* (2000) 181(4):1401-13. PMID: 10762571.
722 45. Arroll TW, Centurion-Lara A, Lukehart SA, Van Voorhis WC. T-cell responses to
723 *Treponema pallidum* subsp. *pallidum* antigens during the course of experimental syphilis
724 infection. *Infect Immun* (1999) 67(9):4757-63. PMID: 10456928
725 46. Cameron CE, Castro C, Lukehart SA, Van Voorhis WC. Function and protective capacity of
726 *Treponema pallidum* subsp. *pallidum* glycerophosphodiester phosphodiesterase. *Infect Immun*
727 (1998) 66(12):5763-70. PMID: 9826352.
728 47. Sun ES, Molini BJ, Barrett LK, Centurion-Lara A, Lukehart SA, Van Voorhis WC. Subfamily
729 I *Treponema pallidum* repeat protein family: sequence variation and immunity. *Microbes Infect*
730 (2004) 6(8):725-37. PMID: 15207819.
731 48. LaFond RE, Molini BJ, Van Voorhis WC, Lukehart SA. Antigenic variation of TprK V
732 regions abrogates specific antibody binding in syphilis. *Infect Immun* (2006) 74(11):6244-51.
733 PMID: 16923793.
734 49. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein
735 structure and function prediction. *Nat Protoc* (2010) 5(4):725-38. Epub 2010/04/03. doi:
736 10.1038/nprot.2010.5. PMID: 20360767.
737 50. Hegedűs T, Geisler M, Lukács G, Farkas B. AlphaFold2 transmembrane protein structure
738 prediction shines. *bioRxiv* (2021):2021.08.21.457196. doi: 10.1101/2021.08.21.457196.
739 51. Anand A, LeDoyt M, Karanian C, Luthra A, Koszelak-Rosenblum M, Malkowski MG, et al.
740 Bipartite Topology of *Treponema pallidum* Repeat Proteins C/D and I: outer membrane

741 insertion, trimerization, and porin function require a c-terminal β-barrel domain. *J Biol Chem*
742 (2015) 290(19):12313-31. Epub 2015/03/26. doi: 10.1074/jbc.M114.629188. PMID: 25805501.
743 52. Leader BT, Hevner K, Molini BJ, Barrett LK, Van Voorhis WC, Lukehart SA. Antibody
744 responses elicited against the *Treponema pallidum* repeat proteins differ during infection with
745 different isolates of *Treponema pallidum* subsp. *pallidum*. *Infect Immun* (2003) 71(10):6054-7.
746 PMID: 14500480.
747 53. Giacani L, Molini B, Godornes C, Barrett L, Van Voorhis WC, Centurion-Lara A, et al.
748 Quantitative analysis of *tpr* gene expression in *Treponema pallidum* isolates: differences among
749 isolates and correlation with T-cell responsiveness in experimental syphilis. *Infect Immun* (2007)
750 75(1):104-12. PMID: 17030565.
751 54. *Guide for the Care and Use of Laboratory Animals*: The National Academic Press (2011).
752 55. Lukehart SA, Marra CM. Isolation and laboratory maintenance of *Treponema pallidum*.
753 *Curr Protoc Microbiol* (2007) Chapter 12:7:12A.1.1–A.1.8. PMID: 18770607.
754 56. Studier FW. Protein production by auto-induction in high density shaking cultures.
755 *Protein Expr Purif* (2005) 41(1):207-34. PMID: 15915565.
756 57. Miles AJ, Ramalli SG, Wallace BA. DichroWeb, a website for calculating protein
757 secondary structure from circular dichroism spectroscopic data. *Protein Sci* (2021). Epub
758 2021/07/04. doi: 10.1002/pro.4153. PMID: 34216059.
759 58. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold -
760 Making protein folding accessible to all. *bioRxiv* (2021):2021.08.15.456425. doi:
761 10.1101/2021.08.15.456425.

762 59. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. UniRef clusters: a comprehensive
763 and scalable alternative for improving sequence similarity searches. *Bioinformatics* (2015)
764 31(6):926-32. Epub 2014/11/16. doi: 10.1093/bioinformatics/btu739. PMID: 25398609.
765 60. Mitchell AL, Almeida A, Beracochea M, Boland M, Burgin J, Cochrane G, et al. MGnify:
766 the microbiome analysis resource in 2020. *Nucleic Acids Res* (2020) 48(D1):D570-d8. Epub
767 2019/11/07. doi: 10.1093/nar/gkz1035. PMID: 31696235.
768 61. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the
769 analysis of massive data sets. *Nat Biotechnol* (2017) 35(11):1026-8. Epub 2017/10/17. doi:
770 10.1038/nbt.3988. PMID: 29035372.
771 62. Janson G, Zhang C, Prado MG, Paiardini A. PyMod 2.0: improvements in protein
772 sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics* (2017)
773 33(3):444-6. Epub 2017/02/06. doi: 10.1093/bioinformatics/btw638. PMID: 28158668.
774 63. Holm L. Using Dali for Protein Structure Comparison. *Methods Mol Biol* (2020) 2112:29-
775 42. Epub 2020/02/02. doi: 10.1007/978-1-0716-0270-6_3. PMID: 32006276.
776

Table 1. Synthetic peptides used in this study. Amino acids encompassing ExLs (predicted by AlphaFold) are highlighted in red with yellow text.

		Percentage Identity with Peptide by Strain and Allele*																	
		Nichols		Chicago		Bal73-1		MexicoA		Sea81-4		Bal3		UW249		SamoaD		IraqB	
Name	Peptide Sequence	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D
C1	GVLTPQVSGTAQLQWGIAFQ	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C2	AQLQWGIAFQKNPRTGPGKH	100	100	100	100	100	100	100	100	100	100	100	100	100	100	90	90	90	90
C2 (SamoaD/IraqB)	AQLQWGIAFQKNPHTVPGKH	90	90	90	90	90	90	90	90	90	90	90	90	90	90	100	100	100	100
C3	KNPRTGPGKHTHGFRTTNSL	100	100	100	100	100	100	100	100	100	100	100	100	100	100	90	90	90	90
C3 (SamoaD/IraqB)	KNPHTVPGKHTHGFRTTNSL	90	90	90	90	90	90	90	90	90	90	90	90	90	90	100	100	100	100
C4	THGFRTTNSLTISLPLVSKH	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C5	TISLPLVSKHTHTRRGEARS	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C6	THTRRGEARSGVWAQLQLKD	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C7	GVWAQLQLKDLAVELASSKS	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C8	LAVELASSKSSTALSFTKPT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C9	STALSFTKPTASFQATLHCY	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C10	ASFQATLHCYGAYLTVGTSP	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C11	GAYLTVGTSPSCVVNFAQLW	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C12	SCVVNFAQLWKPFVTRAYSE	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C13	KPFVTRAYSEKDTRYAPGFS	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C14	KDTRYAPGFSGSGAKLGYQA	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C15	GSGAKLGYQAHNVGNSGVDV	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C16	HNVGNSGVDVDIGFLSFLSN	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C17	DIGFLSFLSNGAWDSTDTTH	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C18	GAWDSTDTTHSKYGFGADAT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C19	SKYGFGADATLSYGVDRQRL	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C20	LSYGVDRQRLLTLELAGNAT	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C21	LTLELAGNATLDQNYVKGTE	100	100	100	100	100	100	85	100	100	100	100	100	100	100	85	85	90	90
C21 (MexicoA/SamoaD)	LTLELAGNATLEQHYRKGTE	85	85	85	85	85	85	100	85	85	85	85	85	85	85	100	100	90	90
C21 (IraqB)	LTLELAGNATLEQNYLKGTE	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	100	100
C22	LDQNYVKGTEDSKNENKTAL	100	100	100	100	100	100	80	100	100	100	100	100	100	100	80	80	85	85
C22 (MexicoA/SamoaD)	LEQHYRKGTEDSTNENKTAL	80	80	80	80	80	80	100	80	80	80	80	80	80	80	100	100	80	80
C22 (IraqB)	LEQNYLKGTEDPKNENKTAL	85	85	85	85	85	85	80	85	85	85	85	85	85	85	80	80	100	100
C23	DSKNENKTALLWGVGGRLTL	100	100	100	100	100	100	95	100	100	100	100	100	100	100	95	95	95	95
C23 (MexicoA/SamoaD)	DSTNENKTALLWGVGGRLTL	95	95	95	95	95	95	100	95	95	95	95	95	95	95	100	100	90	90

		Percentage Identity with Peptide by Strain and Allele																	
		Nichols		Chicago		Bal73-1		MexicoA		Sea81-4		Bal3		UW249		SamoaD		IraqB	
Name	Peptide Sequence	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D
C23 (IraqB)	DPKNENKTALLWGVGGRLTL	95	95	95	95	95	95	90	95	95	95	95	95	95	95	90	90	100	100
C24	LWGVGGRLTLEPGAGFRFSF	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C25	EPGAGFRFSFALDAGNQHQS	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C26	ALDAGNQHQSNAHAQTQERA	100	100	100	100	100	100	100	60	100	60	100	60	100	60	85	50	85	50
C26 (IraqB)	ALDAGNQHQSNADAQTQEER	85	85	85	85	85	85	85	60	85	60	85	60	85	60	95	50	100	50
C26 (SamoaD)	ALDAGNQHQSNAHAQTQKER	80	80	80	80	80	80	80	60	80	60	80	60	80	60	100	50	95	50
C27	NAHAQTQERAILKAREVFRR	100	100	100	100	100	100	100	20	100	20	100	20	100	10	40	20	55	20
C27 (IraqB)	NADAQTQEERVSLAGEVFGQ	55	55	55	55	55	55	55	20	55	20	55	20	55	20	90	10	100	10
C27 (SamoaD)	NAHAQTQKERVSLAGEVFGQ	65	50	65	65	50	65	65	10	65	10	65	10	65	10	100	10	90	10
C28	ILKAREVFRRVEGKLVQNLP	100	100	100	100	100	100	100	15	100	15	100	15	100	15	65	15	65	15
C28 (SamoaD/IraqB)	VSLAGEVFGQVVGKLVQNLP	65	65	65	65	65	65	65	15	65	15	65	15	65	15	100	15	100	15
C29	VEGKLVQNLPNIMMPPGITE	100	100	100	100	100	100	100	5	100	5	100	5	100	5	90	5	90	5
C29 (SamoaD/IraqB)	VVGKLVQNLPNIMMPLGITE	90	90	90	90	90	90	90	5	90	5	90	5	90	5	100	5	100	5
C30	NIMMPPGITEQTTLIEMVGL	100	100	100	100	100	100	100	20	100	20	100	20	100	20	95	20	95	20
C31	QTTLIEMVGLAALIAEGTLG	100	100	100	100	100	100	100	15	100	15	100	15	100	15	95	15	95	15
C32	AALIAEGTLGSAIQTVLAAG	100	100	100	100	100	100	100	15	100	15	100	15	100	15	100	15	100	15
C33	SAIQTVLAAGALAALVSQLV	100	100	100	100	100	100	100	10	100	10	100	10	100	10	100	10	100	10
C34	ALAALVSQLVPNIEQGVRDV	100	100	100	100	100	100	100	5	100	5	100	5	100	5	95	5	95	5
C34 (SamoaD/IraqB)	ALAALVSQLVPHIEQGVRDV	95	95	95	95	95	95	95	5	95	5	95	5	95	5	100	5	100	5
C35	PNIEQGVRDVFRSSDPRVVT	100	100	100	100	100	100	100	20	100	20	100	20	100	20	95	20	95	20
C35 (SamoaD/IraqB)	PHIEQGVRDVFRSSDPRVVT	95	95	95	95	95	95	95	20	95	20	95	20	95	20	100	20	100	20
C36	FRSSDPRVVTAKLLAFLERA	100	100	100	100	100	100	100	25	100	25	100	25	100	25	100	25	100	25
C37	AKLLAFLERAPMNALNIDAL	100	100	100	100	100	100	100	40	100	40	100	40	100	40	100	40	100	40
C38	PMNALNIDALLRMQWKWLSS	100	100	100	100	100	100	100	75	100	75	100	75	100	75	100	70	100	80
C39	LRMQWKWLSSGIYFATAGTN	100	100	100	100	100	100	100	95	100	95	100	95	100	95	100	90	100	100
C40	GIYFATAGTNIFGKRVFATT	100	100	100	100	100	100	100	80	100	80	100	80	100	80	100	80	100	80
C41	IFGKRVFATTRAHYFDFAGF	100	100	100	100	100	100	100	60	100	60	100	60	100	60	100	70	100	70
C42	RAHYFDFAGFLKLETKSGDP	100	100	100	100	100	100	100	80	100	80	100	80	100	80	100	80	100	80
C43	LKLETKSGDPYTHLLTGLNA	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

		Percentage Identity with Peptide by Strain and Allele																	
		Nichols		Chicago		Bal73-1		MexicoA		Sea81-4		Bal3		UW249		SamoaD		IraqB	
Name	Peptide Sequence	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D
C44	YTHLLTGLNAGVEARVYIPL	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C45 (Nichols)	GVEARVYIPLTYIRYRNNGG	100	100	100	100	100	100	90	95	90	95	90	95	90	90	90	95	90	95
C45 (MexicoA/Sea814/ Bal3/UW249)	GVEARVYIPLTYVFYRNNGG	90	90	90	90	90	90	100	90	100	90	100	90	100	90	100	90	100	90
C45 (SamoaD/IraqB)	GVEARVYIPLTYVFYKNNGG	90	90	90	90	90	90	100	95	100	95	95	95	100	95	100	95	100	95
C46 (Nichols)	TYIRYRNNGGYELNGAVPPG	100	100	100	100	100	100	75	50	75	50	75	50	75	50	70	45	80	45
C46 (MexicoA/Sea814/ Bal3/UW249)	TYVFYRNNGGYELNRVVPSG	75	75	75	75	75	75	100	45	100	45	100	45	100	45	85	45	85	45
C46 (IraqB)	TYVFYKNNGGYELNGVVPPG	80	80	80	80	80	80	100	45	100	45	95	45	100	45	90	50	100	50
C46 (SamoaD)	TYVFYKNNGGYPLNGVVPSG	70	70	70	70	70	70	80	50	80	50	80	50	85	50	100	50	90	50
C47 (Nichols)	YELNGAVPPGTINMPILGKA	100	100	100	100	100	100	80	45	80	45	80	45	80	45	85	45	90	45
C47 (MexicoA/Sea814/ Bal3/UW249)	YELNRVVPSGIINMPILGKA	80	80	80	80	80	80	100	40	100	40	100	40	100	40	85	45	90	45
C47 (IraqB)	YELNGVVPPGIINMPILGKA	75	75	75	75	75	75	85	50	85	50	85	50	85	50	85	50	100	50
C47 (SamoaD)	YPLNGVVPSGTINMPILGKA	85	85	85	85	85	85	85	50	85	50	85	50	85	50	100	50	85	50
C48 (Nichols/SamoaD)	TINMPILGKAWCSYRIPLGS	100	100	100	100	100	100	95	85	95	85	95	85	95	85	100	85	95	85
C48 (MexicoA/Sea814/ Bal3/UW249/IraqB)	IINMPILGKAWCSYRIPLGS	95	95	95	95	95	95	100	85	100	85	100	85	100	85	95	85	100	85
C49	WCSYRIPLGSHAWLAPHTSV	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C50	HAWLAPHTSVLGTTNRFNII	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C51 (Nichols)	LGTTNRFNIINPAGNLLNER	100	100	100	100	100	100	95	45	95	45	95	45	95	45	90	40	90	40
C51 (MexicoA/ Sea81-4/Bal3/UW249)	LGTTNRFNIINAAGNLLNER	95	95	95	95	95	95	100	45	100	45	100	45	100	45	95	45	95	45
C51 (SamoaD/IraqB)	LGTTNRFNIINAAGNLVNER	90	90	90	90	90	90	95	45	95	45	95	45	95	45	100	40	100	40
C52 (Nichols)	NPAGNLLNERALQYQVGLTF	100	100	100	100	100	100	95	65	95	65	95	65	95	65	90	60	90	60
C52 (MexicoA/ Sea81-4/Bal3/UW249)	NAAGNLLNERALQYQVGLTF	95	95	95	95	95	95	100	65	100	65	100	65	100	65	95	60	95	60
C52 (SamoaD/IraqB)	NAAGNLVNERALQYQVGLTF	90	90	90	90	90	90	95	60	95	60	95	60	95	60	100	60	100	60
C53	ALQYQVGLTFSPFEKVELSA	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
C54 (Nichols/MexicoA/ UW249)	SPFEKVELSAQWEQGVLADA	100	100	100	100	100	100	100	100	90	100	90	100	100	100	90	90	90	90
C54 (Sea81-4/Bal3/ SamoaD/IraqB)	SPFEKVELSAQWEQGVLSDV	90	90	90	90	90	90	90	90	100	90	100	90	90	90	100	100	100	100
C55 (Nichols)	QWEQGVLADAPYMGIAESIW	100	100	100	100	100	100	85	100	90	100	90	100	85	100	80	90	90	90
C55 (MexicoA/UW249)	QWEQGVLADAPYMGITQSIG	90	90	90	90	90	90	100	85	80	85	80	85	100	85	90	90	80	90

		Percentage Identity with Peptide by Strain and Allele																	
		Nichols		Chicago		Bal73-1		MexicoA		Sea81-4		Bal3		UW249		SamoaD		IraqB	
Name	Peptide Sequence	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D	Tpr C	Tpr D
C55 (Sea81-4/Bal3/IraqB)	QWEQGVLSDVPYMGIAESIW	85	85	85	85	85	85	75	85	100	90	100	90	75	90	85	100	100	100
C55 (SamoaD)	QWEQGVLSDVPYMGITQSIW	80	80	80	80	80	80	85	90	85	75	85	75	85	80	100	90	90	90
C56 (Nichols/Sea81-4/Bal3/IraqB)	PYMGIAESIWSERHFGTLVCGMKVTW	100	100	100	100	100	100	85	100	100	100	100	100	85	100	88	100	100	100
C56 (MexicoA/UW249)	PYMGITQSIGSDRHFGTLVCGMKVTW	85	85	85	85	85	85	100	85	85	85	85	85	100	85	92	85	85	85
C56 (SamoaD)	PYMGITQSIWSERHFGTFVCGMKVTW	88	88	88	88	88	88	88	88	88	88	88	88	88	88	100	88	88	88
D26	ALDAGNQHQSNAQFYARMAP	60	60	60	60	60	60	60	100	60	100	60	100	60	100	50	50	50	50
D26 (IraqB)	ALDAGNQHQSDTKFYFRMAP	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	100	50	100
D27	QFYARMAPSQRVHEVITSLG	10	10	10	10	10	10	10	100	10	100	10	100	10	100	10	85	10	85
D27 (IraqB)	KFYFRMAPSQRVHEVITSLG	10	10	10	10	10	10	10	85	10	85	10	85	10	85	10	95	10	100
D28	RVHEVITSLGDTLLTSPQQD	15	15	15	15	15	15	15	100	15	100	15	100	15	100	15	95	15	95
D29	DTLLTSPQQDVVSFFVQELS	10	10	10	10	10	10	10	100	10	100	10	100	10	100	10	100	10	100
D30	VVSFFVQELSKGSLLEKAGL	25	25	25	25	25	25	25	100	25	100	25	100	25	100	25	100	25	100
D31	KGSLLEKAGLVTLLAQRTIV	35	35	35	35	35	35	35	100	35	100	35	100	35	100	35	100	35	100
D32	VTLLAQRTIVGLASSGGYLR	20	20	20	20	20	20	20	100	20	100	20	100	20	100	20	100	20	100
D33	GLASSGGYLRHLNGKGLEIN	10	10	10	10	10	10	10	100	10	100	10	100	10	100	10	100	10	100
D34	HLNGKGLEINMRLIEQQKNP	5	5	5	5	5	5	5	100	5	100	5	100	5	100	5	100	5	100
D35	MRLIEQQKNPDARMRTALFI	20	20	20	20	20	20	20	100	20	100	20	100	20	100	20	100	20	100
D36	DARMRTALFISWLQFTYTKT	25	25	25	25	25	25	25	100	25	100	25	100	25	100	25	100	25	100
D37	ALFISSWLQFTYTKTLNIDAL	10	10	10	10	10	10	10	100	10	100	10	100	10	100	10	100	10	100
D38	YTKTLNIDALLRMQWRWLSS	75	75	75	75	75	75	75	80	75	80	75	80	75	80	80	80	75	75
D39	LRMQWRWLSSGIYFATAGTN	95	95	95	95	95	95	95	100	95	100	95	100	95	100	95	95	95	95
D40	GIYFATAGTNIFGERVFFKN	80	80	80	80	80	80	80	100	80	100	80	100	80	100	80	100	80	100
D41	IFGERVFFKNQADHFDFAGF	65	65	65	65	65	65	65	100	65	100	65	100	65	100	65	100	65	100
D42	QADHFDFAGFLKLETKSGDP	80	80	80	80	80	80	80	100	80	100	80	100	80	100	80	100	80	100
D45	GVEARVYIPLTYIFYINNGG	90	90	90	90	90	90	90	100	90	100	90	100	90	100	90	100	90	100
D46	TYIFYINNGGAQYKGSNSDG	45	45	45	45	45	45	45	100	45	100	45	100	45	100	45	100	45	100
D47	AQYKGSNSDGVINTPILSKA	50	50	50	50	50	50	50	100	50	100	50	100	50	100	50	100	50	100
D48	VINTPILSKAWCSYRIPLGS	90	90	90	90	90	90	90	100	90	100	90	100	90	100	90	100	90	100
D50	HAWLAPHTSVLWATNRFNHN	80	80	80	80	80	80	80	100	80	100	80	100	80	100	80	100	80	100
D51	LWATNRFNHNQSGDALLREH	45	45	45	45	45	45	45	100	45	100	45	100	45	100	40	100	40	100

D52	QSGDALLREHALQYQVGLTF	65	65	65	65	65	65	65	100	65	100	65	100	65	100	60	100	60	100

*In Figure 2-5, When necessary, strain names are abbreviated as follows: N: Nichols; M: MexicoA; Sea: Sea81-4; B: Bal3; U: UW249; S: SamoaD; I: IraqB.

"C" indicates peptides mapping to conserved portions of the TprC/D/D_2 proteins. "D" peptides map to TprD_2 central region, which greatly differ from TprC/D variants. Peptides are numbered sequentially (e.g. C1 shares the last 10 aa with C2, C2 shares the last 10 aa with C3 and so forth). If multiple peptides share the same name (e.g. C47) the strain name (Table 1) or abbreviation (Figures) follows the peptide number.

Table 2. Sequence of reactive peptides identified by infected-rabbit sera

Peptide or peptide range	Experimentally determined Epitope-containing sequence	Location per AlphaFold	B-cell epitope predicted by		
			IEDB	BCpreds	BepiPred2.0
Sequences of reactive peptides based on Fig.2A (subspecies *pallidum*)					
C1-C3	GVLTPQVSGTAQLQWGIAFQ KNPRTGPGKHTHGFRTTNSL	Scaffold (C1) and ExL1 (C2-3)	X	X	X
C6	THTRRGEARSGVWAQLQLKD	Scaffold	X	X	X
C13-C14	KPFVTRAYSEKDTRYAPGFSGSGAKLGYQA	ExL3	X	X	
C18	GAWDSTDTTHSKYGFGADAT	ExL4	X	X	X
C20	LSYGVDRQRLLTLELAGNAT	Scaffold	X		
C25-C29	EPGAGFRFSFALDAGNQHQS NAHAQTQERAILKAREVFRR VEGKLVQNLPNIMMPPGITE	Scaffold (C25) and ExL6 (C26-C29)	X	X	X
C39	LRMQWKWLSSGIYFATAGTN	Scaffold	None		
C43	LKLETKSGDPYTHLLTGLNA	Scaffold	X	X	X
N–C46-47	TYIRYRNNGGYELNGAVPPGTINMPILGKA	ExL8	X		
C50-C51	HAWLAPHTSVLGTTNRFNIINAAGNLLNER	ExL9	X	X	X
C53-C55	ALQYQVGLTFSPFEKVELSA QWEQGVL**S/A**D**V/A**PYMGIAESIW (N) or QWEQGVLSDVPYMGIAESIW (Sea81-4/Bal3)*	Scaffold (C53) and ExL10 (C54-C55)	X	X	X
Sequences of reactive peptides based on Fig.2B (subspecies *pertenue* and *endemicum*)					
C1	GVLTPQVSGTAQLQWGIAFQ	Scaffold	None		
C6	THTRRGEARSGVWAQLQLKD	Scaffold		X	X
C13-C14	KPFVTRAYSEKDTRYAPGFSGSGAKLGYQA	ExL3	X	X	X
C18	GAWDSTDTTHSKYGFGADAT	ExL4	X	X	
C20	LSYGVDRQRLLTLELAGNAT	Scaffold	X		
S – C22	LEQHYRKGTEDSTNENKTAL	ExL5	X	X	
C25	EPGAGFRFSFALDAGNQHQS	Scaffold		X	
C33	SAIQTVLAAGALAALVSQLV	ExL6	X		
C36	FRSSDPRVVTAKLLAFLERA	ExL6	None		
C43	LKLETKSGDPYTHLLTGLNA	Scaffold	X	X	
S – C46	TYVFYKNNGGYPLNGVVPSG	ExL8	X	X	X
I – C46	TYVFYKNNGGYELNGVVPPG	ExL8	X	X	X
S – C47	YPLNGVVPSGTINMPILGKA	ExL8	X	X	X
I – C47	YELNGVVPPGIINMPILGKA	ExL8	X	X	X
C51	LGTTNRFNIINAAGNLLNER	ExL9	X	X	X
C53	ALQYQVGLTFSPFEKVELSA	Scaffold	None		
S – C54	SPFEKVELSAQWEQGVLSDV	ExL10	X	X	X
I – C55	QWEQGVLSDVPYMGIAESIW	ExL10	X	X	X
Sequences of reactive peptides based on Fig.2C					
D34-D36	HLNGKGLEINMRLIEQQKNP DARMRTALFISWLQFTYTKT	ExL6	X	X	X
I – C39	LRMQWKWLSSGIYFATAGTN	Scaffold	None		

C43	LKLETKSGDPYTHLLTGLNA	Scaffold	X	X	X
D46-D47	TYIFYINNGGAQYKGSNSDGVINTPILSKA	ExL8	X	X	X
D49	WCSYRIPLGSHAWLAPHTSV	Scaffold	None		
D51	LWATNRFNHNQSGDALLREH	ExL9	X	X	X

Lightly shaded peptides are in the NH_2-terminal portion of the protein.

*S and V in the Sea 81-4/Bal3 strain, A and A in Nichols

Table 3. Sequences of reactive peptides following rabbit immunization

Sequences of reactive peptides based on results shown in Fig.4A. Immunization with TprC full-length Nichols variant		
Peptide or peptide range	**Experimentally determined Epitope-containing sequences**	**Location per AlphaFold**
C1-C3	GVLTPQVSGTAQLQWGIAFQ KNPRTGPGKHTHGFRTTNSL	Scaffold (C1) and ExL1 (C2-C3)
C5-C6	TISLPLVSKHTHTRRGEARS THTRRGEARSGVWAQLQLKD	Scaffold
C9[1]	**STALSFTKPTASFQATLHCY**	ExL2/Scaffold
C16-18	HNVGNSGVDVDIGFLSFLSN GAWDSTDTTHSKYGFGADAT	Scaffold (c16-17)/ExL4 (C17-18)/Scaffold (C18)
C28	ILKAREVFRRVEGKLVQNLP	ExL6
C32	AALIAEGTLGSAIQTVLAAG	ExL6
C35	PNIEQGVRDVFRSSDPRVVT	ExL6
N/Sea/U - C47	YELNGAVPPGTINMPILGKA	ExL8
C53	ALQYQVGLTFSPFEKVELSA	Scaffold
C55	QWEQGVLSDVPYMGIAESIW	ExL10
Sequences of reactive peptides based on results shown in Fig.4B. Immunization with TprC full-length Bal3 variant		
C1-C3	GVLTPQVSGTAQLQWGIAFQ KNPRTGPGKHTHGFRTTNSL	Scaffold (C1) and ExL1 (C2-C3)
C6-C7	THTRRGEARSGVWAQLQLKDLAVELASSKS	Scaffold (C6) and ExL2 (C7)
C13-C14	KPFVTRAYSEKDTRYAPGFSGSGAKLGYQA	ExL3
C16[1]-C18	**HNVGNSGVDVDIGFLSFLSN** GAWDSTDTTHSKYGFGADAT	Scaffold (C16) and ExL4 (C17-C18)
C20	LSYGVDRQRLLTLELAGNAT	Scaffold
C43	LKLETKSGDPYTHLLTGLNA	Scaffold
C47	YELNRVVPSGIINMPILGKA	ExL8
C49	WCSYRIPLGSHAWLAPHTSV	Scaffold
C51	LGTTNRFNIINAAGNLLNER	ExL9
Sequences of reactive peptides based on results shown in Fig.4C. Immunization with TprC full-length MexicoA variant		
C1-C3	GVLTPQVSGTAQLQWGIAFQ KNPRTGPGKHTHGFRTTNSL	Scaffold (C1) and ExL1 (C2-C3)
C5-C6	TISLPLVSKHTHTRRGEARS THTRRGEARSGVWAQLQLKD	Scaffold
C9[1]	**STALSFTKPTASFQATLHCY**	Scaffold/ExL2
C28	ILKAREVFRRVEGKLVQNLP	ExL6

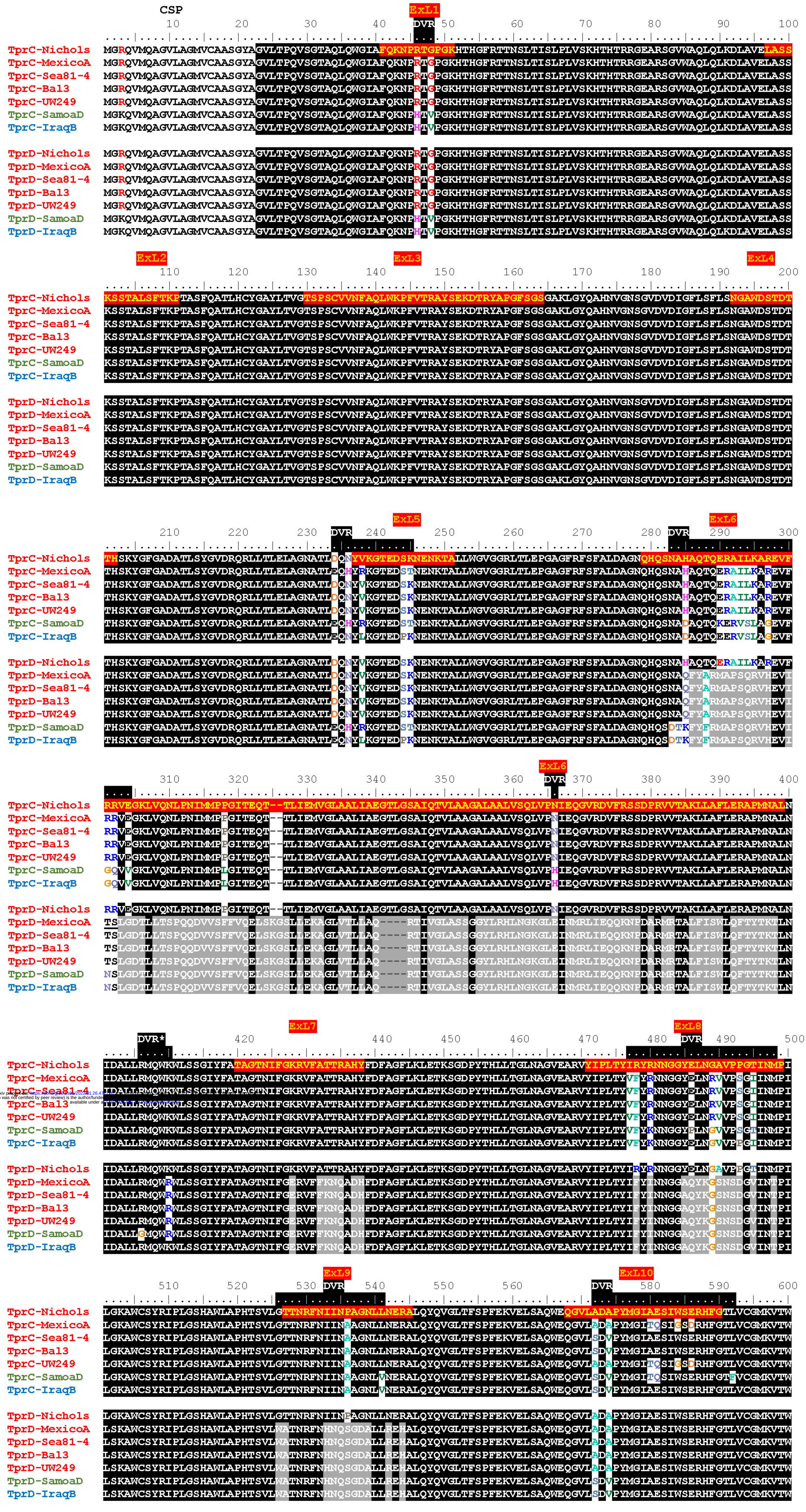
[1] C9/C16 peptides were not predicted to harbor B-cell epitopes by IEDB, BCpreds and bepiPred2.0

Table 4. Treponemal strains used in this study

Species, subspecies	Strain name	Source	Location	Year of isolation
Tp. pallidum	Nichols[a]	Cerebrospinal fluid	Washington DC	1912
	Sea81-4[b]	Primary chancre	Seattle	1980
	Bal3[c]	Blood, congenital	Baltimore	Unknown
	MexicoA[c]	Primary chancre	Mexico	1953
	Street14[d]	Skin	Atlanta	1977
	Bal73-1[c]	Aqueous humor, congenital	Baltimore	1973
	UW249C[e]	Cerebrospinal fluid	Seattle	2004
Tp. endemicum	IraqB[c]	Oral mucous patches	Iraq	1951
Tp. pertenue	SamoaD[c]	Skin lesion	Western Samoa	1953

[a] Originally provided by James N. Miller, University of California, Los Angeles, CA.
[b] Strain isolated in Seattle by Sheila A. Lukehart, University of Washington, Seattle, WA.
[c] Strains provided by Paul Hardy and Ellen Nell, Johns Hopkins University, Baltimore, MD.
[d] Provided by Sandra A. Larsen, Center for Disease Control and Prevention, Atlanta, GA.
[e] Provided by Christina Marra, University of Washington, Seattle, WA.

A


CSP
ExL1
DVR
ExL2
ExL3
ExL4
ExL5
ExL6
DVR
ExL6
DVR
ExL7
ExL8
DVR*
DVR
ExL9
DVR
ExL10
DVR
TprC-Nichols
TprC-MexicoA
TprC-Sea81-4
TprC-Bal3
TprC-UW249
TprC-SamoaD
TprC-IraqB
TprD-Nichols
TprD-MexicoA
TprD-Sea81-4
TprD-Bal3
TprD-UW249
TprD-SamoaD
TprD-IraqB

A

Membrane

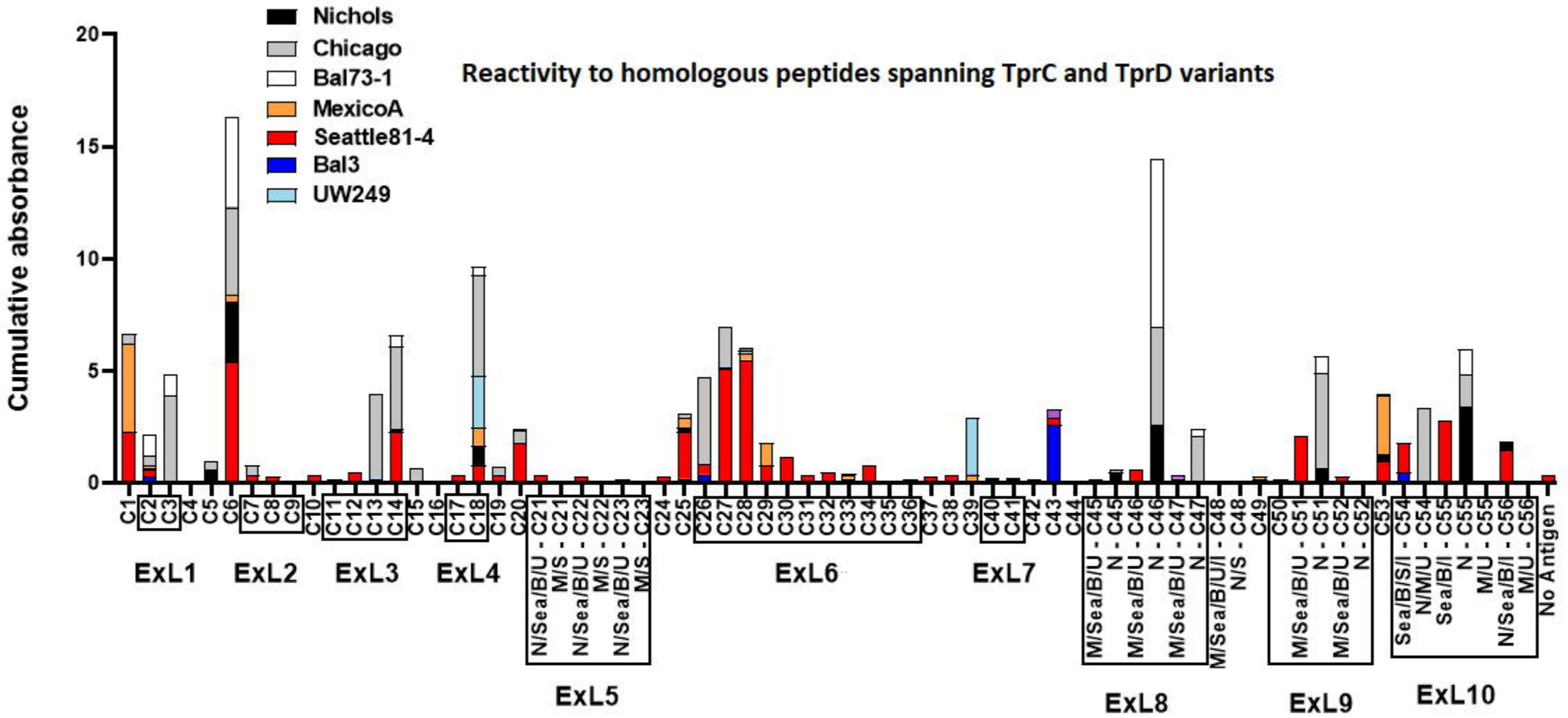
ExL5
ExL6
ExL7
ExL4
ExL8
ExL3
ExL9
ExL10
ExL2
ExL1

B

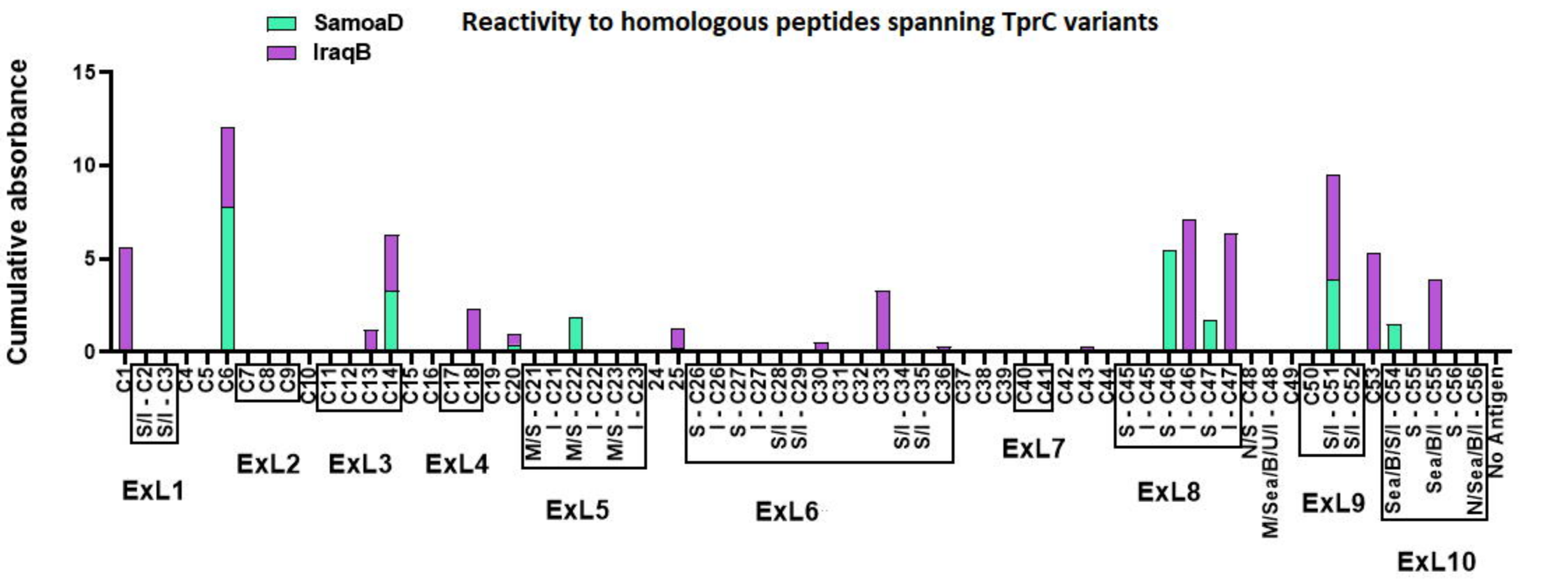
100
pLDDT
30

C

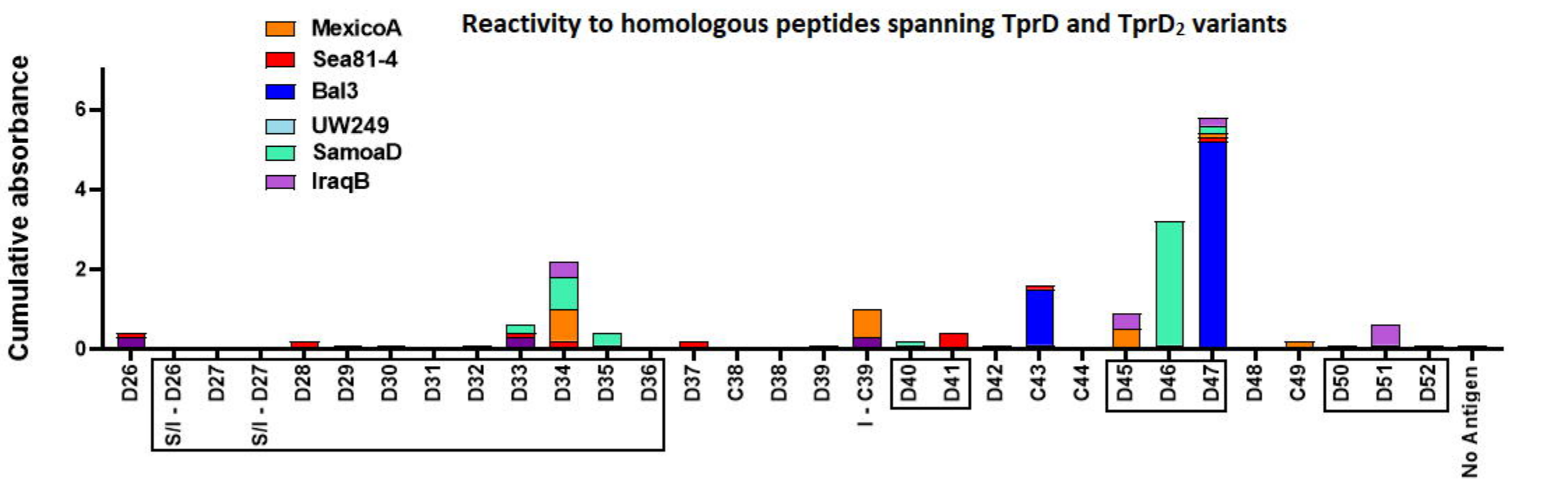
A



B

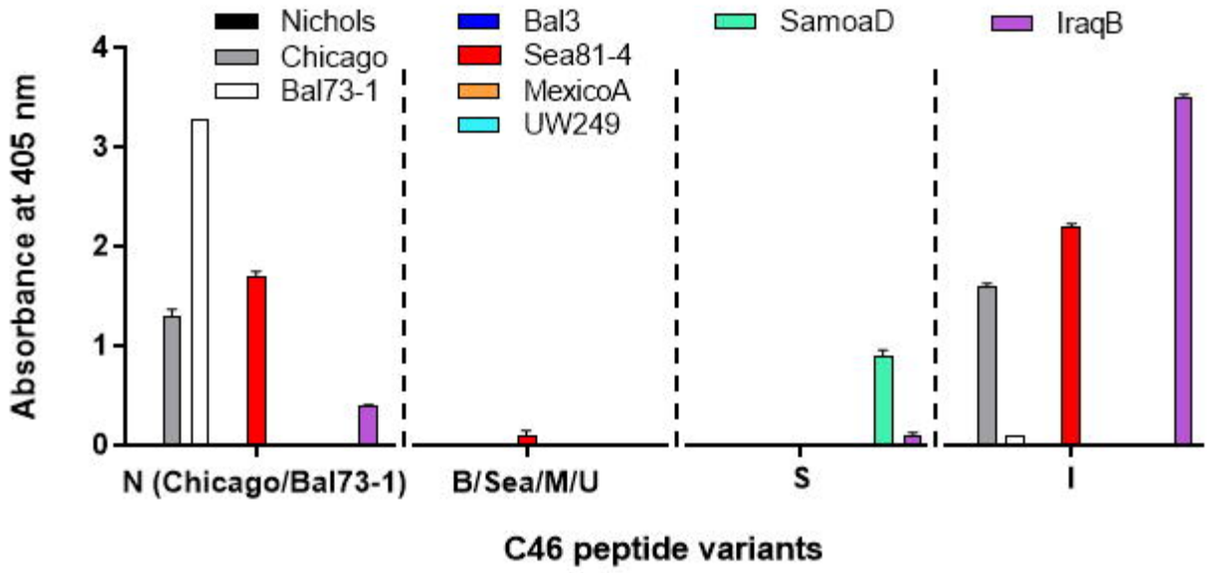


C



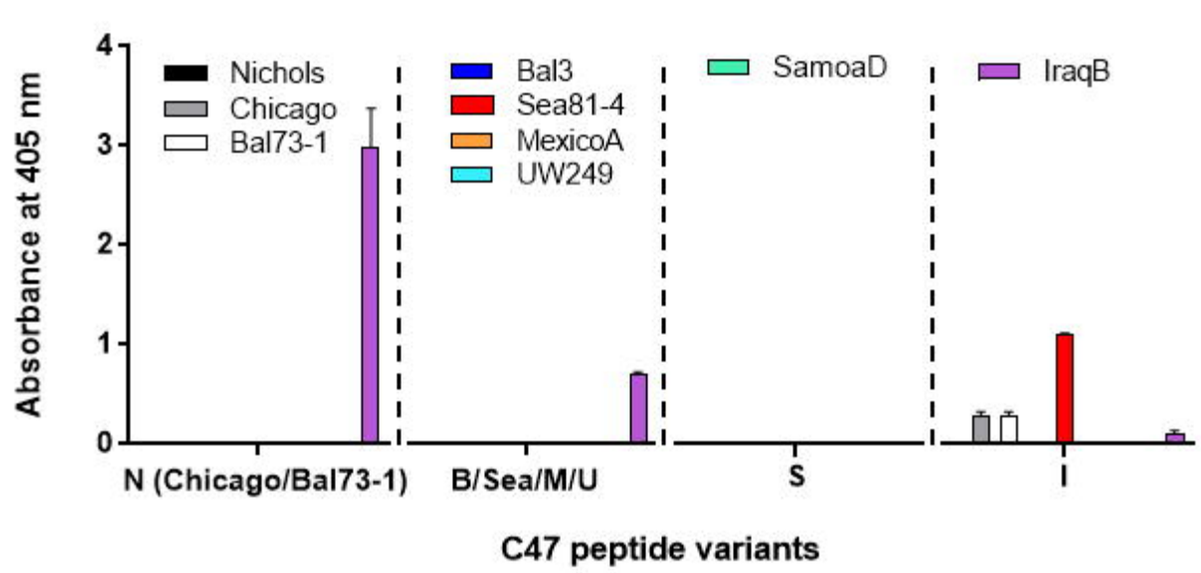
Peptide Variant Legend
B: Bal3; **I**: IraqB; **M**: MexicoA; **N**: Nichols, Chicago, Bal73-1;
S: SamoaD, **Sea**: Seattle81-4; **U**: UW249

A



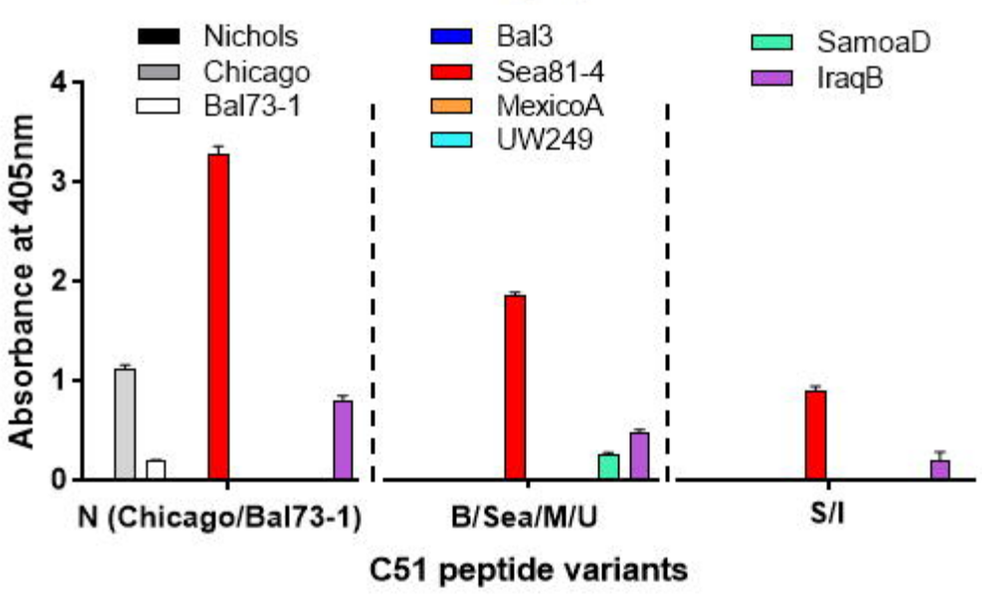
Nichols/Chicago/Bal73-1 C46	TY I RYRNNGGYELNGAVPPG
Bal3/Seattle81-4/MexicoA/UW249 C46	TYVFYRNNGGYELNRVVPSG
SamoaD C46	TYVFYKNNGGYPLNGVVPSG
IraqB C46	TYVFYKNNGGYELNGVVPPG

B



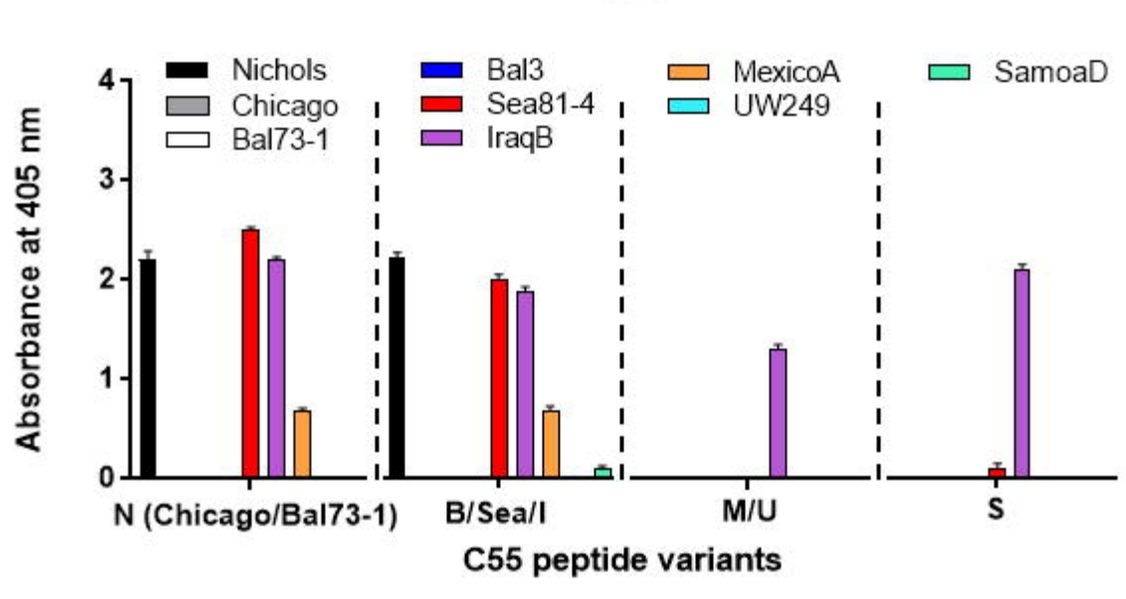
Nichols/Chicago/Bal73-1 C47	YELNGAVPPGTINMPILGKA
Bal3/Seattle81-4/MexicoA/UW249 C47	YELNRVVPSG I INMPILGKA
SamoaD C47	YPLNGVVPSGTINMPILGKA
IraqB C47	YELNGVVPPG I INMPILGKA

C



Bal3/Seattle81-4/MexicoA/UW249 C51	LGTTNRFNIINAAGNLLNER
IraqB/SamoaD C51	LGTTNRFNIINAAGNLVNER
Nichols/Chicago/Bal73-1 C51	LGTTNRFNIINPAGNLLNER

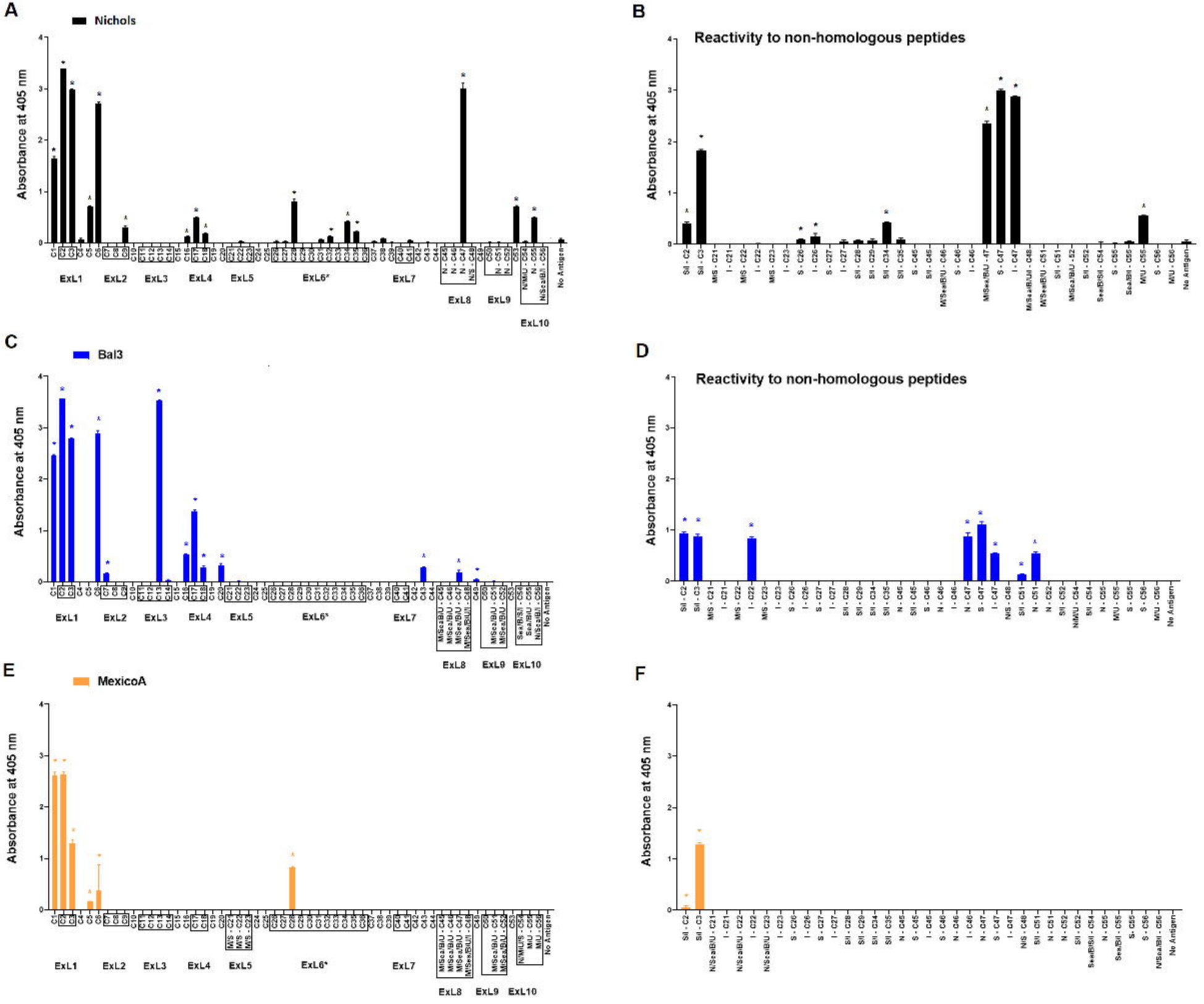
D



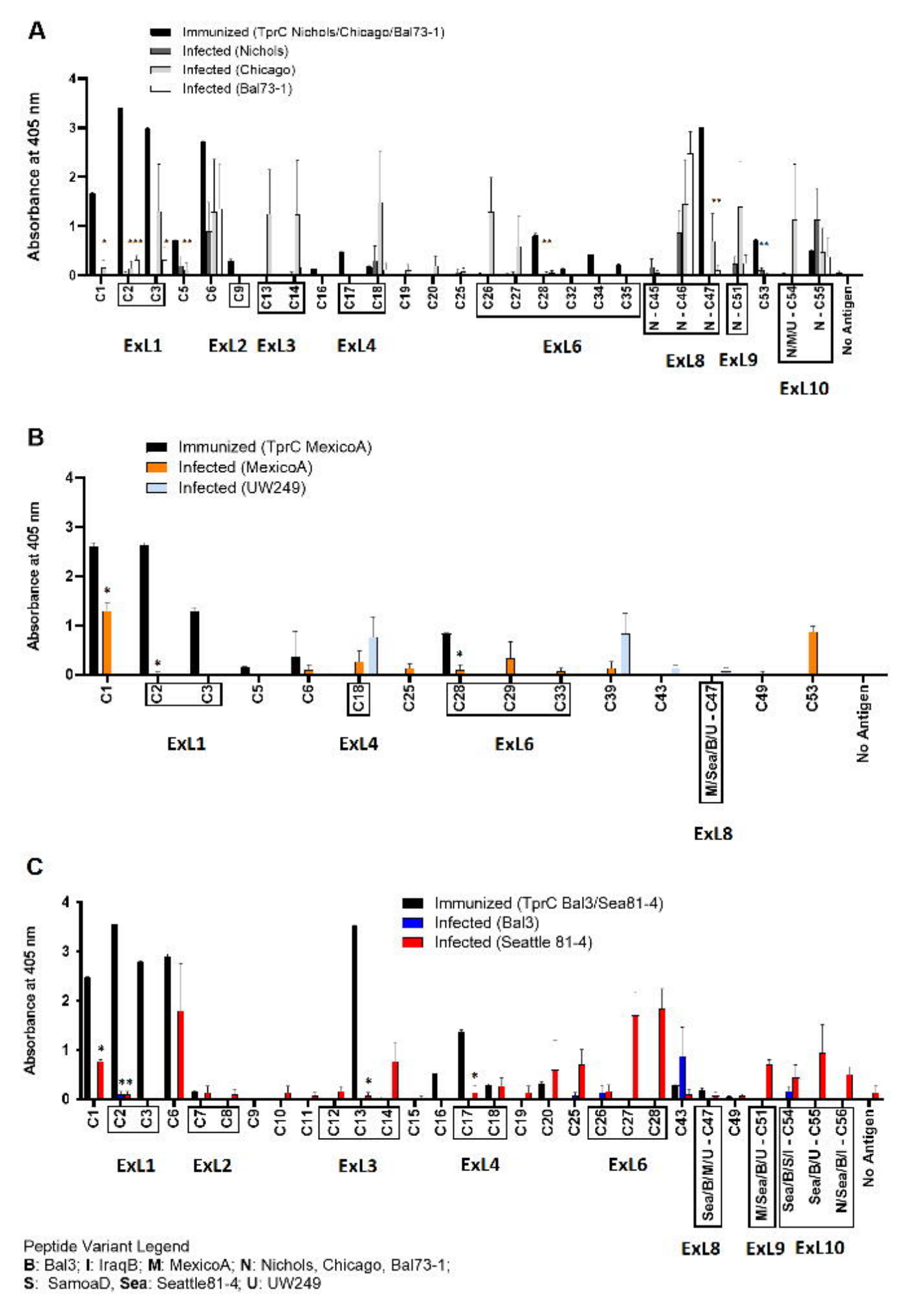
Bal3/Seattle81-4/IraqB C55	QWEQGVLSDVPYMGIAESIW
Nichols/Chicago/Bal73-1 C55	QWEQGVLADAPYMGIAESIW
SamoaD C55	QWEQGVLSDVPYMGITQSIW
MexicoA/UW249 C55	QWEQGVLADAPYMGITQSIG

Peptide Variant Legend

B: Bal3; **I**: IraqB; **M**: MexicoA; **N**: Nichols, Chicago, Bal73-1;
S: SamoaD, **Sea**: Seattle81-4; **U**: UW249



Peptide Variant Legend
B: Bal3; I: IraqB; **M**: MexicoA; **N**: Nichols, Chicago, Bal73-1;
S: SamoaD, **Sea**: Seattle81-4; **U**: UW249



Peptide Variant Legend
B: Bal3; **I**: IraqB; **M**: MexicoA; **N**: Nichols, Chicago, Bal73-1;
S: SamoaD, **Sea**: Seattle81-4; **U**: UW249

Table S1. TprC – IEDB B-cell epitope prediction results

No.	Start	End	Peptide	Peptide(s)	Location	Length
1	41	59	FQKNPRTGPGKHTHGFRTT	C2-C3	ExL1	19
2	71	82	KHTHTRRGEARS	C5-C6	Scaffolding	12
3	95	113	VELASSKSSTALSFTKPTA	C7-C9	ExL2	19
4	132	180	PSCVVNFAQLWKPFVTRAYSEKDTRYAPGFSGSGAKL GYQAHNVGNSGV	C11-C16	ExL3	49
5	192	208	NGAWDSTDTTHSKYGFG	C17-C18	ExL4	17
6	215	220	YGVDRQ	C19-C20	Scaffolding	6
7	233	251	LDQNYVKGTEDSKNENKTA	C21-C23	ExL5	19
8	277	295	GNQHQSNAHAQTQERAILK	C26-C27	ExL6	19
9	309	321	QNLPNIMMPPGIT	C29	ExL6	13
10	343	352	SAIQTVLAAG	C32-C33	ExL6	10
11	359	377	SQLVPNIEQGVRDVFRSSD	C34-C36	ExL6	19
12	393	396	PMNA	C37-C38	ExL6	4
13	421	436	TNIFGKRVFATTRAHY	C40	ExL7	16
14	448	452	KSGDP	C42-C43	Scaffolding	5
15	476	493	FYRNNGGYELNRVVPSGI	C45-C47	ExL8	18
16	527	539	NRFNIINAAGNLL	C50-C51	ExL9	13
17	564	585	WEQGVLSDVPYMGIAESIWSER	C54-C56	ExL10	22

Table S2. $TprD_2$ – IEDB B-cell epitope prediction results

No.	Start	End	Peptide	Peptide(s)	Location	Length
1	41	59	FQKNPRTGPGKHTHGFRTT	C2-C3	ExL1	19
2	71	82	KHTHTRRGEARS	C5-C6	Scaffolding	12
3	95	112	VELASSKSSTALSFTKPT	C7-C9	ExL2	18
4	132	180	PSCVVNFAQLWKPFVTRAYSEKDTRYAPGFSGSGAK LGYQAHNVGNSGV	C11-C16	ExL3	49
5	192	208	NGAWDSTDTTHSKYGFG	C17-C18	ExL4	17
6	215	220	YGVDRQ	C19-C20	Scaffolding	6
7	233	255	LDQNYVKGTEDSKNENKTALLWG	C21-C23	ExL5	23
8	277	297	GNQHQSNAQFYARMAPSQRVH	D26-D27	ExL6	21
9	308	316	LTSPQQDVV	D29	ExL6	9
10	320	332	VQELSKGSLLEKA	D30	ExL6	13
11	339	362	AQRTIVGLASSGGYLRHLNGKGLE	D32-D34	ExL6	24
12	366	376	RLIEQQKNPDA	D34-D35	ExL6	11
15	419	434	TNIFGERVFFKNQADH	D40	ExL7	16
16	446	450	KSGDP	C42-C43	Scaffolding	5
17	475	490	YINNGGAQYKGSNSDG	D45-D46	ExL8	16
18	525	537	NRFNHNQSGDALL	C50-C51	ExL9	13
19	562	570	WEQGVLADA	C54-C55	ExL10	9
20	578	583	SIWSER	C56	ExL10	6

Table S3. TprC – FBCPred B-cell epitope prediction results

Position	Epitope	Peptide(s)	Location	Score
42	QKNPRTGPGKHTHG	C2-C3	ExL1	0.999
70	SKHTHTRRGEARSG	C5-C6	Scaffolding	0.895
105	ALSFTKPTASFQAT	C8-C9	ExL2	0.962
130	TSPSCVVNFAQLWK	C11-C12	ExL3	0.985
156	RYAPGFSGSGAKLG	C14-C15	ExL3	0.998
171	QAHNVGNSGVDVDI	C15-C16	Scaffolding	0.992
191	SNGAWDSTDTTHSK	C17-C18	ExL4	0.995
241	TEDSKNENKTALLW	C22-C23	ExL5	0.923
259	RLTLEPGAGFRFSF	C24-C25	Scaffolding	1
308	VQNLPNIMMPPGIT	C29	ExL6	0.809
362	VPNIEQGVRDVFRS	C34-C35	ExL6	0.935
441	GFLKLETKSGDPYT	C42-C43	Scaffolding	0.997
464	VEARVYIPLTYVFY	C45	ExL8	0.825
524	GTTNRFNIINAAGN	C51-C52	ExL9	0.996
561	SAQWEQGVLSDVPY	C54-C55	ExL10	0.919

Table S3. TprD$_2$ – FBCPred B-cell epitope prediction results

Position	Epitope	Peptide(s)	Location	Score
42	QKNPRTGPGKHTHG	C2-C3	ExL1	0.999
70	SKHTHTRRGEARSG	C5-C6	Scaffolding	0.895
105	ALSFTKPTASFQAT	C8-C9	ExL2	0.962
130	TSPSCVVNFAQLWK	C11-C12	ExL3	0.985
156	RYAPGFSGSGAKLG	C14-C15	ExL3	0.998
171	QAHNVGNSGVDVDI	C15-C16	Scaffolding	0.992
191	SNGAWDSTDTTHSK	C17-C18	ExL4	0.995
241	TEDSKNENKTALLW	C22-C23	ExL5	0.923
259	RLTLEPGAGFRFSF	C24-C25	Scaffolding	1
308	LTSPQQDVVSFFVQ	D29	ExL6	0.8
365	MRLIEQQKNPDARM	D34-D35	ExL6	0.829
439	GFLKLETKSGDPYT	D42	Scaffolding	0.997
462	VEARVYIPLTYIFY	D45-D46	ExL8	0.975
483	YKGSNSDGVINTPI	D47	ExL8	0.996
523	ATNRFNHNQSGDAL	D51-D52	ExL9	0.956
558	LSAQWEQGVLADAP	C54-C55	ExL10	0.895

Fig. S1. TprC- BepiPred B-cell epitope prediction results

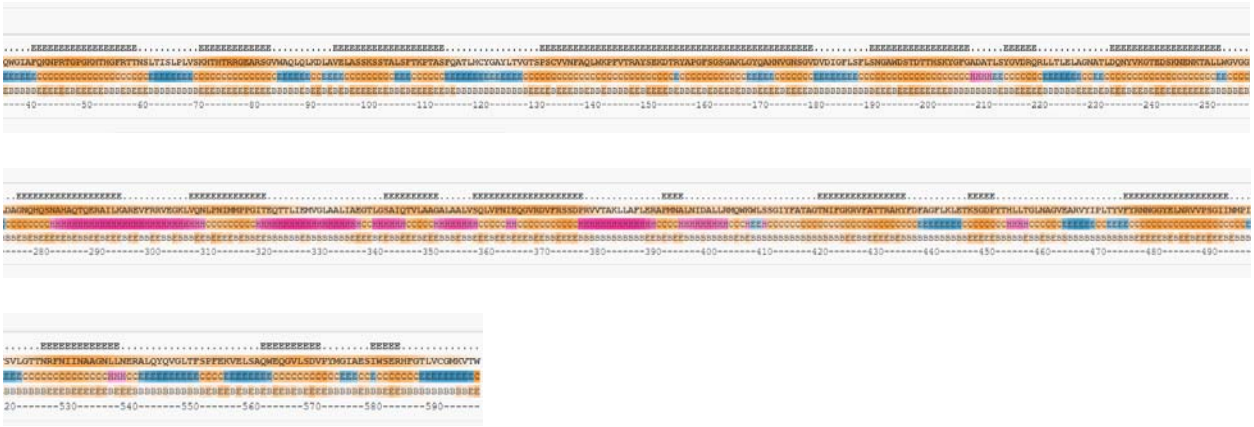
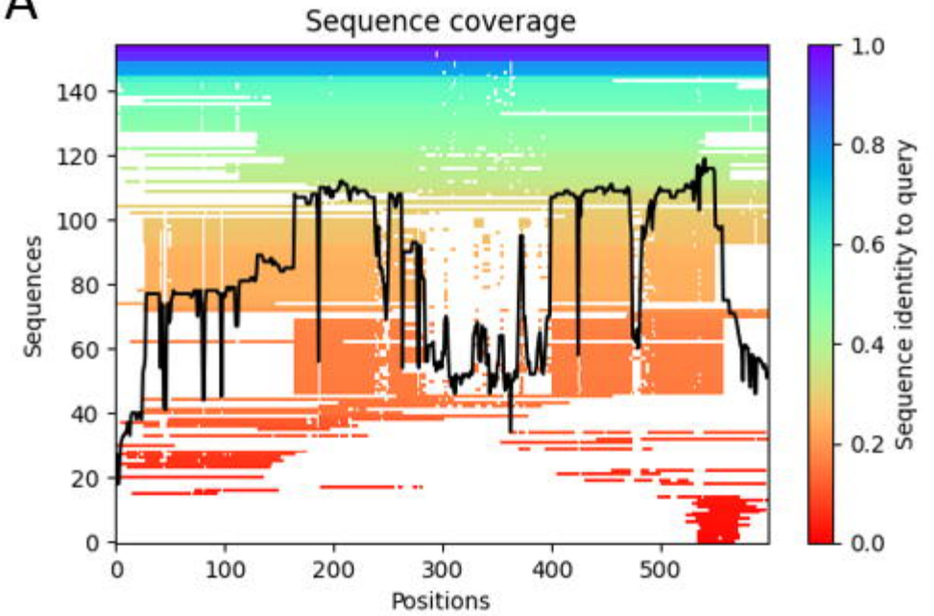


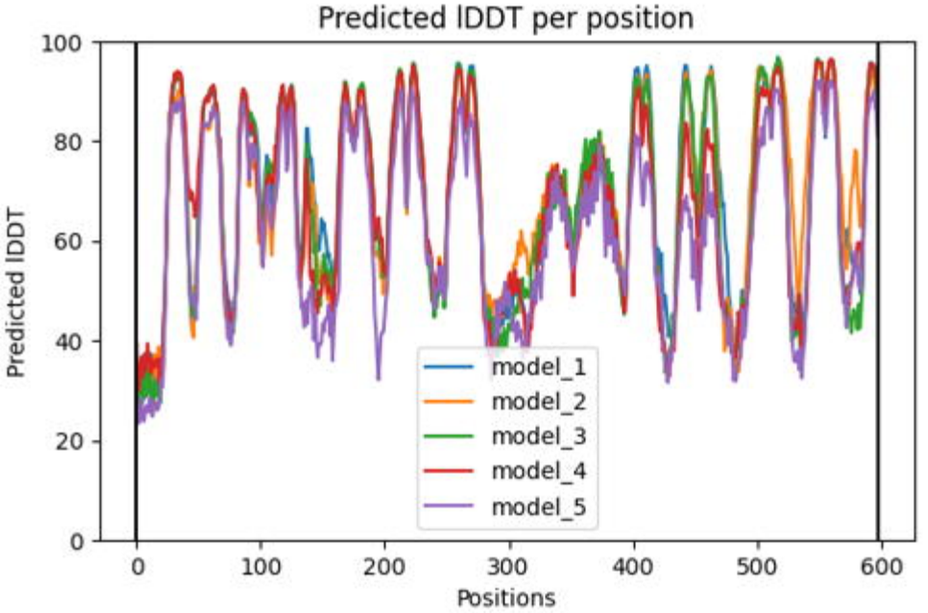
Fig. S2. TprD$_2$- BepiPred B-cell epitope prediction results

Fig.S1/2 Legend

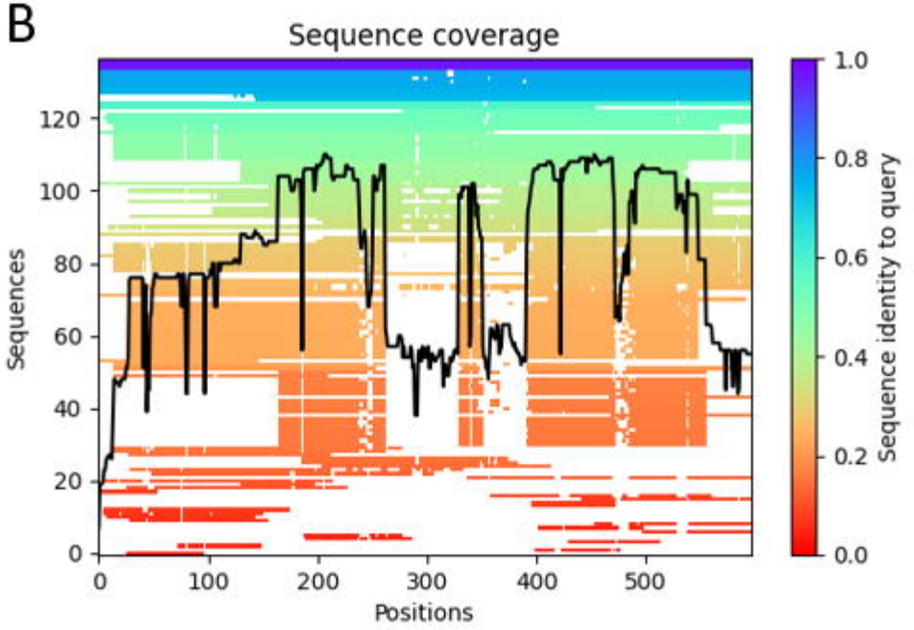
The output format shows the BepiPred-2.0 predictions and epitope classification for each sequence. The BepiPred-2.0 predictions are used to set the background color of the protein sequences. All predictions greater than a user-defined threshold (by default 0.5) are marked as 'E' in the 'Epitopes' line above the protein sequence itself.

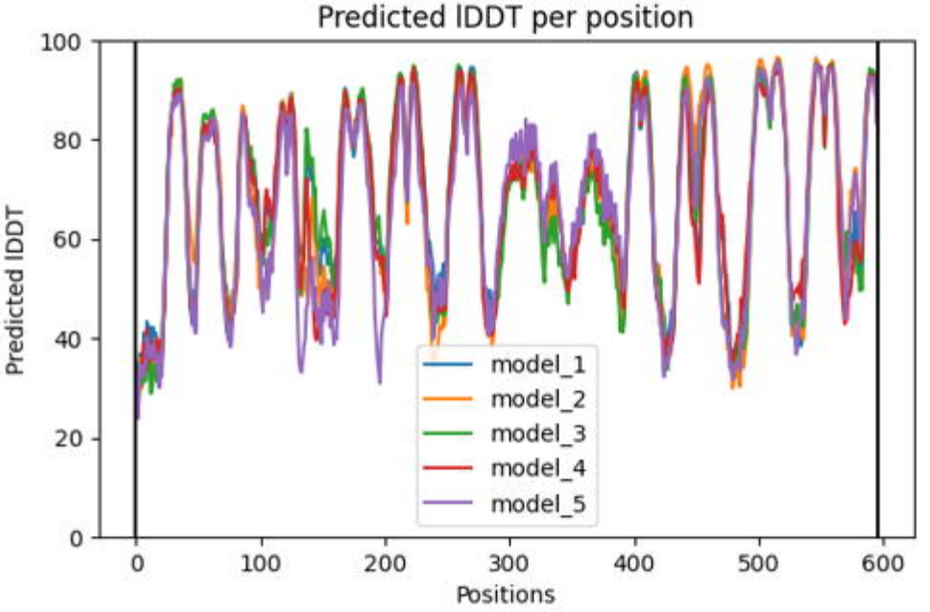
A





B





pLDDT analysis for TprC (A) and $TprD_2$ (B) for all models generated

TprC

Chain	Z	RMSD	lali	nres	% ID	Description	Number of strands
4frt-A	15.7	4.2	273	370	9	PROBABLE PORIN	18
4frt-B	15.6	4.2	270	374	9	PROBABLE PORIN	18
3t20-A	15.5	3.9	261	362	10	CIS-ACONITATE PORIN OPDH	18
2y0h-A	15.5	3.9	265	361	10	PROBABLE PORIN	18
3t0s-A	15.5	3.9	265	361	10	PORIN;	18
2y0k-A	15.5	3.9	262	364	7	PYROGLUTATMATE PORIN OPDO	18
3szv-A	15.4	3.7	257	364	8	PYROGLUTATMATE PORIN OPDO	18
2y0l-A	15.4	3.9	261	362	9	CIS-ACONITATE PORIN OPDH	18
3t24-B	15.4	4	268	368	9	PORIN;	18
4fso-B	15.2	3.9	249	338	8	PROBABLE PORIN	18
4fso-A	15.1	4	251	354	9	PROBABLE PORIN	18
4fms-B	15.1	4.2	265	368	7	PROBABLE PORIN	18
3szd-A	15.1	3.8	259	368	7	PORIN;	18
3szd-B	15	4.1	262	371	7	PORIN;	18
4fms-A	15	3.9	260	369	7	PROBABLE PORIN	18
3jty-B	14.9	4.2	272	383	8	BENF-LIKE PORIN	18
5ldt-C	14.9	3.5	274	403	12	MOMP PORIN	18
3sys-B	14.9	3.8	257	374	7	VANILLATE PORIN OPDK	18
4fsp-A	14.9	3.6	249	364	11	PROBABLE PORIN	18
2qtk-A	14.9	3.7	257	371	6	PROBABLE PORIN	18
1af6-C	14.9	4.6	281	421	9	MALTOPORIN;	18
2qtk-B	14.9	3.7	254	368	6	PROBABLE PORIN	18
1ujw-A	14.9	5.1	294	576	6	VITAMIN B12 RECEPTOR	22
5ldt-A	14.8	3.7	276	403	12	MOMP PORIN	18
3sy9-D	14.8	4.1	259	361	11	HISTIDINE PORIN OPCD	18
4kra-B	14.8	4.3	251	340	6	OUTER MEMBRANE PROTEIN F	16

4ft6-A	14.8	4	273	387	9	PROBABLE PORIN	18
2y2x-B	14.8	3.8	258	378	6	VANILLATE PORIN OPDK	18
3sys-A	14.8	3.6	254	374	7	VANILLATE PORIN OPDK	18
5ldt-B	14.8	3.7	274	403	12	MOMP PORIN	18
3jty-C	14.8	4	267	384	7	BENF-LIKE PORIN	18
5dl8-A	14.8	4.5	277	391	11	BENZOATE TRANSPORT PORIN BENP	18
3jty-D	14.8	3.9	267	384	7	BENF-LIKE PORIN	18
2y2x-A	14.8	3.7	255	374	7	VANILLATE PORIN OPDK	18
3t24-A	14.8	4	268	377	9	PORIN;	18
2xe5-D	14.7	4.2	245	343	9	OUTER MEMBRANE PROTEIN C	16
2xe5-C	14.7	4.2	245	343	9	OUTER MEMBRANE PROTEIN C	16
2xe2-C	14.7	4.2	245	343	9	OUTER MEMBRANE PROTEIN C	16
2xe5-E	14.7	4.2	245	343	9	OUTER MEMBRANE PROTEIN C	16
3jty-A	14.7	4.3	273	388	7	BENF-LIKE PORIN	18
2xe5-A	14.7	4.2	245	343	9	OUTER MEMBRANE PROTEIN C	16
5dl8-B	14.7	4.4	276	391	12	BENZOATE TRANSPORT PORIN BENP	18
3k19-I	14.6	4.1	242	340	13	OUTER MEMBRANE PROTEIN F	16
3nsg-C	14.6	4.3	248	341	6	OUTER MEMBRANE PROTEIN F	16
3o0e-C	14.6	4.1	243	340	13	PORIN OMPF	16
7nst-C	14.6	4.9	248	340	13	OUTER MEMBRANE PROTEIN F	16
5o77-A	14.6	4.1	239	337	8	OMPK35;	16
2xe5-F	14.6	4.2	245	343	9	OUTER MEMBRANE PROTEIN C	16
3k19-L	14.6	4.1	242	340	13	OUTER MEMBRANE PROTEIN F	16
3k19-H	14.6	4.1	242	340	13	OUTER MEMBRANE PROTEIN F	16

TprD2

Chain	Z	RMSD	lali	nres	%ID	Description	Number of strands

1ujw-A	14.9	4.8	298	576	8	VITAMIN B12 RECEPTOR	22
5onu-A	14.6	4	248	320	6	OUTER MEMBRANE PROTEIN OMPU	16
6ehb-C	14.6	4.3	240	320	6	OUTER MEMBRANE PROTEIN U	16
5onu-C	14.6	4	247	320	6	OUTER MEMBRANE PROTEIN OMPU	16
6ehe-A	14.6	5.3	245	302	11	OMPT PROTEIN	16
6ehb-B	14.5	4	236	317	7	OUTER MEMBRANE PROTEIN U	16
5onu-B	14.5	4	243	320	7	OUTER MEMBRANE PROTEIN OMPU	16
4frt-A	14.5	3.8	257	370	7	PROBABLE PORIN	18
6ehf-A	14.5	5.2	239	318	10	OMPT PROTEIN	16
6ehd-A	14.4	5	240	322	10	OMPT PROTEIN	16
5o78-C	14.4	4.4	249	331	8	PHOSPHOPORIN PHOE	16
5o78-B	14.4	4.4	244	331	8	PHOSPHOPORIN PHOE	16
3t24-B	14.4	4.1	263	368	9	PORIN;	18
6ehb-A	14.4	3.8	233	318	6	OUTER MEMBRANE PROTEIN U	16
2y0l-A	14.4	3.8	256	362	9	CIS-ACONITATE PORIN OPHD	18
5o78-A	14.4	4.4	246	331	8	PHOSPHOPORIN PHOE	16
3t20-A	14.4	3.8	256	362	9	CIS-ACONITATE PORIN OPHD	18
6v78-C	14.3	4.1	246	353	10	OMPK37;	16
1gfn-A	14.3	4.4	245	327	12	MATRIX PORIN OUTER MEMBRANE PROTEIN F	16
6ehc-A	14.3	3.9	240	307	7	OUTER MEMBRANE PROTEIN U	16
5o9c-B	14.3	4.6	247	348	11	OSMOPORIN OMP36	16
6ehc-B	14.3	4.1	243	307	6	OUTER MEMBRANE PROTEIN U	16

4fms-B	14.3	4	262	368	7	PROBABLE PORIN	18
4frt-B	14.2	3.8	255	374	7	PROBABLE PORIN	18
3szd-A	14.2	3.8	260	368	7	PORIN;	18
3wi5-A	14.2	4	238	312	11	MAJOR OUTER MEMBRANE PROTEIN P.IB	16
5mdq-B	14.2	4.4	242	350	8	CHITOPORIN;	16
1af6-B	14.2	4.2	274	421	10	MALTOPORIN;	18
1e54-A	14.1	3.6	233	332	8	OUTER MEMBRANE PORIN PROTEIN 32	16
5ldt-A	14.1	3.8	279	403	12	MOMP PORIN	18
6ene-A	14.1	4.4	245	329	11	PHOSPHOPORIN PHOE	16
4bum-X	14.1	3.5	241	283	7	VOLTAGE-DEPENDENT ANION CHANNEL 2	19
5o77-A	14.1	4.4	245	337	8	OMPK35;	16
6rcp-D	14.1	4.5	247	349	9	OMPK36;	16
6rcp-F	14.1	4.5	248	349	9	OMPK36;	16
2fgr-A	14.1	3.7	233	332	8	OUTER MEMBRANE PORIN PROTEIN 32	16
3szd-B	14.1	3.9	259	371	7	PORIN;	18
6rcp-J	14.1	4.7	250	349	8	OMPK36;	16
4fms-A	14.1	3.8	259	369	7	PROBABLE PORIN	18
6ene-C	14.1	4.3	249	329	10	PHOSPHOPORIN PHOE	16
3szv-A	14.1	4	253	364	8	PYROGLUTATMATE PORIN OPDO	18
7nie-C	14	3.6	235	294	9	GLYCEROL KINASE	19
7nie-D	14	3.6	235	294	9	GLYCEROL KINASE	19
4kra-C	14	4.6	252	337	7	OUTER MEMBRANE PROTEIN F	16
7nie-L	14	3.6	235	294	9	GLYCEROL KINASE	19
4lsh-B	14	4.4	249	334	13	OUTER MEMBRANE PROTEIN F	16
4kra-B	14	4.3	252	340	7	OUTER MEMBRANE PROTEIN F	16
2qtk-B	14	4	257	368	6	PROBABLE PORIN	18
6rcp-L	14	4.8	249	349	8	OMPK36;	16

						HISTIDINE PORIN	
3sy9-C	14	4	251	360	11	OPDC	18