Discovery of Marburg virus neutralizing antibodies from virus-naïve human antibody repertoires using large-scale structural predictions

Nina G. Bozhanova[a,b,1], Amandeep K. Sangha[a,b,1], Alexander M. Sevy[b,c], Pavlo Gilchuk[c], Kai Huang[d,e], Rachel S. Nargi[c], Joseph X. Reidy[c], Andrew Trivette[c], Robert H. Carnahan[c,f], Alexander Bukreyev[d,e,g], James E. Crowe Jr[c,f,h], and Jens Meiler[a,b,i,2]

[a]Department of Chemistry, Vanderbilt University, Nashville, TN 37235; [b]Center for Structural Biology, Vanderbilt University, Nashville, TN 37235; [c]Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, TN 37232; [d]Department of Pathology, University of Texas Medical Branch, Galveston, TX 77555; [e]Galveston National Laboratory, University of Texas Medical Branch, Galveston, TX 77550; [f]Department of Pediatrics, Vanderbilt University Medical Center, Nashville TN 37232; [g]Department of Microbiology & Immunology, University of Texas Medical Branch, Galveston, TX 77555; [h]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232; and [i]Institute for Drug Discovery, Leipzig University Medical School, Leipzig, SAC 04103, Germany

Edited by Daniel W. Kulp, Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, CA, and accepted by Editorial Board Member Robert Langer October 22, 2020 (received for review December 26, 2019)

Marburg virus (MARV) disease is lethal, with fatality rates up to 90%. Neutralizing antibodies (Abs) are promising drug candidates to prevent or treat the disease. Current efforts are focused in part on vaccine development to induce such MARV-neutralizing Abs. We analyzed the antibody repertoire from healthy unexposed and previously MARV-infected individuals to assess if naïve repertoires contain suitable precursor antibodies that could become neutralizing with a limited set of somatic mutations. We computationally searched the human Ab variable gene repertoire for predicted structural homologs of the neutralizing Ab MR78 that is specific to the receptor binding site (RBS) of MARV glycoprotein (GP). Eight Ab heavy-chain complementarity determining region 3 (HCDR3) loops from MARV-naïve individuals and one from a previously MARV-infected individual were selected for testing as HCDR3 loop chimeras on the MR78 Ab framework. Three of these chimerized antibodies bound to MARV GP. We then tested a full-length native Ab heavy chain encoding the same 17-residue-long HCDR3 loop that bound to the MARV GP the best among the chimeric Abs tested. Despite only 57% amino acid sequence identity, the Ab from a MARV-naïve donor recognized MARV GP and possessed neutralizing activity against the virus. Crystallization of both chimeric and full-length native heavy chain-containing Abs provided structural insights into the mechanism of binding for these types of Abs. Our work suggests that the MARV GP RBS is a promising candidate for epitope-focused vaccine design to induce neutralizing Abs against MARV.

Marburg virus | neutralizing antibodies | computational antibody function prediction | P3SM approach | Rosetta

With the advent of high-throughput immune repertoire sequencing, the number of available human antibody (Ab) sequences is exploding rapidly from thousands to billions. These datasets provide a resource for understanding the natural process of Ab maturation through somatic mutations, quick identification of novel functional Abs, and engineering of improved Abs. Ultimately, such Ab repertoires may provide or add templates for developing epitope-focused (1) and germline-targeting (2, 3) vaccines. While Ab function sometimes can be predicted from sequence homology, often Abs of very different sequence have the same function by virtue of adopting a similar structure. However, despite substantial progress in computational methods of modeling Abs (4, 5) and Ab–antigen interactions (6–10), and the use of high-performance computing, it is still beyond the available computational resources to predict and study the structure of billions of Abs one by one. Therefore, in order to take advantage of the rapidly increasing Ab sequence databases, it is essential to find more effective ways of relating Ab sequence to function.

The recently proposed position-specific structure scoring matrix (P3SM) approach is a new computational method specifically designed for rapid screening of large Ab sequence libraries (11, 12). This approach aims to predict whether a given Ab sequence can adopt the desired 3D conformation and thus correctly place critical-for-activity functional groups. This prediction is based on modeling of a small subset of structures using Rosetta (13) followed by evaluation of the compatibility of each of 20 amino acids in each of the analyzed positions with the desired structure and function. The resulting P3SM then can be used for rapid screening of the remaining Ab sequences. The best-scoring candidate sequences are fed back into Rosetta for a

Significance

Marburg virus (MARV) is a member of the Filoviridae family that causes severe hemorrhagic fevers with high fatality rates in humans. Despite the major efforts made to develop an effective vaccine after the most recent filovirus outbreak, still no Food and Drug Administration-licensed MARV disease vaccines exist. A good vaccine should be capable of eliciting neutralizing antibodies in MARV-naïve humans. This requires preexistence of antibodies with some affinity to the antigen. Here we used a large-scale computational antibody structure prediction approach to identify whether the epitope of the human anti-MARV monoclonal antibody MR78 can be recognized by antibodies from MARV-naïve humans and thus whether it is a promising candidate for vaccine design.

Author contributions: A.K.S., R.H.C., J.E.C., and J.M. designed research; N.G.B., A.K.S., A.M.S., P.G., K.H., and R.S.N. performed research; P.G. and K.H. contributed new reagents/analytic tools; N.G.B., A.K.S., A.M.S., P.G., K.H., J.X.R., A.T., A.B., J.E.C., and J.M. analyzed data; and N.G.B., J.E.C., and J.M. wrote the paper.

Competing interest statement: J.E.C. is on the Scientific Advisory Boards of CompuVax and Meissa Vaccines, a recipient of previous unrelated research grants from Moderna and Sanofi, and founder of IDBiologics, Inc. Vanderbilt University has applied for a patent that includes the original MR78 Ab.

This article is a PNAS Direct Submission. D.W.K. is a guest editor invited by the Editorial Board.

[1]N.G.B. and A.K.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: jens@meilerlab.org.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1922654117/-/DCSupplemental.

First published November 23, 2020.

detailed energetic analysis and prioritization for experimental validation.

The human monoclonal Ab MR78 was isolated previously from a B cell in the peripheral blood of an otherwise healthy individual with a prior history of naturally acquired Marburg virus (MARV) infection (14). A crystal structure of MARV glycoprotein (GP) in complex with MR78 (Protein Data Bank [PDB] ID 5UQY) revealed that the Ab heavy-chain complementarity determining region 3 (HCDR3) contacts the receptor binding domain (RBD) on MARV GP that interacts with the natural receptor on human cells (Niemann–Pick disease, type C1 [NPC1] protein) (15). Here, we applied the P3SM approach to search for structurally homologous Abs to MR78 based on this Ab–antigen cocrystal structure.

Results and Discussion

We sought to determine if MR78-like Abs are present in healthy individuals' B cell repertoires and if these Abs have GP binding and virus neutralization functions similar to those of MR78. A sequence-based search of ∼394 million naturally occurring human antibody sequences present in the Vanderbilt Vaccine Center next-generation sequence repository (November 2016) for antibody heavy chains encoded by *IGHV4-39* and *IGHJ6* genes and showing more than 70% amino acid similarity to MR78 HCDR3 loop failed to identify any similar antibodies. Therefore, we used a computational approach to compare the structure of the MR78 HCDR3 with the predicted structures of naturally occurring HCDR3s from large Ab gene repertoires obtained by next-generation sequencing (NGS) (16). Accurate modeling for each of the Ab genes individually would not be possible at this scale. Therefore, we used the P3SM approach (11) to interrogate the collection of human Ab variable region sequences.

Modeling. We focused our search for Abs with HCDR3s similar to that of the HCDR3 of MARV-neutralizing Ab MR78. This strategy was chosen for several reasons. First, the previously determined MR78–MARV GP cocrystal structure (15) reveals that mAb MR78 interacts with the GP antigen mainly via its elongated 17-residue-long HCDR3 loop: based on Rosetta calculations, the HCDR3 loop is responsible for 84.4% of the heavy-chain interaction energy and contributes to 72.5% of the heavy-chain interfacial surface area. We also tested this observation experimentally by comparing the binding of MR78 and its reverted unmutated ancestor MR78-RUA (germline V and J genes with the MR78 HCDR3 loop) to MARV GP. Reverting framework, HCDR1, and HCDR2 loops to the germline sequence did not cause a decrease in binding affinity (*SI Appendix*, Fig. S1). That finding supports our hypothesis that the MR78 HCDR3 loop plays a dominant role in the Ab–antigen interaction. Second, focusing on just the HCDR3 region increases the coverage of the available HCDR3 sequence repertoire sampled by including HCDR3 loops fused to different V or J genes. We searched within a database containing NGS sequencing data for Ab variable region sequences from peripheral blood cells of MARV-naïve individuals or a MARV disease survivor and found around 380,000 unique 17-residue-long HCDR3 loops. Since structural modeling for even this highly reduced dataset would be excessively expensive in terms of computational time, we reduced the size of the dataset using the previously described P3SM approach (11). In brief, a small training dataset of 1,200 randomly chosen loops was threaded on the MR78 Ab framework and scored in the context of MARV GP using Rosetta. We used these results to build a P3SM, a prediction of compatibility of the given amino acid in the given position with the desired structure and function (Fig. 1). The HCDR3 loop scores calculated using the P3SM give a good approximation of the Rosetta HCDR3 loop scores (*SI Appendix*, Fig. S2).

We then scored all previously found 17-residue-long loop sequences (∼361,000 from virus-naïve individuals and ∼18,000 from the MARV infection survivor) using the P3SM (*SI Appendix*, Fig. S3) and prioritized sequences for further analysis, choosing the 100 best-scoring HCDR3 loops from virus-naïve individuals and the 50 best-scoring HCDR3 loops from the MARV-immune individual. The P3SM prediction is less accurate than an energetic evaluation within Rosetta. That result is due to several factors, including the unlimited number of amino acid substitutions allowed and the possibility to predict whether the given amino acid is capable of adopting the given conformation and maintaining favorable contacts in this conformation, but not whether the given conformation is the most favorable one in the given amino acid environment. That is why we

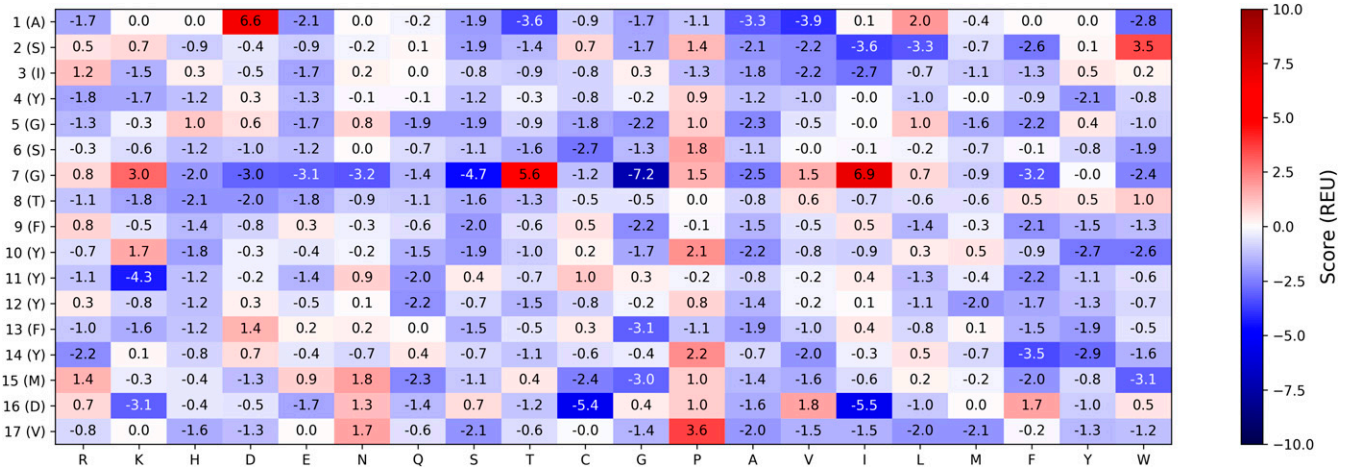

Fig. 1. The P3SM for 17-residue-long HCDR3 loop of MARV-neutralizing Ab MR78. Each row represents a position in the HCDR3 loop. The corresponding number and an amino acid identity in the original MR78 HCDR3 loop are shown on the left. Each column shows one of the 20 canonical amino acids. The number in each cell represents a score for placing the amino acid in the particular position in the HCDR3 loop. A positive score (colored red) means that the amino acid was predicted to be structurally or functionally unfavorable in that position; a negative score (colored blue) suggests potential compatibility with the given structure and function. If none of the 1,200 analyzed sequences contained a particular amino acid in a particular position of the loop, a zero value was assigned to that cell.

threaded the HCDR3 loops selected using P3SM on the MR78 Ab framework and scored in the context of MARV GP. These models were used to calculate numerous parameters that were shown earlier to predict successful discrimination of functional molecules: Ab–antigen interface score, HCDR3 Cα rmsd, shape complementarity, and Rosetta scores for the HCDR3 loop and the overall Ab structure (*SI Appendix*, Fig. S4). Surprisingly, our analysis revealed that some HCDR3 sequences from virus-naïve individuals were predicted to bind better than the best-scoring sequences found in the sequence pool derived from the MARV infection survivor.

While the absence of better scored Abs in virus-naïve individuals would be expected, the absence of predicted high-affinity clones in the Ab repertoire of the previously infected individual was somewhat unanticipated. Lower scores of the HCDR3 sequences from virus-naïve individuals might be explained by the size of the repertoires we studied. The healthy individuals' repertoires were determined from large leukopak samples that yielded tens of millions of unique heavy-chain clonotypes, while the relatively small overall number of available 17-residue-long HCDR3 loop sequences from the MARV-immune individual were obtained from a conventional peripheral blood draw with many fewer B cells. The low count of MR78-related 17-residue-long HCDR3 loop sequences could be due to multiple factors, including low frequencies of the MARV GP-specific B cells in the survivor's bloodstream at the moment of sample collection, fast convergence of the immune response on the MR78 Ab sequence, or, alternatively, generation of the antibody sequence through an insertion or deletion event.

To select candidate HCDR3 loops for further experimental validation, we chose the best-scoring sequences and clustered them using a phylogenetic cladogram. One sequence from each of the nine clusters obtained was selected for testing.

Functional Analysis of Chimeric Ab Variants. Based on in silico modeling, we chose eight HCDR3 loops from the repertoires of virus-naïve individuals and one from the previously MARV-infected individual for further experimental characterization (Fig. 2*A*). These sequences differed from the MR78 HCDR3 loop in 5 to 10 positions (*SI Appendix*, Fig. S5). The recombinant chimeric Abs we tested recapitulated the in silico modeling approach: the MR78 Ab HCDR3 loop was replaced by the selected HCDR3 sequences, and the resulting chimeric heavy chains were paired with the natural MR78 light chain and expressed recombinantly. All nine chimeric Abs were produced successfully as recombinant IgG1 proteins. We then assessed the ability of the expressed IgGs to bind to recombinant MARV GP using unpurified supernatants from Ab-expressing cells by biolayer interferometry (BLI). Three of the Abs (designated chimera-1, -2, and -3) showed binding affinity above background, with one of them, chimera-1, exhibiting binding comparable to that of the original MR78 Ab (*SI Appendix*, Fig. S6). All these three sequences derived from the repertoires of virus-naïve individuals.

We further evaluated these three MR78-like Abs using purified IgGs. We first assessed binding of the purified IgGs to MARV GP using BLI. All three chimeric Abs tested showed binding to the recombinant MARV GP ectodomain that was similar to that of the MR78 Ab in this assay (Fig. 2*B*). Next, we assessed the ability of these Abs to bind to full-length MARV GP displayed on the surface of a mammalian cell line. This assay allows for more native-like display of the GP antigen by displaying a membrane-bound trimeric form of full-length MARV GP similar to that on virions. Binding to cell surface-displayed filovirus GP molecules can predict neutralizing activity for human Abs (17). This analysis showed that chimera-1 Ab demonstrated the best binding among all chimeric Abs. The binding level for chimera-3 Ab was only slightly above the background, likely reflecting the fast off-rate of this Ab from intact GP (Fig. 2*C* and *SI Appendix*, Fig. S7). Neither of the Abs bound as well in this format as the original MR78 Ab. This finding agrees with the less favorable scores of the selected HCDR3 sequences compared to those for the MR78 HCDR3 loop sequence.

Functional Analysis of a Full-Length Ab from a Naïve Individual. The recombinant Abs tested so far were chimeric Abs in which HCDR3s from MARV-naïve donors were grafted onto the MR78 Ab framework, which was encoded by the human *IGHV4-39*01* and *IGHJ6*03* antibody gene segments. Next, we determined whether the HCDR3 loops selected from virus-naïve individuals exhibited MARV GP binding activity only in the context of the MR78 framework, or, alternatively, if the P3SM search can predict the function of a full-length naïve Ab. To test this, we focused on studies of the chimera-1 Ab, which bound best to MARV GP among the three chimeric Abs tested. We searched the MARV-naïve sequence database a second time, this time looking for all full-length Ab genes encoding the same 17-residue-long HCDR3 loop. The search yielded only one Ab sequence (designated naïve-1). The framework of this Ab originated from the same IGHJ gene (*IGHJ6*03*) as the original MR78 Ab but differed significantly overall, being encoded by the *IGHV3-43* gene segment (Fig. 3*A*). A limitation for analyzing the binding of the natural Ab from the original B cell encoding

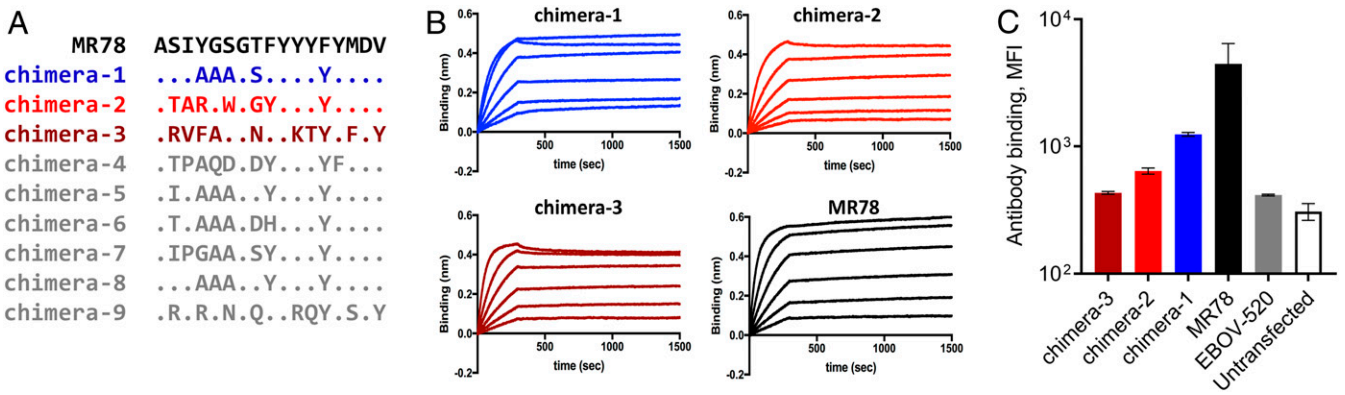

Fig. 2. Characterization of chimeric Abs. (*A*) Alignment of the amino acid sequences of the nine Abs selected for experimental testing HCDR3 loops to the MR78 HCDR3 loop sequence. Dots indicate the same amino acid as in the MR78 Ab. The sequences of the three Abs that were selected for further characterization after the initial screening are highlighted with color. (*B*) Binding of the MR78 and three selected chimeric MR78-like Abs to MARV GP by BLI. Lines represent twofold dilutions of the Abs from 166 nM to 5.2 nM. (*C*) Comparison of the binding of three selected chimeric Abs and MR78 Ab using cell-surface MARV GP display. A previously described mAb EBOV-520 that is specific to ebolavirus GP (17) and untransduced cells were used as negative controls. Shown are mean values ± SD of two replicates.

naïve-1 is that native light-chain pairing information is missing in the collection of NGS data we used. NGS analysis of paired heavy and light chains currently is not technically feasible at the scale of millions to billions. Therefore, we expressed the full-length heavy-chain Ab sequence for naïve-1 paired with the original MR78 light chain.

The cDNA encoding naïve-1 Ab heavy chain successfully produced recombinant IgG, despite the pairing with a potentially nonnative light chain. Display of the same HCDR3 loop on the native Ab sequence (naïve-1 Ab) rather than on the MR78 background (chimera-1 Ab) resulted in approximately a twofold decrease in the signal level for binding to cell surface-displayed MARV GP (Fig. 3*B*). This finding shows that most of the binding to the MARV GP was mediated by the original HCDR3 from the naïve-1 IgG and was not simply due to chimerization of an HCDR3 loop to the MR78 framework as in chimera-1 Ab. The observed decrease in binding of the wild-type heavy chain compared to the chimeric Ab is most likely caused by altered non-HCDR3 loop interactions. In addition to the HCDR3, residues from both the HCDR1 and HCDR2 loops of MR78 make contacts with MARV GP (*SI Appendix*, Fig. S9*B*). The HCDR1 loop of naïve-1 Ab is two amino acids shorter than that of MR78 and chimera-1 Abs. This length difference might make it impossible for the naïve-1 HCDR1 loop to contact the antigen. Alternatively, a different relative orientation in the Ab of heavy and light chains can change the way in which a single light chain participates in the antigen recognition.

We next assessed MARV-neutralizing activity of the chimera-1 Ab and the naïve-1 Ab from the virus-naïve individual using recombinant chimeric Ebola virus engineered to incorporate MARV GP that replaced its own GP (18). The chimera-1 Ab showed neutralization activity, with a half-maximal inhibitory concentration (IC_{50}) value ~15-fold lower than that of the original MR78 Ab. In agreement with activity from the MARV GP cell surface display binding analysis, the full-length naïve-1 Ab was less potent for virus neutralization. Only ~20% neutralization was observed at the highest concentration tested of 200 μg/mL (Fig. 3*C*).

Structural Analysis of the Chimera-1 and Naïve-1 Fabs

Overall Structure. To confirm that the functional similarity of the Abs was indeed mediated by a structural similarity predicted by the P3SM approach, we determined the structures of the unbound chimera-1 and naïve-1 Fabs at 3.48-Å and 1.39-Å resolution, respectively. As expected, the overall Cα rmsd between the HCDR3 chimera-1 and the unbound MR78 Fabs (PDB ID 5JRP) is low (0.84 Å). The relative orientation of the constant and variable regions in the naïve-1 Fab structure differs from that in the MR78 Fab. However, the variable fragments (Fvs) align very well (Cα rmsd 1.07 Å for 227 structurally aligned Fv residues) despite only 57% homology in the heavy-chain variable domain region (Fig. 3*A*) and a different length of the HCDR1 and HCDR2 loops (Fig. 4*A*).

We focused our attention on the HCDR3 loop, based on its major role in the interaction between the MR78 heavy chain and MARV GP antigen (accounting for 72.5% of the total buried surface area and 84.4% of the estimated interface energy). We found that the conformation of the HCDR3 loop selected by the P3SM approach in chimera-1 is overall closer to the conformation of the loop in the naïve-1 structure than to the MR78 HCDR3 loop conformation. The HCDR3 Cα rmsd between MR78 and chimera-1 is 5.67 Å, while the HCDR3 Cα rmsd between naïve-1 and chimera-1 is 3.7 Å. The torso regions show only small perturbations; the main contribution to the difference is the conformation of the tip of the HCDR3 loops (Fig. 4*B*).

HCDR3 Torso. As the original Ab MR78, the chimera-1 and naïve-1 Fabs maintain a kinked torso conformation with τ_{101} and α_{101} (19) equal to 103.7° and 28.4° or 107.9° and 51.1°, respectively. The obtained refined model of the chimera-1 Fab agrees with an assumption that Asp101–Tyr32 interaction contributes to the formation of the kink as it was observed in the MR78 Ab, although the lack of the atomic details prevents us from stating this conclusion with high confidence. However, neither a classical basic Arg/Lys residue at the second position of the HCDR3 loop (Ser94 in naïve-1) nor an Asp101–HCDR1 Tyr (Tyr32 in naïve-1) interaction found in the original MR78 Ab were observed in the naïve-1 Fab structure. The density map suggests the presence of at least two stable conformations of the Asp101 side chain. One of them can form a hydrogen bond with the Ser94 (the second position of the HCDR3 loop) similar to the MR78 Asp101. Light-chain Lys55 might serve as an interaction partner for the other. For the deposited structure, we chose a conformation that fails to allow Lys55 to interact with Asp101, but it should be noted that the last two atoms of the Lys55 side chain are characterized by higher-than-average B factors (39.05 and 43.48 for CE and NZ, respectively, compared with the average B

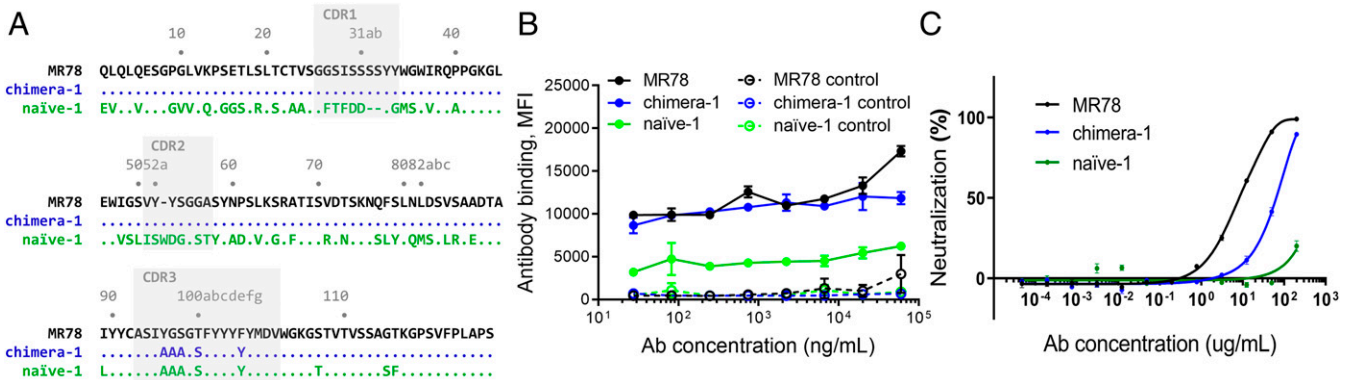

Fig. 3. Characterization of the full-length MR78-like Ab. (*A*) Alignment of the amino acid sequences of the heavy-chain variable regions of MR78, chimera-1, and naïve-1 Abs. Dots indicate the same amino acid as in the MR78 Ab sequence. Positions of the HCDR loops (IMGT definition) are shaded. Chothia numbering of residues is used and indicated above the alignment. (*B*) Comparison of the binding efficiency of different concentrations of MR78, chimera-1 Ab, and the full-length Ab naïve-1 using cell surface-displayed MARV GP antigen. Solid symbols show Abs binding to the MARV GP-expressing Jurkat cells. Hollow symbols show Abs binding to untransduced cells (negative control). Data shown are the mean values ± SD of technical duplicates. (*C*) Neutralization of MARV by MR78 Ab, chimeric MR78-like Ab chimera-1, or the corresponding full-length Ab naïve-1. Predicted IC_{50}: MR78, 8.3 μg/mL (95% CI, 7.4 to 9.3); chimera-1, 111.1 μg/mL (95% CI, 87.9 to 140.5). Data shown are the mean values ± SD of technical triplicates.

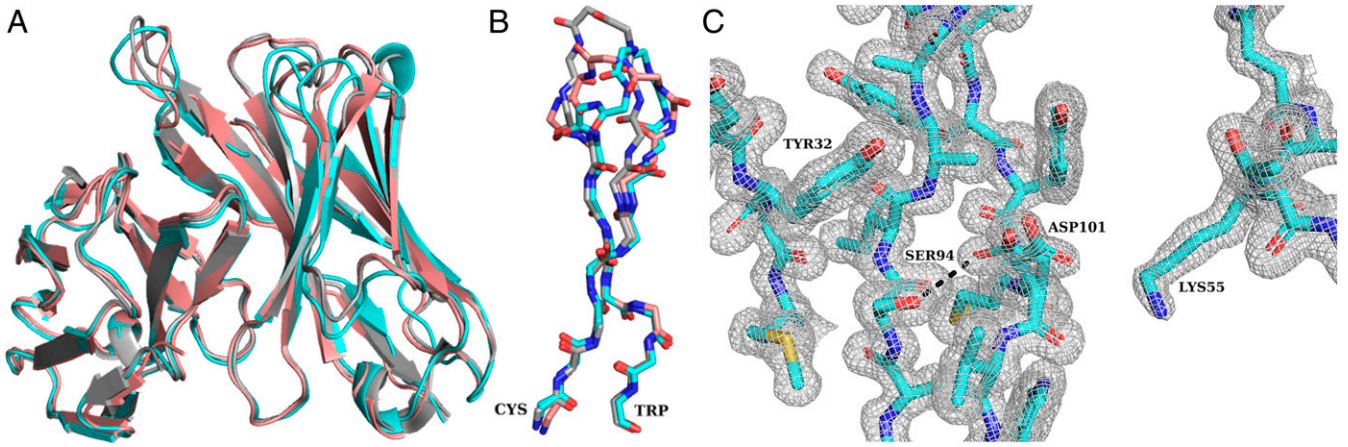

Fig. 4. Crystal structure analysis. Comparison of MR78 (PDB ID 5JRP; gray), chimera-1 (salmon), and naïve-1 (cyan) Fv domains (*A*) and HCDR3 loop backbone conformations (*B*). (*C*) Side-chain conformations and possible interactions of the naïve-1 Ab residue Asp101. The 2Fo-Fc electron density map contoured at 1σ is shown as gray mesh. Black dotted line represents possible hydrogen bond between Asp101 and Ser94 side chains.

factor of 16.79 for the model). This uncertainty for the placement of these atoms, combined with the close proximity of the light-chain Lys55 to the Asp101, makes this residue a possible interaction partner. On the other hand, there is no evidence for the presence of the second conformation or high uncertainty of the Tyr32 side-chain placement in naïve-1 Fab structure (Fig. 4*C*). This finding suggests that, despite the presence of both Tyr32 and Asp101 that form a hydrogen bond in MR78 and stabilize its kinked torso conformation (20), the naïve-1 Ab establishes an alternate paradigm.

HCDR3 Loop Tip. In all three cases, the tip of the HCDR3 loop of the unbound Fab is involved in the crystal packing interface (*SI Appendix*, Fig. S8). While it was shown previously that the apex of the MR78 HCDR3 loop undergoes a backbone conformational change upon binding to the antigen, the difference between the MR78 unbound and bound forms is less dramatic than that between the MR78 and the experimentally determined structures of chimera-1 and naïve-1. However, it is impossible to tell whether the HCDR3 loop tip conformations observed in the crystal structures are the actual dominant conformations present in solvent, which could explain lower binding affinities of these Abs to the antigen. A second explanation would be that the HCDR3 loop in these Abs is just less rigid as a result of the presence of the longer sequence of small and flexible amino acids (ASI**AAAGS**FYYYYYMDV), which was previously hypothesized to mediate polyspecificity of Abs (21). This feature, in turn, could make HCDR3 loops of the chimera-1 and naïve-1 Abs more prone to the environment- or crystal packing-induced conformational changes. Higher flexibility increases the entropic cost of the antigen binding and thus can reduce the binding affinity.

Possible Interaction of the Abs with MARV GP. To estimate possible interaction modes of chimera-1 and naïve-1 Abs with MARV GP, we superimposed the obtained Fab structures with MR78 Ab in the MR78–MARV GP cocrystal structure and performed interface energy minimization using Rosetta (*SI Appendix*, Fig. S9). While the relative orientation of heavy and light chains remained unchanged in the new Abs, in both cases, light chains seem to be capable of maintaining the same contacts with MARV GP as seen in the MR78–MARV GP structure. In addition to the HCDR3 residues, MR78 Ab residues from HCDR1 and HCDR2 loops contribute to the binding to MARV GP. Having the same framework, chimera-1 Ab is theoretically capable of maintaining these interactions. However, even a small perturbation that can be caused by slightly different HCDR3 loop placement in the RBD might make these additional interactions less favorable or even impossible. The shorter HCDR1 loop of naïve-1 Ab and its different amino acid content almost certainly prevents it from making contacts with MARV GP if the Ab binds to the antigen in a similar fashion as MR78. Despite also belonging to a different structural class, the HCDR2 loop of naïve-1 remains within a reasonable distance from the GP, suggesting its involvement in the interaction.

Conclusions

Here we demonstrated that the P3SM approach is capable of identifying not only new HCDR3 sequences that are functional when chimerized on the "parental" Ab framework, but also an entirely different Ab heavy chain-variable gene sequence encoding an Ab that can bind the antigen. Our computational survey of a large database containing sequences of Ab variable regions revealed three HCDR3 loop sequences from virus-naïve individuals compatible with binding to the MARV GP. Further evaluation of the full heavy-chain sequence function confirmed that at least two IGHV gene segments (*IGHV4-39*01* and *IGHV3-43*) are capable of providing the required framework for proper positioning and structuring of the HCDR3 loop for MARV GP binding and virus neutralization. This finding proposes that potential Ab precursors exist in virus-naïve individuals that can recognize the MR78 MARV GP epitope and might evolve into highly potent protective Abs. Thus, the MR78 Ab-binding epitope could be an attractive candidate for epitope-focused vaccine design.

Structural characterization of two unbound Abs further confirmed the predictive power of the structure-based similarity search. Both the chimeric Ab and the Ab containing the full-length native heavy chain-variable region displayed high overall conformational similarity. Only the tip of the HCDR3 loop differed from that of MR78 Ab. Interestingly, while the torso conformation in the naïve-1 Ab is very close to the one in the MR78 Ab, it is not stabilized by the same amino acid pair (Asp101–Tyr32) interaction that was suggested to be important for the MR78 Ab structure (20). That observation again emphasizes the lack of the full understanding of principles of the long loops' structure formation. Increasing the number of known structurally similar Abs with diverse sequences should improve our understanding in the future and hence improve the quality of de novo Ab structure prediction. Superimposition of the obtained structures with MR78 Ab in the MR78–MARV GP cocrystal structure and analysis of the theoretical binding

interfaces suggests that chimera-1 should be capable of maintaining most of the MR78 interactions. In contrast, the naïve-1 Ab most likely lacks some of the heavy-chain interactions found in MR78 because of the shorter HCDR1 loop. This finding might explain the weaker binding to the antigen that was observed.

The P3SM approach was shown to be successful in template-based search of functional Abs with low sequence similarity to the starting Ab. This result corroborates the earlier discovery of new PG9-like Abs capable of binding to HIV (11) and CH65-like Abs against influenza hemagglutinin (12). In all these test cases, the Ab–antigen interaction is mainly driven by the HCDR3 loop. However, interactions of many highly potent Abs demonstrate much more complex interaction patterns that often involve all six complementarity determining region (CDR) loops as well as framework amino acids (22, 23). While, technically, the extension of the P3SM approach to cover multiple loop regions or even nonloop regions is possible and straightforward, such an extension might face some obstacles. Some CDR loop lengths are more common than others. Inclusion of multiple CDR loop regions, especially those with relatively rare loop lengths, can significantly decrease the available sequence search space and hence lower the probability of finding desired Abs. This limitation might be overcome, for example, through the employment of the recent finding of the length-independent structural CDR canonical classes (24). Building chimeric models with structurally similar loops of different lengths, and their use as additional templates for the P3SM search, will increase the number of scored sequences. However, strengths and limitations of such an approach have yet to be explored.

Methods

In Silico Analysis. Before the start of the computational analysis, the structure of MARV GP in complex with MR78 (PDB ID 5UQY) was refined using the Phenix-Rosetta protocol (20, 25, 26).

Estimation of the relative contribution of different parts of Abs into Ab–MARV GP interaction energy and interaction surface area was performed using Rosetta InteractionEnergyMetric and SasaMetric SimpleMetrics, respectively.

To generate a P3SM for MR78, 1,200 sequences of 17-amino-acid-long HCDR3 loops were selected randomly from a database of NGS data containing Ab sequences obtained from sequencing the antibody genes expressed in peripheral blood mononuclear cells of healthy individuals and the bulk peripheral blood NGS data from the Marburg virus infection survivor. Each of these 1,200 sequences was threaded using Rosetta onto the MR78 structure in the context of MARV GP and then relaxed in order to generate 10 models for each HCDR3 sequence. Based on the Rosetta total score, the five top-scoring models were selected for further analysis (a total of 6,000 models). The HCDR3 loop total score and sequence data were used to calculate a score for each of the 20 amino acids at every position in the HCDR3 using linear regression. If none of the 1,200 analyzed sequences contained a particular amino acid in a particular position of the loop, a zero value was assigned to that position. The scores obtained are shown in Fig. 1.

All unique 17-residue-long HCDR3 loop sequences then were scored by the P3SM. Scoring by P3SM amounts to summing the score values for all 17 amino acids in the loop. The top-scoring 150 sequences were threaded onto the MR78 structure in the context of MARV GP. The resulting models were analyzed based on total score of the complex, predicted binding affinity, affinity per solvent-accessible surface area, HCDR3 loop score, HCDR3 rmsd to MR78, overall rmsd of the complex, and shape complementarity. The top 100 HCDR3 loop sequences were clustered based on the heavy-chain variable (V_H) and joining (J_H) gene segment assignments.

Antibody Genes Cloning and Expression. To create chimeric MR78-like Abs, we exchanged the nucleotides in the HCDR3 loop-encoding region of a cDNA encoding MR78 for nucleotides encoding the HCDR3 amino acids from our models. We modified nucleotides with a QuikChange protocol using long primers containing nonoverlapping sequences at the 3′ end and primer–primer complementary regions at the 5′ end. In some cases, two consequent cloning steps were performed. A cDNA specifying the sequence of a mAb that we designated naïve-1 was synthesized by GenScript and cloned into an IgG expression vector (27) using the Gibson Assembly Master Mix reagent (New England BioLabs). All constructs obtained were verified by Sanger sequence analysis.

Recombinant Ab proteins were expressed by transient transfection using Expi293F human embryonic kidney cells in serum-free medium (ThermoFisher Scientific). The supernatants were harvested after 7 d and filter-sterilized with a 0.2-μm filter, and IgG concentration in supernatant was measured on the FortéBio Octet instrument using anti-IgG anti-human IgG quantitation (AHQ) sensors. Ab variants were analyzed directly from supernatants for binding to recombinant MARV GP using Octet anti-IgG biosensors. Variants that showed binding in the screening step were produced at a larger scale and purified using a 5-mL HiTrap MabSelectSure protein A column (GE Healthcare).

Fab fragments were expressed recombinantly by cloning the cDNA encoding the heavy-chain variable regions into Fab expression vectors. Fabs were produced by transient cotransfection of the Fab heavy-chain and light-chain cDNA into Expi293F cells (ThermoFisher Scientific). Recombinant Fabs were purified from centrifuged, 0.45-μM-filtered supernatant using an anti-CH1 column (GE Healthcare).

Recombinant MARV GP ΔTM Expression. The ectodomain of MARV GP aa 1 to 648 (strain Angola2005; GenBank accession DQ447653) was expressed transiently in Expi293F cells with a C-terminal Strep-tag II and purified using a 5-mL StrepTrap HP column (GE Healthcare). The GP was purified further, and buffer exchanged in phosphate-buffered saline (PBS) using a Superose 6 Increase (GE Healthcare) size-exclusion chromatography column.

BLI Assay. Binding kinetics were determined using BLI with an Octet RED96 instrument (FortéBio). Abs were loaded onto AHQ biosensors at a concentration of 10 μg/mL in kinetics buffer (PBS + 1% BSA, 0.05% Tween 20). The binding experiments were performed with the following steps: 1) baseline in kinetics buffer for 60 s, 2) loading of IgG for 120 s, 3) baseline for 60 s, 4) association of recombinant MARV GP for 300 s, and 5) dissociation of MARV GP into kinetics buffer for 1,200 s. A reference well was run in all experiments, in which IgG was loaded onto the biosensor but antigen was not present, and the result was subtracted from all sample wells to correct for drift and buffer evaporation. MARV GP was diluted twofold starting from a concentration of 166 nM. Six dilutions were used to fit kinetic curves. Curves were fit to a 1:1 model using the FortéBio software. Curve fits were accepted only if they fulfilled an R^2 of > 0.9.

ELISA Binding Assay. Wells of microtiter plates were coated with purified MARV GP ΔTM at 2 μg/mL and incubated at 4 °C overnight. Plates were blocked with 2% nonfat dry milk and 2% normal goat serum in Dulbecco's PBS containing 0.05% Tween-20 for 1 h. Serial threefold dilutions of purified mAbs were applied to the wells in quadruplicate and incubated for 1 h at ambient temperature. The bound antibodies were detected using goat anti-human IgG conjugated with horseradish peroxidase (Southern Biotech) and 3,3′,5,5′-tetramethylbenzidine substrate (ThermoFisher Scientific). Color development was monitored, 1N hydrochloric acid was added to stop the reaction, and the absorbance was measured at 450 nm using a spectrophotometer (Biotek).

Cell Surface MARV GP Antigen Display. Jurkat cells stably expressing MARV (Angola) GP were briefly proteolytically primed with thermolysin (5 min at 34 °C, 0.5 ng/mL thermolysin in Ca^{2+} Mg^{2+} PBS) to facilitate receptor binding site (RBS) exposure; this binding method was validated in previous studies with RBS-specific MARV mAbs (28). This approach resulted in surface exposure of partially cleaved EBOV-MARV GP_{CL}. Approximately 10^5 cells per staining reaction were incubated in 50 μL of FACS buffer (PBS containing 2% of heat-inactivated fetal bovine serum) with serial threefold dilutions of individual mAbs. Incubation with Abs was performed at ambient temperature for 40 min in V-bottom 96-well plates. Untransduced Jurkat cells that were incubated with the same mAb dilutions served as a control for specificity of mAb binding. Cells were washed twice with PBS and fixed with 4% paraformaldehyde in PBS for 10 min at ambient temperature. Cells were washed twice with FACS followed by incubation with secondary detection anti-human PE-conjugated Ab (Southern Biotech) at 1:500 dilution in a 30-μL volume of diluted Ab per well for 40 min at ambient temperature. After washes, cells were analyzed on a high-throughput flow cytometer (iQue Screener; IntelliCyt).

Virus Neutralization Assay. The neutralization assay was performed in the biosafety level 4 facility of the Galveston National Laboratory, University of Texas Medical Branch. The recombinant chimeric EBOV strain Mayinga-expressing enhanced green fluorescent protein from an added gene in which the EBOV GP was replaced with its counterpart from MARV (strain

200702854 Uganda) (29) was used (EBOV/MARV-GP) (18). Fourfold serial dilutions of MR78, chimera-1, or naïve-1 Abs were tested at final concentrations from 200 to 4.8×10^{-5} μg/mL in triplicate Vero-E6 cell monolayers. The assay was performed in a high-throughput format as previously described (18).

Crystallization, Data Collection, and Structure Determination. Chimera-1 Fab (9.5 to 13 mg/mL in 20 mM Tris, 7.5, 50 mM NaCl) was crystalized at 14 °C in 4.3 to 5 M NaCl, 0.1 M Hepes, pH 7.5, with a protein-to-precipitant volume ratio of 1:2 using the hanging-drop vapor diffusion technique. Crystals grew within 1 to 2 wk and were flash-frozen in liquid nitrogen using Parabar 10312 oil as cryoprotectant.

The naïve-1 Fab (13.5 mg/mL in 20 mM Tris, 7.5, 50 mM NaCl) was crystallized at 18 °C in 1 to 8% wt/vol PEG 8000, 20% vol/vol 2-propanol, 0.1 M Tris, pH 7.5 to 7.9, with a 1:1 protein-to-precipitant volume ratio using the sitting-drop vapor diffusion technique. Crystals grew within 3 to 4 d and were flash-frozen in liquid nitrogen using cryo-solution made of mother liquor supplemented with 10% glycerol.

Diffraction data were collected at the beamline 21-ID-G at the Advanced Photon Source. The diffraction data were processed using the xia2 software suite (30). The crystal structures were solved by molecular replacement with MOLREP (31) using the MR78 structure (PDB ID 5JRP) for chimera-1 and Hyb3 Fab from its complex with HLA-A1:MAGE-A1 (PDB ID 1W72) for naïve-1 as search models. Model building and iterative refinement were performed with Coot (32), REFMAC (33), and Phenix (34), respectively. The Rosetta-Phenix low-resolution refinement protocol (25) was applied to assist in the refinement of the chimera-1 structure.

The final statistics of the structures are shown in *SI Appendix*, Table S1. The models have been deposited into the Protein Data Bank (PDB ID 6V4Q and 6V4R). Structure figures were prepared using PyMOL (v.2.2.3; Schrodinger).

Quantification and Statistical Analysis. The descriptive statistics mean ± SEM or mean ± SD were determined for continuous variables as noted. Statistical analyses were performed using Prism v8.2.1 (GraphPad).

Data and Software Availability. The accession numbers for the crystal structures reported in this paper have been deposited to the Protein Data Bank under accession codes 6V4Q (35) and 6V4R (36). The Rosetta-refined MR78–MARV GP complex structure is available in the PDB-Dev (PDBDEV_00000048). Sequence data, the subset of which were used as MARV-naïve individuals' sequences in this study, are available at the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under code SRP174305 or Bioproject number PRJNA511481. The PyIR processing pipeline and all FASTA files from the Adaptive Biotechnologies sequencing can be found in Github at https://github.com/crowelab/PyIR. All of the processed data are available on VDJSERVER (https://vdjserver.org/). Requests for computational resources should be directed to and will be fulfilled by the lead contact, J.M. (jens@meilerlab.org). Requests for MARV disease survivor sequences should be directed to and will be fulfilled by J.E.C. (james.crowe@vumc.org). All other relevant data are included with the manuscript. Materials described in this paper are available for distribution under the Uniform Biological Material Transfer Agreement, a master agreement that was developed by the NIH to simplify transfers of biological research materials.

ACKNOWLEDGMENTS. This work was supported by US NIH Grants R01 AI141661 (to A.B., J.E.C., and J.M.), U19 AI117905 (to J.E.C. and J.M.), and U19 AI109711 (to J.E.C. and A.B.); National Center for Research Resources Grant UL1 RR024975-01, which is now at the National Center for Advancing Translational Sciences, and Grant 2 UL1 TR000445-06; and Defense Threat Reduction Agency Grant HDTRA1-13-1-0034 (to J.E.C. and A.B.). Work in BSL-4/ABSL-4 containment of the Galveston National Laboratory was supported by NIH Grant 5UC7AI094660-07. The Jurkat-MARV GP cell line was a gift from Carl Davis and Rafi Ahmed. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

1. B. E. Correia *et al*., Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201–206 (2014).
2. J. Jardine *et al*., Rational HIV immunogen design to target specific germline B cell receptors. *Science* **340**, 711–716 (2013).
3. A. T. McGuire *et al*., Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies. *J. Exp. Med.* **210**, 655–663 (2013).
4. J. Dunbar *et al*., SAbPred: A structure-based antibody prediction server. *Nucleic Acids Res.* **44**, W474–W478 (2016).
5. B. D. Weitzner *et al*., Modeling and docking of antibody structures with Rosetta. *Nat. Protoc.* **12**, 401–416 (2017).
6. R. Chen, L. Li, Z. Weng, ZDOCK: An initial-stage protein-docking algorithm. *Proteins* **52**, 80–87 (2003).
7. D. Kozakov, R. Brenke, S. R. Comeau, S. Vajda, PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins* **65**, 392–406 (2006).
8. A. Sircar, J. J. Gray, SnugDock: Paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *PLoS Comput. Biol.* **6**, e1000644 (2010).
9. K. Krawczyk, X. Liu, T. Baker, J. Shi, C. M. Deane, Improving B-cell epitope prediction and its application to global antibody-antigen docking. *Bioinformatics* **30**, 2288–2294 (2014).
10. E. Liberis, P. Velickovic, P. Sormanni, M. Vendruscolo, P. Liò, Parapred: Antibody paratope prediction using convolutional and recurrent neural networks. *Bioinformatics* **34**, 2944–2950 (2018).
11. J. R. Willis *et al*., Long antibody HCDR3s from HIV-naïve donors presented on a PG9 neutralizing antibody background mediate HIV neutralization. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4446–4451 (2016).
12. J. A. Finn *et al*., Identification of structurally related antibodies in antibody sequence databases using rosetta-derived position-specific scoring. *Structure* **28**, 1124–1130.e5 (2020).
13. A. Leaver-Fay *et al*., ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
14. A. I. Flyak *et al*., Mechanism of human antibody-mediated neutralization of Marburg virus. *Cell* **160**, 893–903 (2015).
15. T. Hashiguchi *et al*., Structural basis for Marburg virus neutralization by a cross-reactive human antibody. *Cell* **160**, 904–912 (2015).
16. C. Soto *et al*., High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398–402 (2019).
17. P. Gilchuk *et al*., Multifunctional pan-ebolavirus antibody recognizes a site of broad vulnerability on the Ebolavirus glycoprotein. *Immunity* **49**, 363–374.e10 (2018).
18. P. A. Ilinykh *et al*., Chimeric Filoviruses for identification and characterization of monoclonal antibodies. *J. Virol.* **90**, 3890–3901 (2016).
19. B. D. Weitzner, R. L. Dunbrack, Jr, J. J. Gray, The origin of CDR H3 structural diversity. *Structure* **23**, 302–311 (2015).
20. A. K. Sangha *et al*., Role of non-local interactions between CDR loops in binding affinity of MR78 antibody to Marburg virus glycoprotein. *Structure* **25**, 1820–1828.e2 (2017).
21. J. R. Willis, B. S. Briney, S. L. DeLuca, J. E. Crowe, Jr, J. Meiler, Human germline antibody gene segments encode polyspecific antibodies. *PLoS Comput. Biol.* **9**, e1003045 (2013).
22. M. Ono *et al*., Structural basis for tumor necrosis factor blockade with the therapeutic antibody golimumab. *Protein Sci.* **27**, 1038–1046 (2018).
23. J. Y. Lee *et al*., Structural basis of checkpoint blockade by monoclonal antibodies in cancer immunotherapy. *Nat. Commun.* **7**, 13354 (2016).
24. J. Nowak *et al*., Length-independent structural similarities enrich the antibody CDR canonical class model. *MAbs* **8**, 751–760 (2016).
25. F. DiMaio *et al*., Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* **10**, 1102–1104 (2013).
26. A. K. Sangha *et al*, Data from "Refined structure of MR78 antibody in complex with Marburg glycoprotein using Rosetta." PDB-Dev. https://pdb-dev.wwpdb.org/entry.html?PDBDEV_000000485. Accessed 13 November 2020.
27. G. R. McLean, A. Nakouzi, A. Casadevall, N. S. Green, Human and murine immunoglobulin expression vector cassettes. *Mol. Immunol.* **37**, 837–845 (2000).
28. P. A. Ilinykh *et al*., Non-neutralizing antibodies from a Marburg infection survivor mediate protection by Fc-effector functions and by enhancing efficacy of other antibodies. *Cell Host Microbe* **27**, 976–991.e11 (2020).
29. J. S. Towner *et al*., Generation of eGFP expressing recombinant Zaire ebolavirus for analysis of early pathogenesis events and high-throughput antiviral drug screening. *Virology* **332**, 20–27 (2005).
30. G. Winter, xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186–190 (2010).
31. A. Vagin, A. Teplyakov, MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
32. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
33. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
34. P. D. Adams *et al*., PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
35. N. G. Bozhanova, J. E. Crowe, J. Meiler, Crystal structure of a MR78-like antibody naive-1 Fab. *Protein Data Bank*. https://www.rcsb.org/structure/6V4Q. Deposited 28 November 2019.
36. N. G. Bozhanova, J. E. Crowe, J. Meiler, Crystal structure of a chimeric MR78-like antibody chimera-1 Fab. *Protein Data Bank*. https://www.rcsb.org/structure/6V4R. Deposited 28 November 2019.