



OPEN

Broad specificity of immune helminth scFv library to identify monoclonal antibodies targeting *Strongyloides*

Anizah Rahumatullah, Dinesh Balachandra, Rahmah Noordin, Zamrina Baharudeen, Yee Ying Lim, Yee Siew Choong & Theam Soon Lim✉

Antibodies have different chemical properties capable of targeting a diverse nature of antigens. Traditionally, immune antibody libraries are perceived to be disease-specific with a skewed repertoire. The complexity during the generation of a combinatorial antibody library allows for a skewed but diverse repertoire to be generated. *Strongyloides stercoralis* is a parasite that causes strongyloidiasis, a potentially life-threatening disease with a complex diagnosis that impedes effective control and treatment of the disease. This study describes the isolation of monoclonal antibodies against *S. stercoralis* NIE recombinant protein using an immune antibody phage display library derived from lymphatic filaria-infected individuals. The isolated antibody clones showed both lambda and kappa light chains gene usage, with diverse amino acid distributions. Structural analysis showed that electropositivity and the interface area could determine the binding affinity of the clones with NIE. The successful identification of *S. stercoralis* antibodies from the filarial immune library highlights the breadth of antibody gene diversification in an immune antibody library that can be applied for closely related infections.

In the past decade, phage display technology has become a common tool for the discovery of novel binders against various antigen targets¹. The nature of the target antigen can range from proteins, haptens, peptides, enzymes, membrane fractions, liposomes, virus-like particles, cells, tissue sections, or entire tissues^{2,3}. Phage display technology has been the preferred in vitro method for the production of monoclonal antibodies replacing the conventional hybridoma technology. Central to phage display technology is the construction of a phage library displaying a plethora of unique antibodies. An essential aspect of antibody libraries is the source of the antibody repertoire used for library preparation. Phage libraries can be divided into two main types, i.e., natural libraries that consist of antibody genes from immune and non-immune (naïve) donors, and synthetic libraries comprising antibody genes derived from chemical synthesis.

A naïve antibody library repertoire is derived from the IgM repertoire of donors in a healthy state. The advantage of this library is that it can be used for discovery of antibodies against a wide array of antigens. However, the drawback of naïve library-derived antibodies is a general lower affinity compared to antibodies from an immune source and a higher possibility of cross-reactions^{4,5}. Nevertheless, this limitation may be circumvented by in vitro affinity maturation processes. Conversely, an immune library is derived from the IgG repertoire from an infected host. The antibody repertoire in immune libraries consists of B cells that have been exposed to a particular pathogen and have undergone affinity maturation processes². Thus, the resulting antibodies have an increased affinity towards the target antigen, enabling isolation of high-affinity binders. The antibody repertoire from an immune source is generally of a lower diversity making immune libraries not as broadly applicable in terms of their antigen scope as compared to naïve libraries.

In the present study, we utilized a previously constructed Human Antibody Disease ENhanced (HAYDEN)-Filariasis library which is an immune helminth phage display library⁶ to isolate monoclonal antibodies against *Strongyloides stercoralis* (*S. stercoralis*) NIE recombinant protein (rNIE). This parasite causes strongyloidiasis, a potentially life-threatening disease with a complex diagnosis that impedes effective control and treatment of the disease. The helminth library was generated using the blood of individuals infected with lymphatic filaria, specifically *Brugia malayi* (*B. malayi*). The samples were taken from apparently asymptomatic individuals with

Institute for Research in Molecular Medicine (INFORMM), Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia. ✉email: theamsoon@usm.my

parasite larvae (microfilariae) in their blood. Parasites causing lymphatic filariasis (LF) and strongyloidiasis can be found in the blood and tissue of infected individuals. They are both helminths under the phylum of Nematoda (roundworm)⁷. Infection with helminth parasites are commonly associated with an elevation of IgE antibodies, eosinophilia, mucous mastocytosis, and goblet cells hyperplasia⁸. During helminth infection, T helper cell 2 (Th2) type response mediates protection, while B cells play vital roles in antibody secretion⁹, stimulation and control of Th2-type immune responses¹⁰. Following infection, antibodies such as IgG and IgM can act as potent mediators of protective immunity, and Th2-type responses trigger B cell class switching to IgE and IgG¹¹. The commonalities of immune responses of both parasites indicate the possibility of using the antibody library against *B. malayi* to isolate antibodies against *S. stercoralis*.

Strongyloides stercoralis, a human parasitic roundworm is estimated to infect approximately 370 million¹² people globally, with a mortality rate of 16.7% among patients requiring hospitalization. Meanwhile, among immunocompromised patients, the fatality is 60–85%¹¹. *S. stercoralis* mainly infects humans through the penetration of infective filariform larvae into the skin when there is contact with contaminated soil. This nematode's autoinfection mechanism allows infection to be maintained life long as it is able to carry out and restart its life cycle within the human body. Immunosuppressed patients are most at risk of developing the life-threatening clinical symptoms of strongyloidiasis due to hyperinfection syndrome or dissemination¹³. Therefore, the treatment of *S. stercoralis* infections must completely eradicate the parasite from the body to avoid complications of hyperinfection and dissemination syndrome which may prove to be fatal. Strongyloidiasis is listed as a neglected tropical disease (NTD) by the World Health Organization (WHO) in 2004 and more often, the diagnostics of this disease is often overlooked as it requires the utility of complex diagnostic techniques and resources. The difficulty of diagnosing this disease is mainly due to the sporadic output of larvae and serological cross-reactivity to other parasitic infections.

In the present study, the protein used for biopanning was rNIE, a highly antigenic *S. stercoralis*-specific antigen used for serodiagnosis of strongyloidiasis. It was derived from an *S. stercoralis* L3 cDNA library¹⁴. Based on its high diagnostic specificity and sensitivity, rNIE has been commonly used for development of antibody detection assays for strongyloidiasis^{15,16}.

The inherent ability of an immune library generated utilizing the antibody repertoire of a closely related infection, to present useful repertoire for antibody generation is not well-studied. The ability of the LF immune library to isolate *S. stercoralis*-specific antibodies demonstrates a broader coverage of immune libraries than currently understood, thus making it an even more valuable asset for antibody generation. This study highlights the immense genetic diversity of the antibody repertoire in an immune antibody library that yields antibodies against a related organism.

Results

Recombinant protein preparation and verification. rNIE was expressed and purified for biopanning. The protein was successfully expressed and verified using SDS PAGE and Western blot with an expected size of approximately 37 kDa (Fig. S1). The analysis showed the rNIE was produced with good yield and purity suitable for biopanning.

Isolation of monoclonal antibodies. Monoclonal antibody clones against *Strongyloides* rNIE were isolated using previously described biopanning method⁶. The polyclonal ELISA results showed enrichment of rNIE-specific phage in consecutive rounds of the selection process and a significant increase in absorbance value approximately at 3.0 (Fig. 1a). Figure 1b shows the amount of phage used and recovered from each round of panning. The amount of phage particles from rounds 1 to 3 were 8.71×10^{10} , 5.1×10^{10} and 1.21×10^{10} , respectively. The amount of recovered phage particles in rounds 1 to 3 was 1.6×10^6 , 2.6×10^6 and 3.6×10^8 pfu, respectively. The enrichment ratios for rounds 1 to 3 were determined to be 8×10^{-5} , 3×10^{-6} and 1.2×10^3 , respectively. This was translated to an enrichment ratio of 3.25 and 150 folds for rounds 2 and 3, respectively.

A total of 184 antibody clones were screened in the monoclonal ELISA (Fig. 1c) and this resulted in 104 positive binders with absorbance values ranging from 0.61 to 3.95. Positive clones were identified with a cut-off OD_{405nm} value above 0.5 after subtracting the background values with the majority of clones having OD_{405nm} readings above 2.5. This indicates that the helminth phage display library was capable of enriching rNIE-specific monoclonal antibodies.

Antibody gene analysis. Sequencing results of the positive binders were analysed by IMGT/V-QUEST software to determine the identity of the scFv clones based on the human germline sequences available in the database. The sequence analysis showed that only 30 full-length antibody clones were obtained. The remaining 60 clones had only partial scFv sequences with either a complete heavy or light chain with complete complementarity-determining regions (CDR), while another 14 clones had mutations that resulted in frameshifts and truncations. These problematic clones were omitted from the gene pairing and antibody gene analysis.

The 30 scFv clones showed variations in their antibody gene family distributions clustered into 4 unique gene pairings (Fig. 2). In total, 29 out of 30 (97%) sequences showed functional variable lambda (VL) genes with only a single (3%) sequence for variable kappa (VK) gene. In addition, there was a preference seen in the heavy chain (VH) gene usage with majority of the clones originating from VH3 (94%) followed by VH5 (3%) and VH1 (3%). On the other hand, LV6 (87%) was the dominant gene used followed by LV3 (10%) and KV3 (3%) for the light chain. Next, there was a clear enrichment for one heavy and light chain gene pairing. The most common pairing was IgHV3-LV6 (87%) followed by IgHV3-LV3 (7%), IgHV1-LV3 (3%) and IgHV5-KV3 (3%). A comparison of the antibody gene family usage for *B. malayi* antigens is included and described in the discussion.

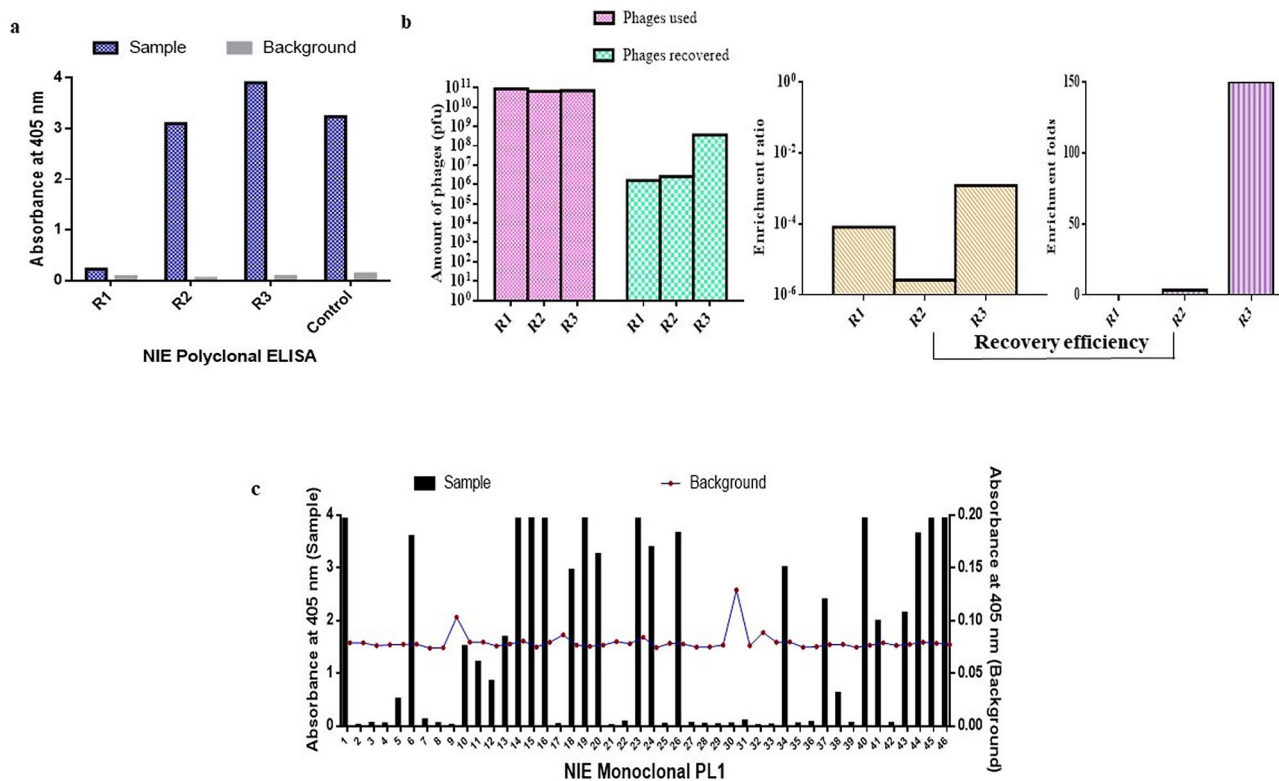
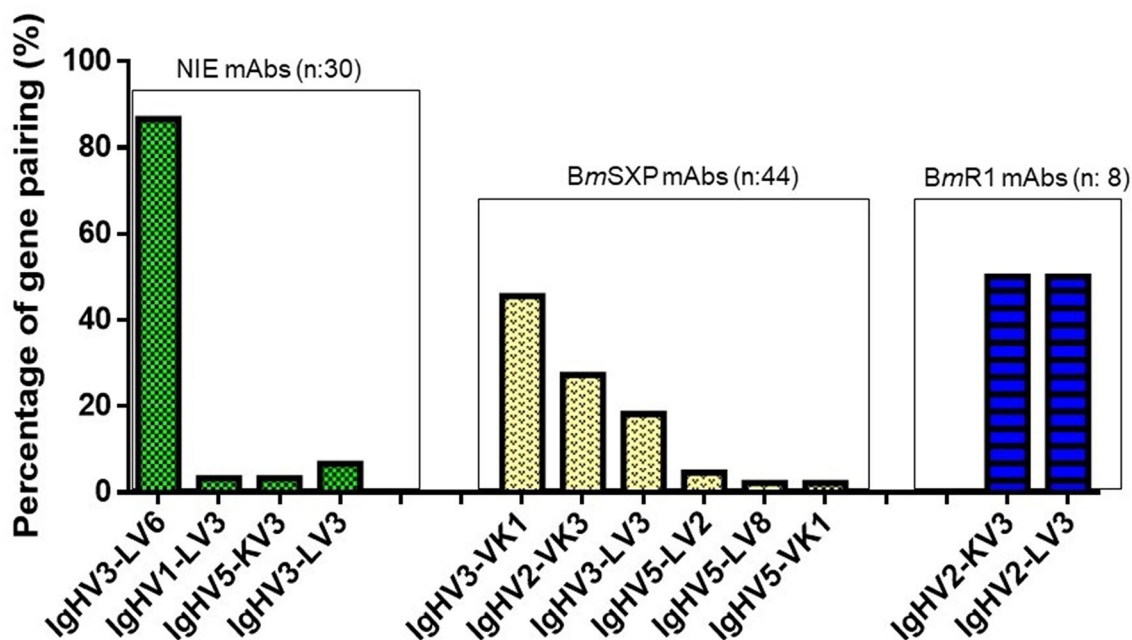


Figure 1. Isolation of rNIE specific monoclonal antibodies. (a) Polyclonal phage ELISA of rNIE antigen during the panning cycles. (b) Phage particles used and recovered in rounds 1 to 3 panning. (c) Representative of monoclonal phage ELISA analysis of scFv clones from helminth library against rNIE antigen. The positive control is a non-related protein that is reactive to the anti-M13-HRP antibodies and BSA was used as the negative control.



Summary of gene usage and comparisons with previously isolated *B. malayi* antigens antibodies

Figure 2. Analysis of gene pairing frequency of isolated of rNIE antigen antibody clones and comparison against previous *B. malayi* antigens antibodies (BmR1 and BmSXP).

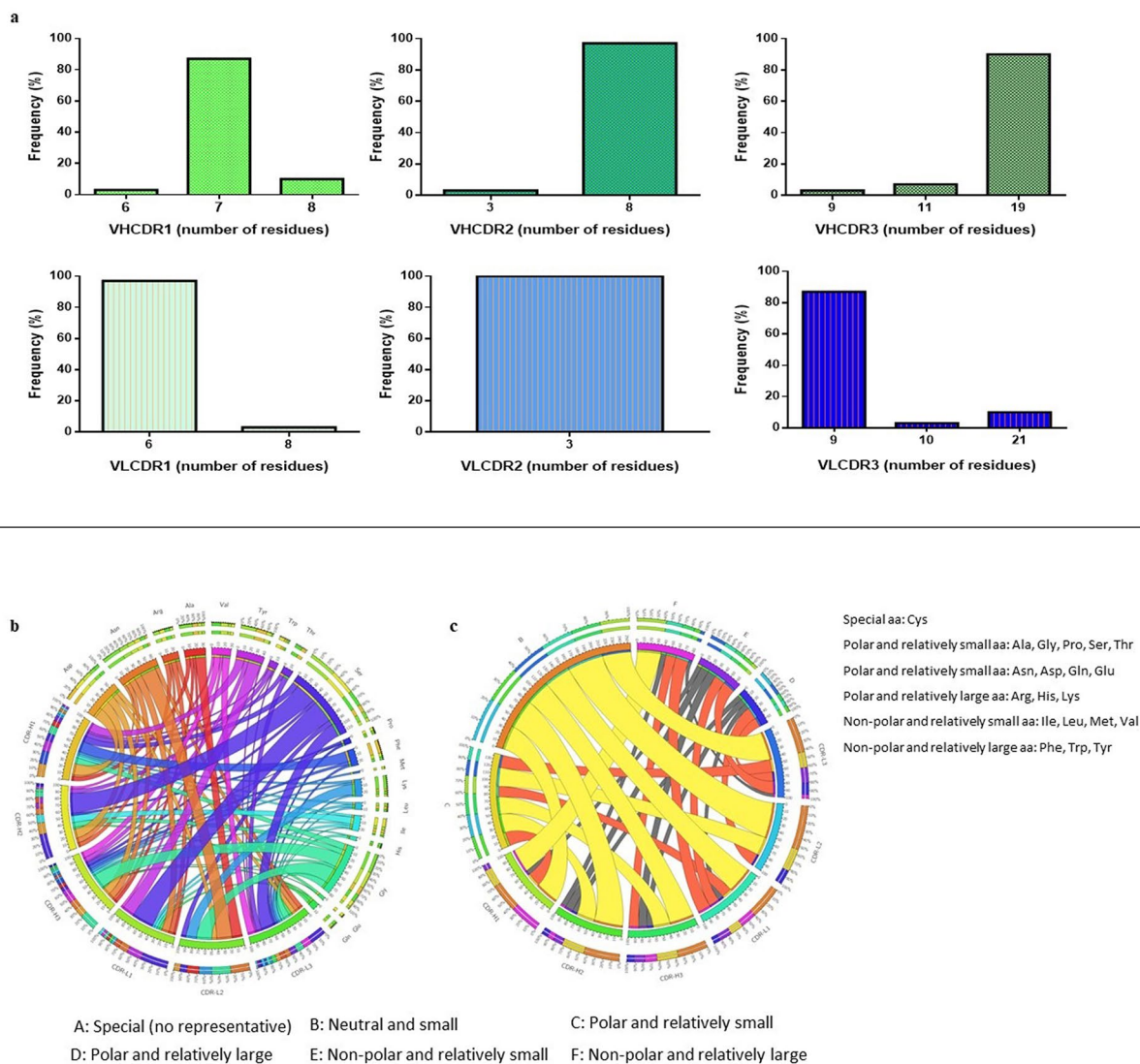


Figure 3. Monoclonal antibody clone characterization. **(a)** Amino acid length variations of heavy and light chains of antibody clones. **(b)** CIRCOS diagram analysis for amino acid frequency. **(c)** CIRCOS diagrams were drawn using Circos Table Viewer v0.63–9 (<http://mkweb.bcgsc.ca/tableviewer/visualize>).

Antibody sequence analysis. The CDR analysis consisting of the length and composition of the CDR region was observed due to its vital role for the binding site topology. The range of amino acid (aa) length for the VH is between 3 to 19 aa (Fig. 3a). The HC CDR1 has a length distribution from 6 to 8 aa with 7 aa being the dominant length. CDR2 had only 3 and 8 aa, whereas for CDR3 the distribution was 9, 11 and 19 aa. The dominant length for CDR2 and CDR3 was 8 aa and 19 aa, respectively. The distribution of the LC CDRs was also varied with 6 and 8 aa for CDR1 and 3 aa for CDR2. However, the length range for CDR3 was broader from 9 to 21 aa with 9 aa as the dominant length.

The aa propensity for the enriched scFv CDR regions was observed and a Circos plot analysis was performed to highlight the distribution patterns in a clearer presentation. It showed some regions with random aa distributions while others showed a skewed aa representation for both heavy and light chains (Fig. 3b). The Circos plot allows the visualization of the frequency of the aa present in each of the CDR. The HC CDR1 showed a higher representation of aspartic acid, phenylalanine and tyrosine compared to other aa, whereas the HC CDR2 showed over representation of serine. The highly diverse HC CDR3 was dominated with glycine, aspartic acid, valine and asparagine. Conversely, for LC CDR1 and CDR3 the distribution of aa was random but with a significant increase in serine while the LC CDR2 showed a skewed pattern of distribution with higher representations of glycine and asparagine. The unique presentation by the Circos plot allows for a global view to analyze relationship patterns between the different CDR. Figure 3b shows the dominance of serine in CDR-H2, CDR-L1 and CDR-L3. However, glycine was dominant in CDR-H3 and CDR-L2. The aa usage in CDR-L2 is unique in comparison to the other CDR with a more even distribution of lysine, glycine, asparagine and alanine. The aa usage in the other CDR are randomly distributed with the utilization of more different aa.

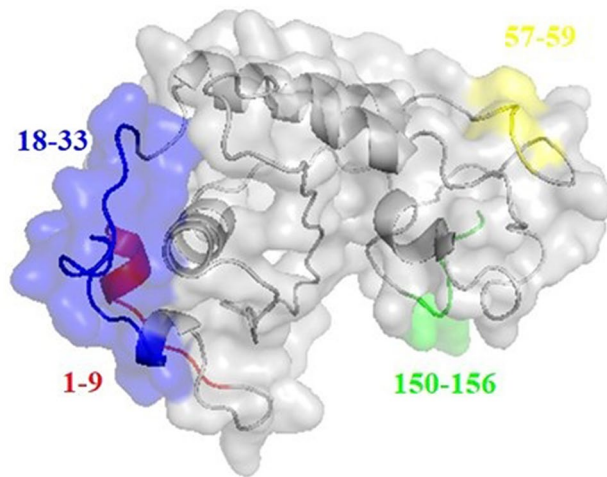


Figure 4. The four predicted consensus B-cell epitopes of NIE (residues 1–9, 18–33, 57–59 and 150–156). Figure was prepared by PyMol ver 2.4.0 (<https://pymol.org/2/>).

The Circos plots analyzing the polarity distribution for all CDRs in both chains showed similar distribution patterns with a higher representation of neutral and small aa in all CDR (Fig. 3c). CDR-H1 has a similarly high presence of non-polar and relatively large aa. CDR-H2 and CDR-L2 also has a significant amount of polar and relatively small aa present in the binding pocket. Moderate amounts of non-polar and relatively large aa is found in CDR-H3 and CDR-L1. Non-polar and relatively small aa are moderately distributed in CDR-L1 and CDR-L3.

In addition, the Kabat numbering scheme was used to further analyze the CDR3 aa position. CDR3 HC was determined to be from position 93 to 102, while position 89 to 97 was designated for CDR3 LC (Fig. S2). The dominance of certain amino acids can be observed at different positions. For CDR3 HC, at position 93 the dominating aa is alanine, 95 and 100C is asparagine, 97, 99, 100D and 100H is glycine, 100A and 100F is phenylalanine while position 100B and 100G was only restricted to tyrosine. The dominating aa at position 100E and 100I was only restricted to valine and methionine, respectively. For CDR3 LC, positions 95B, 95E and 95F were restricted to histidine, phenylalanine, and glycine respectively. Serine was dominating at position 90, 91, 93, and 95. At position 89 the dominating aa is glutamine, 92 is aspartic acid and 95A is asparagine.

Structural analysis. The consensus secondary structure analysis showed that the N-terminal (residues 3–29) of the NIE is a helix while the rest of the region is coils/turns (Table S1). The secondary structure calculations of the NIE structures modelled by different approaches showed similar findings but with the addition of three, one two helices in the comparative, de novo and ab initio models, respectively (Table S2). Structural visualization on the NIE structure modelled using different approaches also showed a similar globular shaped protein and the structural stoichiometry evaluation of the comparative model has the best scores among the structures build from different approaches (Fig. S3). Therefore, the comparative model was selected for further docking simulation with scFvs. B-cell epitope prediction results indicated 4 possible epitopes for NIE (residues 1–9, 18–33, 57–59 and 150–156; Fig. 4). These 4 epitopes were then marked as the active interface residues for the docking simulation with the four antibody clones. On the other hand, the electrostatic energy analysis of the modelled antibody clones showed that Ab5 and Ab14 have more electropositive CDR3 while the CDR3 of Ab6 is more electronegative and CDR3 of Ab23 consists of both electropositive and electronegative residues (Fig. 5).

Docked conformation of the antibodies showed different binding orientations with the NIE (Fig. 6). CDR3 was observed as the core antigen binding domain. Table S3 showed that all clones interact with NIE with the CDR3 except for Ab6. Ab14 and Ab23 have the highest number of interacting residues (36 and 32 interaction, respectively) with NIE compared with Ab5 and Ab6 (both with 26 interactions with NIE). The isoelectric point (pI) of Ab6 is 5.2 while other clones have pI values of more than 8.0 (Table S4). Therefore, Ab6 is the only clone with a negative net charge at pH7. The calculated interface area of Ab6 was also the lowest (614.4 Å²) among the antibody clones. Estimated binding free energies (G_{Bind}) of the antibodies with NIE showed that Ab14 and Ab23 ($G_{\text{Bind}} = -7.8$ and -6.1 kcal/mol, respectively) are better binders compared with Ab5 and Ab6 (both with $G_{\text{Bind}} = -5.6$ kcal/mol).

IMGT Colliers de Perles was also applied to analyse the structure of the scFv variable domains depicted with a unique numbering system and standardized 2D graphical representations of the antibody molecule (data not shown). This tool uses statistical methods to analyze the aa and hierarchic classification of framework positions based on three properties (hydropathy, volume and physiochemical characteristics). The analysis consists of the distribution of hydrophobic aa, hydropathy and other aa properties displaying the key aa that play vital role in protein interactions and domain structures. Based on 2D graphical representations, all four antibody clones showed similar pattern of aa distribution.

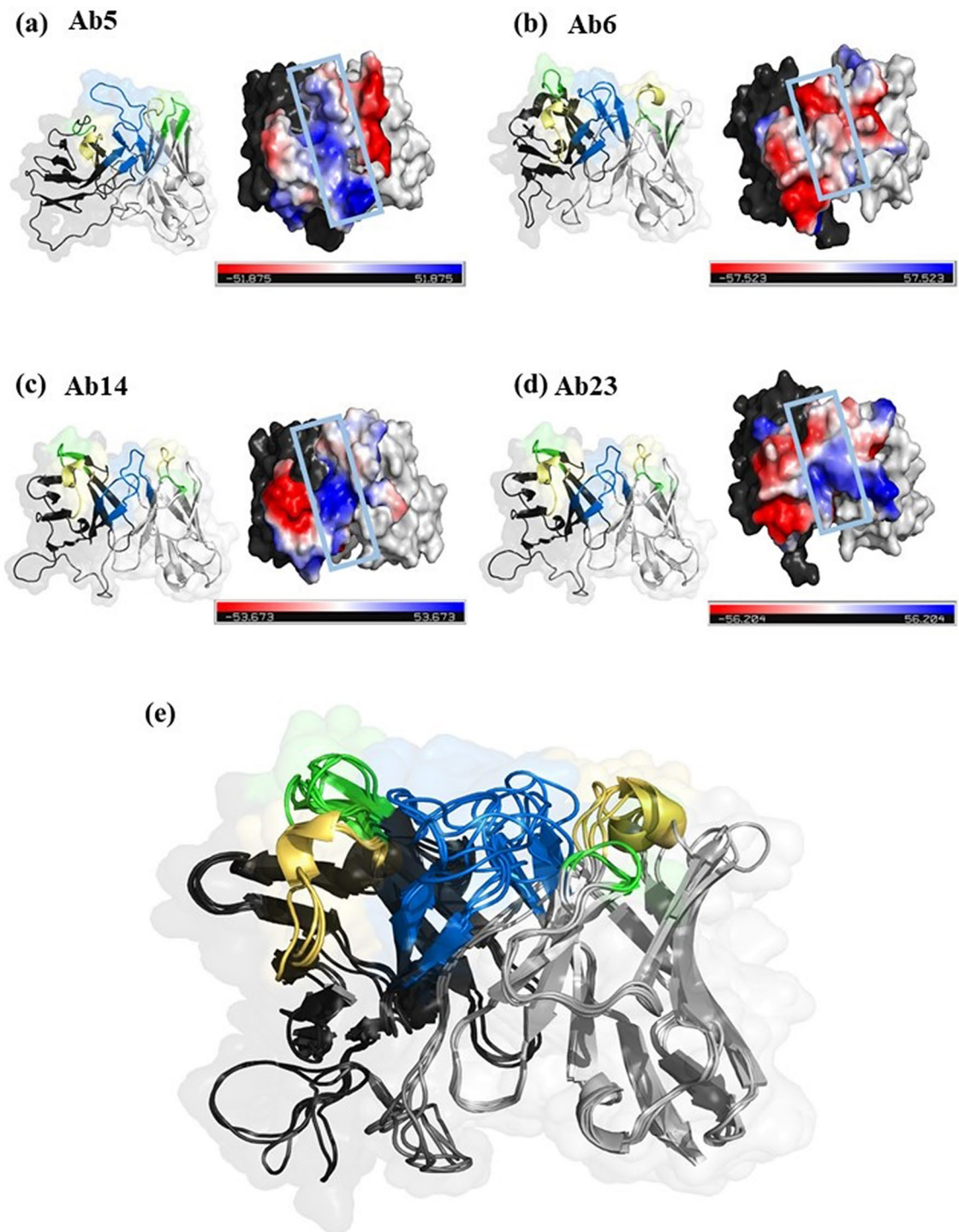


Figure 5. Three dimensional structures of rNIE specific antibodies (a) Ab5, (b) Ab6, (c) Ab14 and (d) Ab23, by surface/ribbon (left) and electrostatic (right) presentations. Variable light (VL) and variable heavy (VH) domains are in black and grey ribbon presentation, respectively. CDR1, CDR2 and CDR3 are in yellow, green and blue ribbon presentation, respectively. Electronegative and electropositive residues are in red and blue surface presentations viewed from top. CDR3 is highlighted with blue box. (e) Superimposition of the four modelled scFvs. Figure was prepared by PyMol ver 2.4.0 (<https://pymol.org/2/>).

Preparation of recombinant monoclonal antibody protein. In total, four clones of unique gene families (Ab5: IgHV3-LV1; Ab6: IgHV3-LV6; Ab14: IgHV5-KV3; Ab23: IgHV3-LV3) were selected and successfully subcloned into pET 51(b)+ vector for expression in SHuffle T7 *Escherichia coli* cells. The SDS PAGE

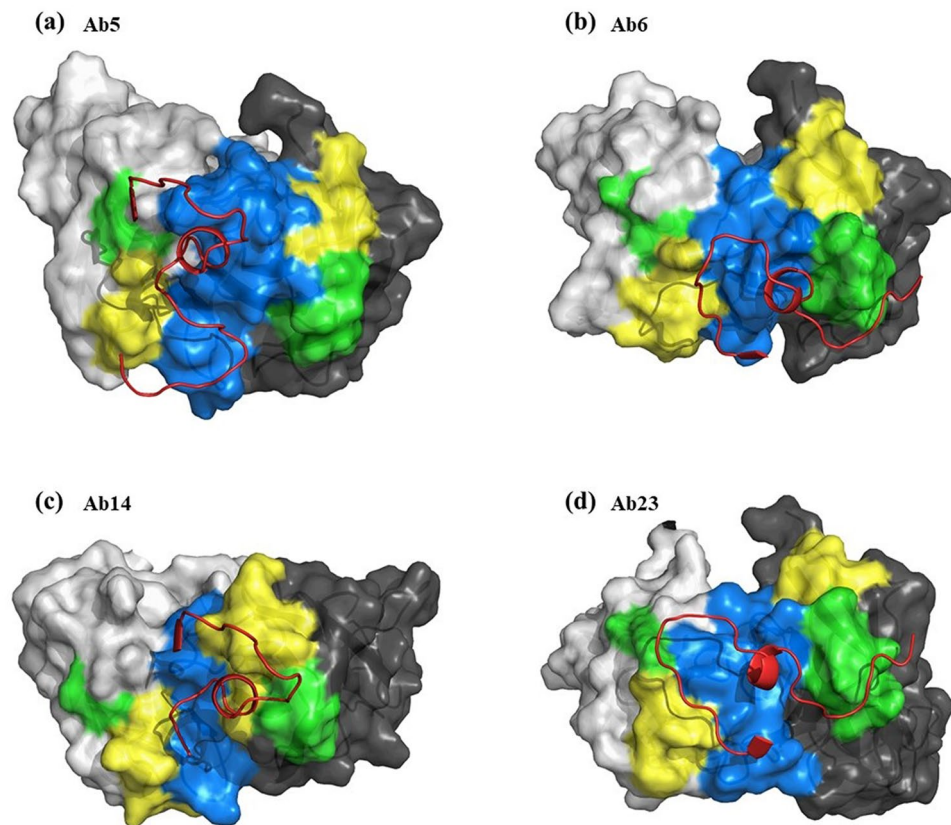


Figure 6. Surface presentation of the docked conformation of the antibody clones (a) Ab5, (b) Ab6, (c) Ab14 and (d) Ab23 with rNIE. VH and VL domains are in black and grey presentation with respective Yellow, green and blue represent CDR1, CDR2 and CDR3 respectively. Red ribbon is the rNIE residues that interacts with the antibody. Figure was prepared by PyMol ver 2.4.0 (<https://pymol.org/2/>).

(Fig. S4) and Western blot showed the expected molecular mass of the recombinant antibodies at approximately 35 kDa (Fig. 7a). The antibodies were expressed and purified at satisfactory levels using the SHuffle T7 *Escherichia coli* cells.

Characterization of antigen–antibody binding. The antigen–antibody binding was characterized using four different experiments. First, it was determined using antigen–antibody Western blot (Fig. 7b), native antigen–antibody Western blot (Fig. S5) and antigen–antibody ELISA (Fig. 7c) and native antigen–antibody ELISA (Fig. S6). These assays were performed using Strep Tactin-HRP conjugated antibody which confirms the binding of the recombinant monoclonal antibodies to the target antigen. All the identified antibody clones were able to specifically bind to the target protein but with different band intensities. The Western blot band intensity of antigen–antibody Western blot was quantified using the ImageJ software (<http://rsb.info.nih.gov/ij/>). Ab23 showed the strongest band intensity with 44.5% followed by Ab14 (23.5%), Ab5 (19.4%), and Ab6 (6.3%).

Western blot and ELISA analysis using the native antigen was also carried out. It was performed to determine the binding capabilities of the newly isolated antibody clones to the native antigen. All four recombinant antibody clones showed binding to *S. stercoralis* native antigen (Figs. S5 and S6). Subsequent analysis was performed using titration ELISA to determine the limit of binding of the antibody clones. The highest binding strength at lowest protein concentration was shown by Ab5 followed by Ab14, Ab6 and Ab23 (Fig. 7d). The range of recombinant monoclonal antibody protein concentration was 10 μg to 0.156 μg /well. Ab5 was able to bind as low as 0.625 μg /well, followed by Ab14 and Ab6 with a binding limit of 1.25 μg /well, and Ab23 at 2.5 μg /well. In addition, the limit of detection of previously isolated recombinant monoclonal antibodies against *Brugia malayi* antigens using the same HAYDEN Filariasis library was included. Ab 5B-XP and Ab 4-R1 are recombinant monoclonal antibodies against BmSXP and BmR1 antigens respectively. BmSXP and BmR1 specific antibodies showed highest binding limits as both antibodies able to bind as low as 0.156 μg /well. Cross-reactivity analysis using ELISA was also carried out. The profiles showed that all four antibody clones were competing for the same epitope on the target antigen highlighting a predominant epitope with high immunoreactivity (Fig. 7e). For all three types of ELISAs performed, the absorbance values of the background protein control were subtracted from the readings of the samples.

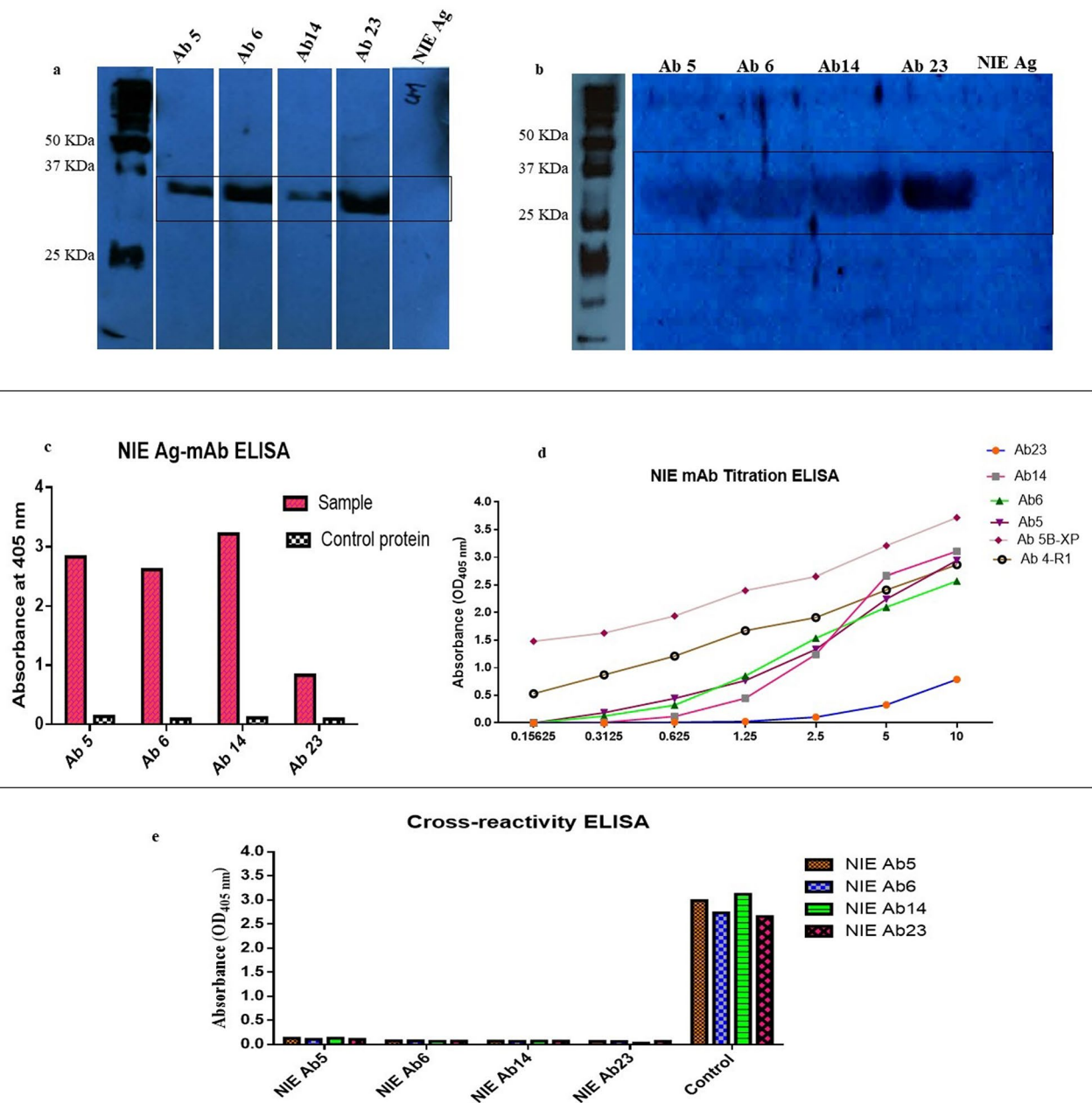


Figure 7. Verification and binding analysis of rNIE specific recombinant monoclonal antibody proteins. (a) Western blot analysis of recombinant monoclonal antibody proteins. (b) Antigen-antibody Western blot analysis. (c) Immunoassay analysis of recombinant monoclonal antibody proteins. (d) Antibody titration ELISA; Ab23, Ab14, Ab6, Ab5 are antibody clones against rNIE, Ab5B-XP and Ab4-R1 are antibody clones against BmSXP and BmR1 antigens respectively (e) Cross-reactivity ELISA. All the binding verifications were detected using Strep Tactin HRP conjugated antibody except for cross-reactivity ELISA was detected using anti-M13 HRP conjugated antibody.

Discussion

Generally, phage display antibody libraries from natural and synthetic repertoires have been useful sources for the generation of antibodies against various targets of interest for medical and research applications¹⁷, and some libraries have been developed into technological platforms¹⁸. One of the key factors for this success is the robustness of the bacteriophage to display diversified antibodies that allows an expanded coverage for a diverse selection of target antigens. The increased understanding of antibody-antigen binding properties has also enabled improved strategies in antibody generation.

The HAYDEN-Filariasis library used in this study is an immune antibody phage display library constructed using LF blood donors⁶. LF is mainly caused by *Wuchereria bancrofti* and *Brugia malayi*, while *S. stercoralis* is the main cause of strongyloidiasis. Both helminth diseases share similarities in terms of immune responses to the parasites. The target used in this study was rNIE, a specific and sensitive protein for detection of *S. stercoralis*

infection that can be produced with high yield and good purity¹⁶. Previously, the HAYDEN-Filariasis immune library was successfully utilized to produce antibodies against two important filarial antigens, BmSXP and BmR1^{6,19}.

An immune antibody library is designed to represent a biased collection of antibody repertoires due to the exposure of the host to a particular pathogen. Therefore, the host antibody repertoire is mainly enriched with disease-specific antibody clones that have undergone affinity maturation processes such as clonal selection and somatic hyper mutation. This increases the likelihood of enriching high affinity antibodies against antigens of the targeted disease. However, when considering the essence of phage antibody library cloning, it is likely that the combinatorial amalgamation of antibody genes will inadvertently invoke unnatural gene combinations to materialize. This has an unprecedented consequence to the way immune library applications are perceived. The additional repertoire generated by unnatural gene combinations can expand the coverage of an immune library to enrich antibodies for non-targeted disease-related antigens. However, the immediacy of predominant target disease-skewed antibody gene repertoire is expected to permit a limited spectrum of non-disease antigens to be covered by the repertoire. Nevertheless, it is conceivable that antigens of closely related diseases to the primary disease of the immune library repertoire could be targeted for antibody generation by using the initial immune library. This is exemplified in a recent study which reported the use of an immune library for the isolation of antibodies against various non-immune antigens²⁰. This is likely due to the presence of large amounts of unimmunized antibody clones, together with biased immunized clones in an immune library repertoire. Hence, it is not surprising that an immune library may be diverse enough to function like a naïve library and be applied to target various antigens. The presence of unimmunized clones emphasizes the broadness of the B-cell memory repertoire in the immune system.

Subsequently, another possible consideration would be the choice of sample collection during immune library constructions. The stage at which patient samples are collected can range from acute, recovering to recovered from an infection. The antibody configuration will vary as a result of the different immune response at the different stages of infection, much like the distribution of different immunoglobulin isotypes in a disease cycle. Therefore, it is conceivable that a sample antibody composition contains a mixed population of activated B cells during encounter with an antigen and also unexposed antibody clones which are naïve B cells. The HAYDEN-filariasis library used in this study consists of blood from apparently asymptomatic individuals who had high density of circulating microfilariae, i.e. ~2600–15,000 larvae per ml of blood. This category of individuals would generally display a broader range of antibody responses compared to symptomatic or asymptomatic individuals with low infection load²¹. In general, intestinal nematode parasites usually stimulate CD4 T cell dependent Th2 type polarized immune responses with potentially overlapping effector mechanisms²².

BLAST analysis of the NIE rAg sequence against *B. malayi* sequences was performed in order to determine the extent of sequence identity. At the nucleotide level both sequences showed no identity. Meanwhile at the protein level there was some similarities ranging from 26.23 to 48.28%. ClustalW alignment was performed to show the conserved and non-conserved regions. This is attached as supplementary file as Fig. S7. The homology between *B. malayi* genome and NIE protein is expected as they are both nematodes. Although low, the percentage of similarity may also explain the ability of the HAYDEN-Filariasis immune library to isolate antibodies against the NIE protein from *S. stercoralis*. This possibility cannot be discounted as a potential reason for the antibody library repertoire to isolate antibodies against proteins from other nematodes.

In total, 30 full length antibody clones representing four unique gene families were identified. It is noticeable that 97% of the isolated clones were dominated by functional variable lambda genes and only 3% of sequences were kappa genes. The preferred V gene family was VH3 for HC and VL6 for LC and the most common gene pairing was IgHV3-VL6. Conversely, antibodies to BmSXP filarial antigen (previously isolated from the same library) showed more functional variable kappa genes compared to functional lambda genes whereas antibodies to BmR1 filarial antigen (also previously isolated from the same library) showed equal representation of variable kappa and lambda genes. The V-gene preference for antibodies raised against the two filarial antigens using the same library showed different gene usage. The BmSXP antibody clones were VH3, similar to that of NIE, while BmR1 antibody clones were mainly represented by VH2. Overwhelming predominance of VH3 antibodies for defence against a variety of bacteria and viruses have been previously reported^{23–25}. The most common gene pairings for LF were IgHV3VK1 for BmSXP and IgHV2VL3 and IgHV2-VK3 for BmR1. This is thus different with rNIE whereby VL6 was dominant.

The gene encoding the kappa gene segments are located on chromosome 2²⁶ while lambda gene segments are encoded on chromosome 22²⁷. Although there are more kappa antibodies in the human peripheral blood, in antigen-selected populations it can vary significantly depending on the class of antibody heavy chain²⁸. Furthermore, over-representation of the lambda or kappa subfamilies is mainly based on the interaction of the antibody gene segments during antigen encounter. Many studies have reported the dominance of lambda subfamilies compared to kappa subfamilies in different infections^{29–31}. For example, antibodies from the mucosal region³² and HIV-specific antibodies³³ showed a strong bias of enriching lambda antibodies. It is because lambda antibodies are more stable when paired with different VH families to produce more stable scFv antibodies. The stability is due to its higher scFv-pIII fusion protein expression levels. In addition, kappa antibodies; showed poor expression in *E. coli* compared to lambda antibodies³⁴.

Complementary determining region (CDR) loops are the most diverse region of an antibody. The length of the CDR determines the topological variation of the binding sites³⁵. The range of CDR length for the HC was 3 to 19 amino acids and LC was 3 to 21 amino acids. The length variations can contribute to the information of the binding site motifs³⁶. For example, deep pocket motif antibodies tend to bind haptens, while peptide binding antibodies usually appear as grooves, and antibody that binds to protein is normally flat³⁷. The CDR-H3 plays the most important role during antibody-antigen binding and the average length of this loop is around 1 and 35 amino acid in a pattern resembling a Gaussian distribution³⁸.

In addition, antibodies raised against the two filarial antigens using the same library showed some variation in terms of CDR length^{6,19}. For heavy chain, the CDR1 length was similar for all three antigens (NIE, BmSXP and BmR1) which was around 6–10 aa. NIE antigen showed similarities with BmSXP antigen for CDR2 and CDR3 regions. The CDR1 ranges from 3–8 aa and CDR3 ranges from 9–20 aa. Conversely for BmR1 the CDR2 and CDR3 lengths were 3 aa and 9–11 aa. Subsequently, light chain CDR1 length for all three antigens was about 6–10 aa. The CDR2 of the NIE showed the shortest length with 3 aa compared to both filarial antigen that was in the range from 7–8 aa.

The docked complex of NIE-Ab in this study has further evidenced the interactions between the antigen and antibody is mainly at the CDR3. Long CDR-H3 are usually associated with autoreactivity as it would have been selected against the target as B cell matures and successfully pass through tolerance checkpoints^{39,40}. However, the role of CDR-L3 should not be overlooked as the physiochemical and structural differences of this region reflect different roles in the immune response²⁸. The stability of the longer CDRs are internally stabilized through the presence of cysteine forming disulfide bridges.

Subsequently, the amino acid usage and polarity of NIE mAbs against previously isolated Filarial antigens mAbs are comparable with some variations. For all the mAbs the aa at CDR3 regions showed highest variations. However, representations of aa were different, for BmR1 antigen the heavy chain of CDR3 was dominated by arginine and alanine. On the other hand, BmSXP antigen showed random equal distribution of all aa but slightly higher serine. Meanwhile for NIE antigen the mAbs showed over representation of glycine, aspartic acid, valine and asparagine. Heavy chain of CDR2 region of BmSXP and NIE showed over representation of serine while BmR1 showed higher aspartic acid. Heavy chain of CDR1 though with the presence of other aa but both filarial antigens showed dominancy of serine and glycine while NIE antigen showed over representation of aspartic acid, phenylalanine and tyrosine. Interestingly for light chain of CDR1 and CDR3 despite with the presence of other aa all three antigens showed a significant increase in serine. CDR2 of the light chain for BmR1 showed abundance of glycine followed by BmSXP with proline and serine meanwhile NIE showed higher representation of glycine and asparagine. Looking at the polarity of all the three antigens, in total all showed higher representation of neutral and small aa due to rich in aa such as serine and glycine.

By using the crystal structure of Ves v 2⁴¹ as the template in comparative modelling, the built NIE 3D structure has the best evaluation scores for its global stereochemistry compared with the structures predicted other approaches. This could be due to NIE and Ves v 2 having sequence similarity of nearly 40%. In addition, the experimental data also showed that NIE shares epitope with Ves v 5⁴¹. The physiochemical properties calculations of Ab6 with the lowest interface area and the only negatively charged antibody at pH7 could have explained the weakest intensity of Ab6 band observed in the Western blot. The antibody clones were further analysed using IMGT Colliers de Perles tool. The bioinformatic analysis of the CDR regions permits the visualization of the aa which are important for the 3D structural configuration and to delineate the standardized framework regions. For example, the length of the helix represented in IMGT Colliers de Perles becomes a crucial information in the domain characterization and are particularly useful for sequence-structure analysis, antibody engineering, visualization and comparison of positions for mutations, polymorphisms and contact analysis⁴². This structural information would be useful for antibody refinement in the future studies.

All four recombinant monoclonal antibody proteins were successfully expressed, purified and the binding analysis conducted confirmed the specificity towards the NIE target antigen. An *in-house* unrelated protein control carrying the same tags as rNIE was used when performing all types of ELISA. In addition, the expression and purification tag for monoclonal antibody proteins in this study was anti-His. However, the detection tag for polyclonal and monoclonal ELISA was anti-M13 horseradish peroxidase (HRP) and for binding assays was Strep Tactin- HRP. Different tags were used for detection in order to ensure correct protein detection.

The titration ELISA showed Ab5, Ab6 and Ab14 exhibiting approximately similar binding strength while Ab23 showed lowest strength of binding. The native antigen Western blot also showed the capabilities of all four antibody clones binding the *S. stercoralis* native antigen. Furthermore, cross-reactivity ELISA shows the ability of all four monoclonal antibody clones with different V-gene family usage and light chain pairing competing for a similar epitope. This is evidence of the diversity and flexibility of antibody molecules to adapt different structural motifs to bind to a similar epitope. This results in the presence of different antibodies against a single epitope that affects the evolution of the pathogen for host recognition. In addition, multiple antibodies targeting one binding site of an antigen allows for rapid recognition and destruction by other components of the immune system. In the present study, the presence of multiple antibodies against a single dominant epitope highlights the immunoreactivity of the epitope for the rNIE.

Existing commercial diagnostics to detect *Strongyloides* infection in blood/serum/plasma are antibody-detection assays. These assays are sensitive but may detect past or treated infections and may not detect early infection. Antigen detection assay can address the above limitations, thus it can complement existing antibody-detection tests. There are only a few reports on antigen detection for strongyloidiasis, and one of them used recombinant antibody to *Strongyloides* HSP60 protein⁴³. The antiHSP60 scFv fragment was reported to detect immune complexes in strongyloidiasis patients' serum samples in an ELISA. Subsequently, in a follow-up study, the ELISA detected six times (27.27%) more positive samples as compared with conventional parasitological techniques (4.54%)⁴⁴. In the future, the recombinant antibody to rNIE from the present study can be used to develop an antigen detection assay; subsequently, the diagnostic sensitivity and specificity can be compared with reported antigen-detection ELISA.

In conclusion, the present study provides a small glimpse into the *Strongyloides*- specific antibodies derived from non-target disease specific antibody library. This study also highlights the depth of the antibody repertoire even from a skewed population to yield antibodies against a disease target that was close to the parent disease where the library repertoire was derived from. The sequence-structure analysis was deciphered in order to characterize the newly isolated antibody clones. Analysis of the antibody physiochemical properties highlighted

the positively charged molecule at pH7 and interface area could determine the amount of bound antibody with NIE observed in the Western blot. In future, further structural analysis especially epitope mapping of the antibody clones is required to elucidate the underlying interactions and identify the key epitope of NIE. It is also important to note that the antibody clones identified in this study are potentially useful for various biomedical applications in the near future.

Methods

Recombinant protein preparation. A single colony of bacteria expressing the rNIE was inoculated into 100 mL Terrific broth (TB) media containing 100 µg/mL ampicillin, and cultured overnight at 37 °C, 200 rpm. The next day, 10 mL of the overnight culture was inoculated into 4 × 500 mL TB supplemented with 100 µg/mL of ampicillin and incubated at 37 °C with agitation at 200 rpm until OD₆₀₀ of 0.6–0.8 was reached. Subsequently, 0.5 mM of isopropyl-B-D-thiogalactopyranoside (IPTG) was added into the culture and incubated for 4 h at 28 °C. The culture was harvested by centrifugation at 10,000×g for 10 min at 4 °C. The cell pellet was resuspended in cold lysis buffer (50 mM NaH₂PO₄, 500 mM NaCl and 10 mM imidazole) containing lysozyme (0.5 mg/mL) and protease inhibitors. The mixture was incubated on ice for 30 min. The mixture was then lysed mechanically using a French press (1500 psi, 1 min). The lysate was then centrifuged at 10,000×g for 5 min at 4 °C to separate the supernatant from the cell debris. DNase I was added at a concentration of 0.5 µg/ml to the supernatant and incubated on ice for 15 min. This was followed by another round of centrifugation at 10,000×g, 4 °C for 30 min. The supernatant was then collected and filtered through a 0.45 µm filter membrane to remove debris that may block the purification column. The filtered supernatant was mixed with the Ni-NTA slurry resin (cComplete His-Tag purification resin, Roche, Switzerland). The resin-lysate mixture was incubated at 4 °C for 1 h with gentle rotation. The wash column was preloaded with resin-lysate mixture and washing was performed using wash buffers containing 20 mM, 30 mM, and 40 mM of imidazole. The target protein was eluted using the elution buffer containing 250 mM imidazole and the fractions were analysed through SDS-PAGE. The protein-containing fractions were pooled, and buffer exchanged into phosphate-buffered saline (PBS, pH 7.4) using Amicon Ultra 10 mL centrifugal filter (Merck, Germany). The presence of the target protein was confirmed via Western blotting using the anti-His6-tagged secondary antibody at 1:3000 dilution. All chemicals used are from Merck (Germany) unless stated otherwise.

Biopanning. Biopanning was carried out as previously reported⁶ using a human helminth scFv phage display library (HAYDEN-immune filariasis) against the rNIE. The construction of the library was previously described and have been applied for the isolation of scFv clones against other target antigens^{6,19}.

Phage ELISA DNA sequencing. Polyclonal and monoclonal phage ELISA were performed as previously reported⁶. The positive clones from the monoclonal ELISA were grown at 37 °C and 200 rpm, overnight (~ 16 h). The cell pellets were harvested, and plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen, Germany), then sent for sequencing (FIRST Base Laboratories Sdn Bhd, Malaysia). The sequencing results were analyzed using IMG/QUEST bioinformatics tool available at IMG, the International ImMunoGeneTics information system^{45–47}. An *in-house* unrelated protein control that can be detected using anti-M13 HRP was included as positive control and for background control the wells were coated with BSA.

Recombinant monoclonal antibody protein expression, purification and verification. Plasmids of positive clones from monoclonal phage ELISA were subcloned into pET 51(b) + vector for expression in SHuffle T7 *Escherichia coli* cells (NEB, USA) and expression was performed in 1 L 2-YT broth supplemented with 100 µg/mL ampicillin and 0.2% glucose. The culture was grown at 37 °C, 200 rpm until the OD₆₀₀ reached 0.6–0.7, then induced with 1 mM IPTG and further cultured for 16 h at 25 °C with shaking at 180 rpm. The pellet was harvested and resuspended in ice cold lysis buffer (50 mM NaH₂PO₄, 500 mM NaCl, 10 mM) containing imidazole and 0.5 mg/mL lysozyme). The mixture was then incubated on ice for 30 min and then subjected to cell breakage. It was performed using a sonicator for 4 cycles at 30 s/cycle with an output of 4.5 Hz. The disrupted cells were centrifuged at 10,000×g for 30 min at 4 °C. The supernatant was carefully collected and filtered through a 0.45 µm filter membrane.

The purification was carried out using C-terminal His-tag with a nitrilotriacetic acid (Ni-NTA) purification column (Qiagen, Germany) according to the manufacturer's instructions. The purified proteins were then verified by SDS PAGE and Western blotting was performed using Strep Tactin-HRP conjugated antibody (1:3000). The bands were developed on film using Pierce ECL Plus substrate (Thermo Scientific, USA).

Antigen-antibody binding assay. Antigen-antibody Western blot was carried out by running the rNIE on 12% SDS PAGE at 100 V for 1 h until the front dye reached the bottom of the gel. Then protein was transferred onto the nitrocellulose membrane using semi dry transfer at 12 V for 30 min. The membrane was then cut into stripes and incubated with the respective monoclonal antibody proteins at 4 °C overnight followed by incubation with Strep Tactin-HRP conjugated antibody (1:3000). The bands were developed on film using Pierce ECL Plus substrate (Thermo Scientific, USA).

For native antigen Western blot, the crude lysate antigen was first prepared. Briefly, a 50X protease inhibitor cocktail tablet (Roche, Switzerland) was dissolved in 2 mL of deionized water to produce a 25X protease inhibitor solution. A volume of 60 µL of the solution was added to 1.5 mL of *S. stercoralis* infective larvae and incubated on ice for 30 min. This mixture was then lysed by sonication on ice using a microtip probe for 3 min (30 s ON, 30 s OFF, 4.5 Hz). The lysed mixture was transferred to a cryovial and dipped into liquid nitrogen for 30 s (snap freeze) and then immediately thawed in a water bath for 20 min. The freeze-thawed mixture was centrifuged at

14,000 rpm for 5 min and the supernatant (soluble larval lysate) was collected. The protein concentration of the lysate was determined by RCDC Protein Assay (BioRad, USA) and stored at -20 °C until further use.

The native antigen Western blot was carried out similarly as the rNIE antigen–antibody Western blot as described above but with slight modifications. The modifications are 100 µg of *S. stercoralis* infective larvae protein was separated on SDS-PAGE and dilutions of the Strep-Tactin HRP conjugate used was 1:2000. Polyclonal anti-NIE was used as positive control and detected using goat-anti-rabbit HRP conjugate (1:5000).

Antigen–antibody ELISA was performed using rNIE and crude lysate antigens according to previously used method with slight modifications¹⁹. ELISA was carried out by coating 10 µg/well of rNIE or 50 µg/mL and 100 µg/mL of crude lysate antigens on Maxisorb ELISA microtiter plate (Nunc, New York, NY, USA) with 1XPBS, pH 7.4 overnight. The coated wells were washed with PBST (PBS with 0.5% Tween 20) and blocked with MPBST (2% skim milk in PBST) for 1 h at 37 °C, 600 rpm. After washing, each well was incubated with the respective monoclonal antibody proteins for 2 h at room temperature, 600 rpm. The plate was then washed and incubated with Strep Tactin HRP conjugate antibody (1:3000) for 1 h at RT, 600 rpm. The plate was washed three times with PBST and ABTS substrate was added and left to incubate in the dark for 30 min. The absorbance value was read at 405 nm using the SkanIT absorbance reader (Thermo Scientific, USA).

Titration ELISA. Antibody titration ELISA was performed to determine the limit of binding of the identified anti-rNIE monoclonal antibody clones. The ELISA was performed similarly as described above but with one modification. After the second washing step, each well was incubated with the scFv protein at concentrations ranging from 10 µg/well to 0.156 µg/well.

Cross-reactivity ELISA. Cross-reactivity ELISA was performed as previously reported⁶. Briefly, the plate was coated with 10 µg/well of rNIE. Then the wells were washed, blocked, and incubated with 50 µg/well of the purified monoclonal antibody proteins. After 1 h, the wells were washed and 10⁵ of scFv phage was added accordingly and incubated for 1 h at room temperature (RT). The wells were then washed and incubated with 100 µL anti-M13 HRP conjugated antibody (1:5000 dilution) for 1 h at RT. Then the plate was washed and ABTS substrate was added. The absorbance value was measured at 405 nm using a SkanIT absorbance reader (Thermo Scientific, USA) after 30 min incubation in the dark. The absorbance values from the control wells were subtracted from test wells to obtain total absorbance.

Structural analysis. The secondary structure of NIE protein was predicted using secondary structure prediction servers- PsiPred⁴⁸, JPred⁴⁹ and Spider⁵⁰. The 3D structure of NIE was modelled via comparative modelling⁴¹, threading (ITasser)⁵¹ and DMPFold⁵² and ab initio (Robetta)⁵³ approaches. The modelled structures from the above mentioned approaches were then evaluated for their compatibility using Procheck⁵⁴, MolProbity⁵⁵, Vadar⁵⁶, Verify3D⁵⁷, Errat⁵⁸ and PROVE⁵⁹. The best model from the structural evaluation results was selected for docking with the scFvs.

On the other hand, the 3D model of the four scFvs (Ab5, Ab6, Ab14, Ab23) were built via comparative modelling (MODELLER v9.20; template PDB id 1RZI, 5WKZ, 4KQ3, 2DD8, 6RCQ, 5J66 for Ab5; 1W72, 3GIZ, 4QF1, 6MG4, 6E63, 4RIR for Ab6; 6MI2, 4KMT, 5WK2, 6GHG, 5I17, 6B9J for Ab14; 3GIZ, 1W72, 4QF1, 5BV7, 5Y9J, 4OD2 for Ab23). Physicochemical characteristics and electrostatic distribution for the scFvs were analysed prior to docking simulation by the jsPISA v.2.0.4⁶⁰. The isoelectric point was calculated using Protein Calculator v3.3 (<http://protcalc.sourceforge.net/>; accessed 6th October 2019) and isoelectric point calculator⁶¹. Delphi program⁶² with atomic charges and radii parameters were set to AMBER force fields to calculate the electrostatic free energies and pKa of scFv CDRs titratable residues.

Molecular docking was applied to elucidate the interactions of the scFvs with NIE. Prior to docking simulation, the epitope for NIE was predicted by consensus results from three different independent predictions, namely BepiPred⁶³, CBTope⁶⁴ and Ellipro⁶⁵. During docking simulation, only scFv CDRs residues and the predicted NIE epitopes were set as the interface residues. The docking simulation was performed using ClusPro Antibody mode⁶⁶ available in ClusPro 2.0 server⁶⁷. The scFv-NIE complex with the most negative binding free energy (ΔG_{bind}) obtained from the most populated cluster with root mean square deviation (RMSD) tolerance of 2.0 Å was selected as the representative structure for further interactions analysis using jsPISA v.2.0.4. All scFv-NIE complexes were depicted using PyMOL molecular visualization system⁶⁸.

The two-dimensional graphical representations of monoclonal antibody protein domains were constructed using IMGT Collier de Perles tool⁶⁹. The CDR3 amino acid position and composition was further analysed using the Kabat numbering scheme and presented as heat map using heatmapper.ca online software⁷⁰.

Received: 26 March 2020; Accepted: 4 January 2021

Published online: 28 January 2021

References

- Smith, G. P. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317. <https://doi.org/10.1126/science.4001944> (1985).
- Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23**, 1105–1116. <https://doi.org/10.1038/nbt1126> (2005).
- Zhou, H., Zhang, Y. L., Lu, G., Ji, H. & Rodi, C. P. Recombinant antibody libraries and selection technologies. *New Biotechnol.* **28**, 448–452. <https://doi.org/10.1016/j.nbt.2011.03.013> (2011).
- Ponsel, D., Neugebauer, J., Ladetzki-Baehs, K. & Tissot, K. High affinity, developability and functional size: The holy grail of combinatorial antibody library generation. *Molecules* **16**, 3675–3700. <https://doi.org/10.3390/molecules16053675> (2011).

5. Lai, J. Y., Lim T. S. Infectious disease antibodies for biomedical applications: A mini review of immune antibody phage library repertoire. *Int. J. Biol. Macromol.* **163**, 640–648. <https://doi.org/10.1016/j.ijbiomac.2020.06.268> (2020).
6. Rahumatullah, A., Ahmad, A., Noordin, R. & Lim, T. S. Delineation of BmSXP antibody V-gene usage from a lymphatic filariasis based immune scFv antibody library. *Mol. Immunol.* **67**, 512–523. <https://doi.org/10.1016/j.molimm.2015.07.040> (2015).
7. Hotez, P. J. *et al.* Helminth infections: The great neglected tropical diseases. *J. Clin. Investig.* **118**, 1311–1321. <https://doi.org/10.1172/JCI34261> (2008).
8. Anthony, R. M., Rutitzky, L. I., Urban, J. F. Jr., Stadecker, M. J. & Gause, W. C. Protective immune mechanisms in helminth infection. *Nat. Rev. Immunol.* **7**, 975–987. <https://doi.org/10.1038/nri2199> (2007).
9. Harris, N. & Gause, W. C. To B or not to B: B cells and the Th2-type immune response to helminths. *Trends Immunol.* **32**, 80–88. <https://doi.org/10.1016/j.it.2010.11.005> (2011).
10. Moreau, E. & Chauvin, A. Immunity against helminths: Interactions with the host and the intercurrent infections. *J. Biomed. Biotechnol.* **2010**, 428593. <https://doi.org/10.1155/2010/428593> (2010).
11. Iriemenam, N. C., Sanyaolu, A. O., Oyibo, W. A. & Fagbenro-Beyioku, A. F. *Strongyloides stercoralis* and the immune response. *Parasitol. Int.* **59**, 9–14. <https://doi.org/10.1016/j.parint.2009.10.009> (2010).
12. Bisoffi, Z. *et al.* *Strongyloides stercoralis*: A plea for action. *PLoS Negl. Trop. Dis.* **7**, e2214. <https://doi.org/10.1371/journal.pntd.0002214> (2013).
13. Marcos, L. A., Terashima, A., Dupont, H. L. & Gotuzzo, E. *Strongyloides* hyperinfection syndrome: An emerging global infectious disease. *Trans. R. Soc. Trop. Med. Hyg.* **102**, 314–318. <https://doi.org/10.1016/j.trstmh.2008.01.020> (2008).
14. Ravi, V., Ramachandran, S., Thompson, R. W., Andersen, J. F. & Neva, F. A. Characterization of a recombinant immunodiagnostic antigen (NIE) from *Strongyloides stercoralis* L3-stage larvae. *Mol. Biochem. Parasitol.* **125**, 73–81. [https://doi.org/10.1016/s0166-6851\(02\)00214-1](https://doi.org/10.1016/s0166-6851(02)00214-1) (2002).
15. Ramanathan, R. *et al.* A luciferase immunoprecipitation systems assay enhances the sensitivity and specificity of diagnosis of *Strongyloides stercoralis* infection. *J. Infect. Dis.* **198**, 444–451. <https://doi.org/10.1086/589718> (2008).
16. Rascoe, L. N. *et al.* Development of Ss-NIE-1 recombinant antigen based assays for immunodiagnosis of strongyloidiasis. *PLoS Negl. Trop. Dis.* **9**, e0003694. <https://doi.org/10.1371/journal.pntd.0003694> (2015).
17. Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J. & Laustsen, A. H. Basics of antibody phage display technology. *Toxins* <https://doi.org/10.3390/toxins10060236> (2018).
18. Almagro, J. C., Pedraza-Escalona, M., Arrieta, H. I. & Perez-Tapia, S. M. Phage display libraries for antibody therapeutic discovery and development. *Antibodies* <https://doi.org/10.3390/antib8030044> (2019).
19. Rahumatullah, A., Abdul Karim, I. Z., Noordin, R. & Lim, T. S. Antibody-based protective immunity against helminth Infections: Antibody phage display derived antibodies against BmR1 antigen. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18112376> (2017).
20. Moon, S. A. *et al.* Antibodies against non-immunizing antigens derived from a large immune scFv library. *Mol. Cells.* **31**, 509–513. <https://doi.org/10.1007/s10059-011-2268-8> (2011).
21. Babu, S. & Nutman, T. B. Immunopathogenesis of lymphatic filarial disease. *Semin. Immunopathol.* **34**, 847–861. <https://doi.org/10.1007/s00281-012-0346-4> (2012).
22. Sorobetea, D., Svensson-Frej, M. & Grecis, R. Immunity to gastrointestinal nematode infections. *Mucosal Immunol.* **11**, 304–315. <https://doi.org/10.1038/mi.2017.113> (2018).
23. Baxendale, H. E. *et al.* Natural human antibodies to pneumococcus have distinctive molecular characteristics and protect against pneumococcal disease. *Clin. Exp. Immunol.* **151**, 51–60. <https://doi.org/10.1111/j.1365-2249.2007.03535.x> (2008).
24. Houimel, M. The analysis of VH and VL genes repertoires of Fab library built from peripheral B cells of human rabies virus vaccinated donors. *Hum. Immunol.* **75**, 745–755. <https://doi.org/10.1016/j.humimm.2014.05.005> (2014).
25. Kearns-Jonker, M. *et al.* Use of molecular modeling and site-directed mutagenesis to define the structural basis for the immune response to carbohydrate xenoantigens. *BMC Immunol.* **8**, 3. <https://doi.org/10.1186/1471-2172-8-3> (2007).
26. Malcolm, S. *et al.* Localization of human immunoglobulin kappa light chain variable region genes to the short arm of chromosome 2 by in situ hybridization. *Proc. Natl. Acad. Sci. USA* **79**, 4957–4961. <https://doi.org/10.1073/pnas.79.16.4957> (1982).
27. Barbie, V. & Lefranc, M. P. The human immunoglobulin kappa variable (IGKV) genes and joining (IGJ) segments. *Exp. Clin. Immunogenet.* **15**, 171–183. <https://doi.org/10.1159/000019068> (1998).
28. Townsend, C. L. *et al.* Significant differences in physicochemical properties of human immunoglobulin kappa and lambda CDR3 regions. *Front. Immunol.* **7**, 388. <https://doi.org/10.3389/fimmu.2016.00388> (2016).
29. Schofield, D. J. *et al.* Application of phage display to high throughput antibody generation and characterization. *Genome Biol.* **8**, R254. <https://doi.org/10.1186/gb-2007-8-11-r254> (2007).
30. Lloyd, C. *et al.* Modelling the human immune response: Performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens. *Protein Eng. Des. Sel.* **22**, 159–168. <https://doi.org/10.1093/protein/gzn058> (2009).
31. Hust, M. *et al.* A human scFv antibody generation pipeline for proteome research. *J. Biotechnol.* **152**, 159–170. <https://doi.org/10.1016/j.jbiotec.2010.09.945> (2011).
32. Mole, C. M., Bene, M. C., Montagne, P. M., Seilles, E. & Faure, G. C. Light chains of immunoglobulins in human secretions. *Clin. Chim. Acta* **224**, 191–197. [https://doi.org/10.1016/0009-8981\(94\)90185-6](https://doi.org/10.1016/0009-8981(94)90185-6) (1994).
33. Sajadi, M. M. *et al.* Lambda light chain bias associated with enhanced binding and function of anti-HIV env glycoprotein antibodies. *J. Infect. Dis.* **213**, 156–164. <https://doi.org/10.1093/infdis/jiv448> (2016).
34. Kim, D. Y. *et al.* Antibody light chain variable domains and their biophysically improved versions for human immunotherapy. *mAbs* **6**, 219–235. <https://doi.org/10.4161/mabs.26844> (2014).
35. Kunik, V. & Ofra, Y. The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops. *Protein Eng. Des. Sel.* **26**, 599–609. <https://doi.org/10.1093/protein/gzt027> (2013).
36. Dunbar, J., Knapp, B., Fuchs, A., Shi, J. & Deane, C. M. Examining variable domain orientations in antigen receptors gives insight into TCR-like antibody design. *PLoS Comput. Biol.* **10**, e1003852. <https://doi.org/10.1371/journal.pcbi.1003852> (2014).
37. Marchuk, D., Drumm, M., Saulino, A. & Collins, F. S. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19**, 1154–1154. <https://doi.org/10.1093/nar/19.5.1154> (1991).
38. Finlay, W. J. & Almagro, J. C. Natural and man-made V-gene repertoires for antibody discovery. *Front. Immunol.* **3**, 342. <https://doi.org/10.3389/fimmu.2012.00342> (2012).
39. Rackovsky, S. & Scheraga, H. A. On the information content of protein sequences. *J. Biomol. Struct. Dyn.* **28**(593–594), 669–674. <https://doi.org/10.1080/073911011010524957> (2011).
40. Larimore, K., McCormick, M. W., Robins, H. S. & Greenberg, P. D. Shaping of human germline IgH repertoires revealed by deep sequencing. *J. Immunol.* **189**, 3221–3230. <https://doi.org/10.4049/jimmunol.1201303> (2012).
41. Ravi, V., King, T. P., Andersen, J. F., Nutman, T. B. & Neva, F. A. *Strongyloides stercoralis* recombinant NIE antigen shares epitope with recombinant Ves v 5 and Pol a 5 allergens of insects. *Am. J. Trop. Med. Hyg.* **72**, 549–553 (2005).
42. Levenhagen, M. A. *et al.* Structural and functional characterization of a novel scFv anti-HSP60 of *Strongyloides* sp. *Sci. Rep.* **5**, 10447. <https://doi.org/10.1038/srep10447> (2015).
43. Miguel, C. B. *et al.* scFv against HSP60 of *Strongyloides* sp. and its application in the evaluation of parasite frequency in the elderly. *Dis Markers* **2020**, 4086929. <https://doi.org/10.1155/2020/4086929> (2020).
44. Vlachakis, D., Feidakis, C., Megalooikonomou, V. & Kossida, S. IMGT/Collier-de-Perles: A two-dimensional visualization tool for amino acid domain sequences. *Theor. Biol. Med. Model.* **10**, 14. <https://doi.org/10.1186/1742-4682-10-14> (2013).

45. Retter, I., Althaus, H. H., Munch, R. & Muller, W. VBASE2, an integrative V gene database. *Nucleic Acids Res.* **33**, D671–674. <https://doi.org/10.1093/nar/gki088> (2005).
46. Brochet, X., Lefranc, M. P. & Giudicelli, V. IMGT/V-QUEST: The highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* **36**, W503–508. <https://doi.org/10.1093/nar/gkn316> (2008).
47. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202. <https://doi.org/10.1006/jmbi.1999.3091> (1999).
48. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–394. <https://doi.org/10.1093/nar/gkv332> (2015).
49. Heffernan, R., Yang, Y., Paliwal, K. & Zhou, Y. Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. *Bioinformatics* **33**, 2842–2849. <https://doi.org/10.1093/bioinformatics/btx218> (2017).
50. Fiser, A. & Sali, A. Modeller: Generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374**, 461–491. [https://doi.org/10.1016/S0076-6879\(03\)74020-8](https://doi.org/10.1016/S0076-6879(03)74020-8) (2003).
51. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **9**, 40. <https://doi.org/10.1186/1471-2105-9-40> (2008).
52. Greener, J. G., Kandathil, S. M. & Jones, D. T. Deep learning extends de novo protein modelling coverage of genomes using iteratively predicted structural constraints. *Nat. Commun.* **10**, 3977. <https://doi.org/10.1038/s41467-019-11994-0> (2019).
53. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–531. <https://doi.org/10.1093/nar/gkh468> (2004).
54. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486. <https://doi.org/10.1007/BF00228148> (1996).
55. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21. <https://doi.org/10.1107/S0907444909042073> (2010).
56. Willard, L. *et al.* VADAR: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **31**, 3316–3319. <https://doi.org/10.1093/nar/gkg565> (2003).
57. Eisenberg, D., Luthy, R. & Bowie, J. U. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **277**, 396–404. [https://doi.org/10.1016/S0076-6879\(97\)77022-8](https://doi.org/10.1016/S0076-6879(97)77022-8) (1997).
58. Colovos, C. & Yeates, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519. <https://doi.org/10.1002/pro.5560020916> (1993).
59. Pontius, J., Richelle, J. & Wodak, S. J. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* **264**, 121–136. <https://doi.org/10.1006/jmbi.1996.0628> (1996).
60. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* **43**, W314–319. <https://doi.org/10.1093/nar/gkv314> (2015).
61. Kozłowski, L. P. IPC—isoelectric point calculator. *Biol. Direct* **11**(1), 55 (2016).
62. Wang, L., Zhang, M. & Alexov, E. DelPhiPKa web server: Predicting pKa of proteins, RNAs and DNAs. *Bioinformatics* **32**, 614–615. <https://doi.org/10.1093/bioinformatics/btv607> (2016).
63. Jespersen, M. C., Peters, B., Nielsen, M. & Marcatili, P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **45**, W24–W29. <https://doi.org/10.1093/nar/gkx346> (2017).
64. Ansari, H. R. & Raghava, G. P. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res.* **6**, 6. <https://doi.org/10.1186/1745-7580-6-6> (2010).
65. Ponomarenko, J. *et al.* ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* **9**, 514. <https://doi.org/10.1186/1471-2105-9-514> (2008).
66. Brenke, R. *et al.* Application of asymmetric statistical potentials to antibody-protein docking. *Bioinformatics* **28**, 2608–2614. <https://doi.org/10.1093/bioinformatics/bts493> (2012).
67. Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **32**, W96–99. <https://doi.org/10.1093/nar/gkh354> (2004).
68. DeLano, W. P. *The PyMOL Molecular Graphic System* (DeLano Scientific LLC, Palo Alto, 2008).
69. Ehrenmann, F., Giudicelli, V., Duroux, P. & Lefranc, M. P. IMGT/Collier de Perles: IMGT standardized representation of domains (IG, TR, and IgSF variable and constant domains, MH and MhSF groove domains). *Cold Spring Harb. Protoc.* **726–736**, 2011. <https://doi.org/10.1101/pdb.prot5635> (2011).
70. Babicki, S. *et al.* Heatmapper: Web-enabled heat mapping for all. *Nucleic Acids Res.* **44**, W147–153. <https://doi.org/10.1093/nar/gkw419> (2016).

Acknowledgements

This study was supported by the Higher Institution Centre of Excellence (HICoE) Grant, Malaysian Ministry of Higher Education (NO. 311/CIPPM/44001005) and Universiti Sains Malaysia RUI Grant (NO. 1001/CIPPM/8012323). T.S.L. would like to acknowledge financial support from the Malaysian Ministry of Higher Education through the Fundamental Research Grant Scheme (203/CIPPM/6711658—Reference Code: FRGS/1/2018/STG05/USM/02/2).

Author contributions

T.S.L., R.N. and A.R. designed the research project. A.R., D.B., and Z.B. performed the experiments, collected the data and analyzed the data. Y.Y.L., and Y. S.C. performed the bioinformatic analysis. A.R., T.S.L., and R.N. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82125-3>.

Correspondence and requests for materials should be addressed to T.S.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021