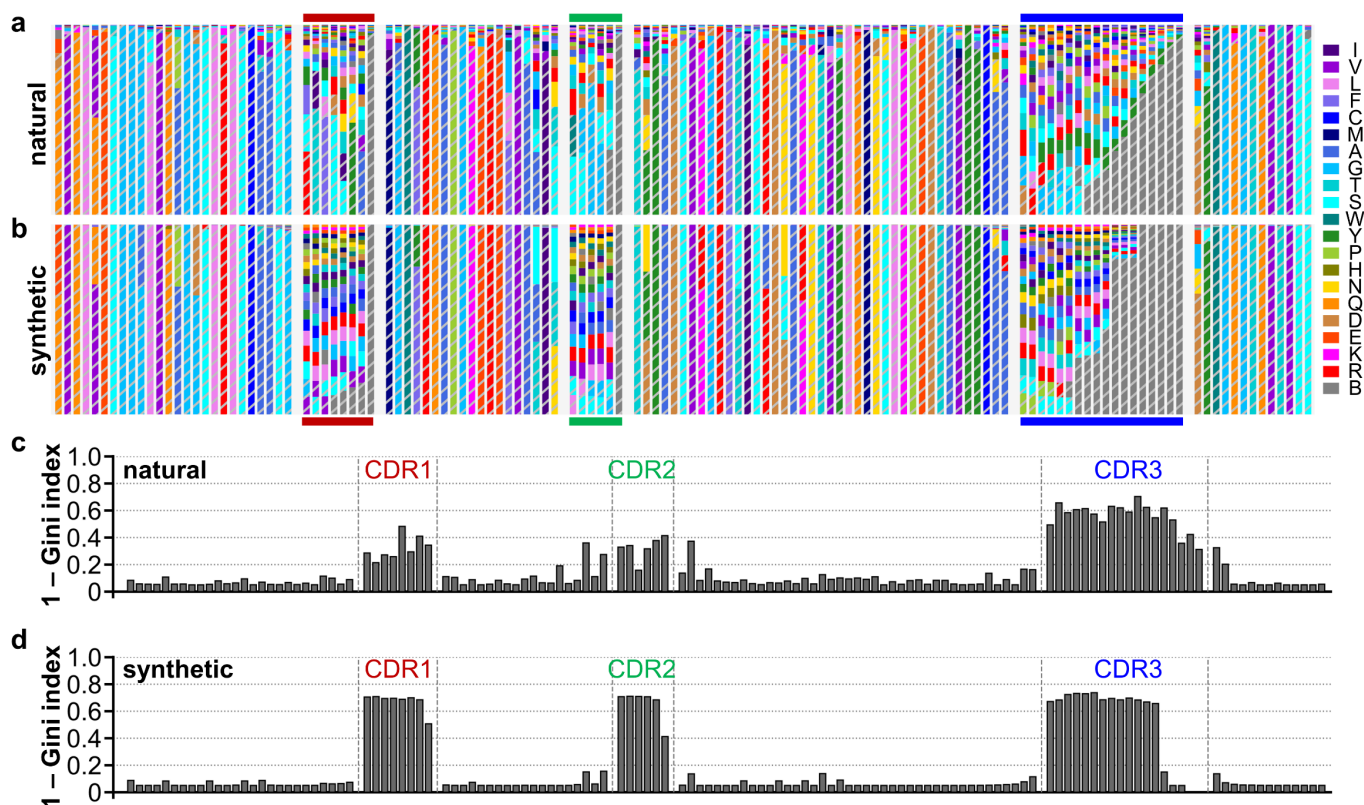


Fig S1



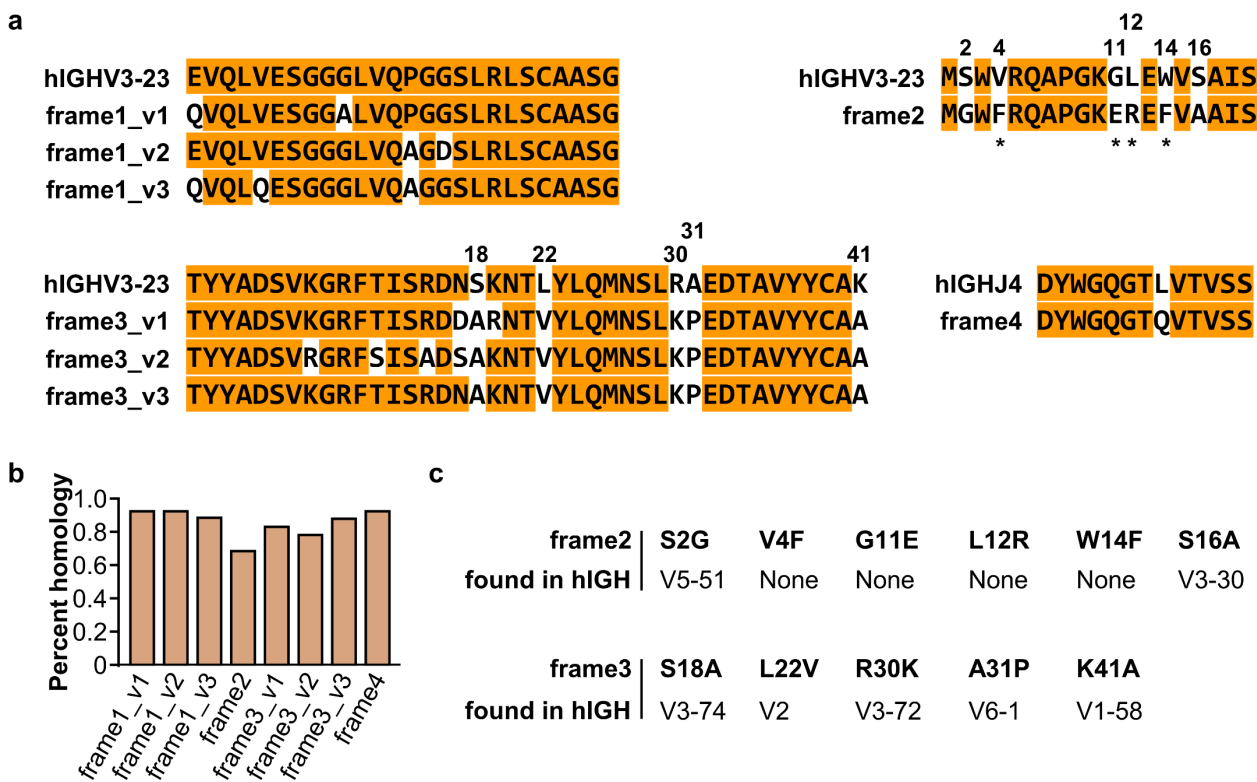
652 **Fig. S1. Amino acid profiles of natural and synthetic VHHs.** (a) Position-wise amino acid
653 profile of natural VHHs (298 VHHs, PDB) and (b) synthetic VHHs. Amino acids were color
654 coded according to labels to the right, B indicates an empty position. Bar height is the relative
655 percentage of each amino acids. The two most common amino acids were shown as patterned
656 bars while others were shown as solid bars. (c) Plot of diversity index (as 1 - Gini index) for
657 each amino acid position of natural VHHs and (d) synthetic VHHs.

658

659

660

Fig S2



661 **Fig. S2. Design of VHH frames and their homology to human IGH genes. (a)** Amino acid
662 sequences encoded by frames that serve as templates for VHH library generation were aligned to
663 the corresponding segments of the human IGHV3-23 (hIGHV3-23) or IGHJ4 (hIGHJ4). Positions
664 in hIGHV3-23/hIGHJ4 that are identical to the corresponding position in at least one VHH frames
665 are highlighted in orange. Positions in VHH frames that are identical to the corresponding position
666 in hIGHV3-23/hIGHJ4 are highlighted in orange. hIGHV3-23 positions not identical to any VHH
667 frames are numbered according to its position within the segment. Asterisks indicate VHH
668 hallmark residues thought to be required for VHH's independence of light chain. **(b)** Percent
669 homology of VHH frames to the closest human gene. **(c)** List of VHH residues at positions
670 numbered in (a) and representative human IGHV genes that encode the same VHH residue at the
671 corresponding position. None: no human IGHV genes has the VHH residue at the corresponding
672 position.

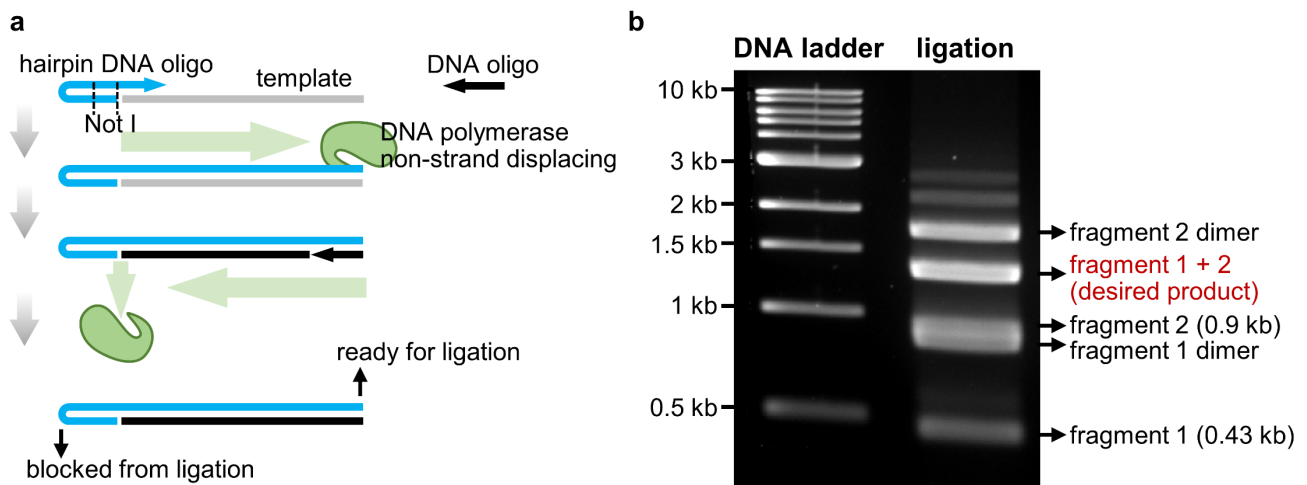
673

674

675

676

Fig S3



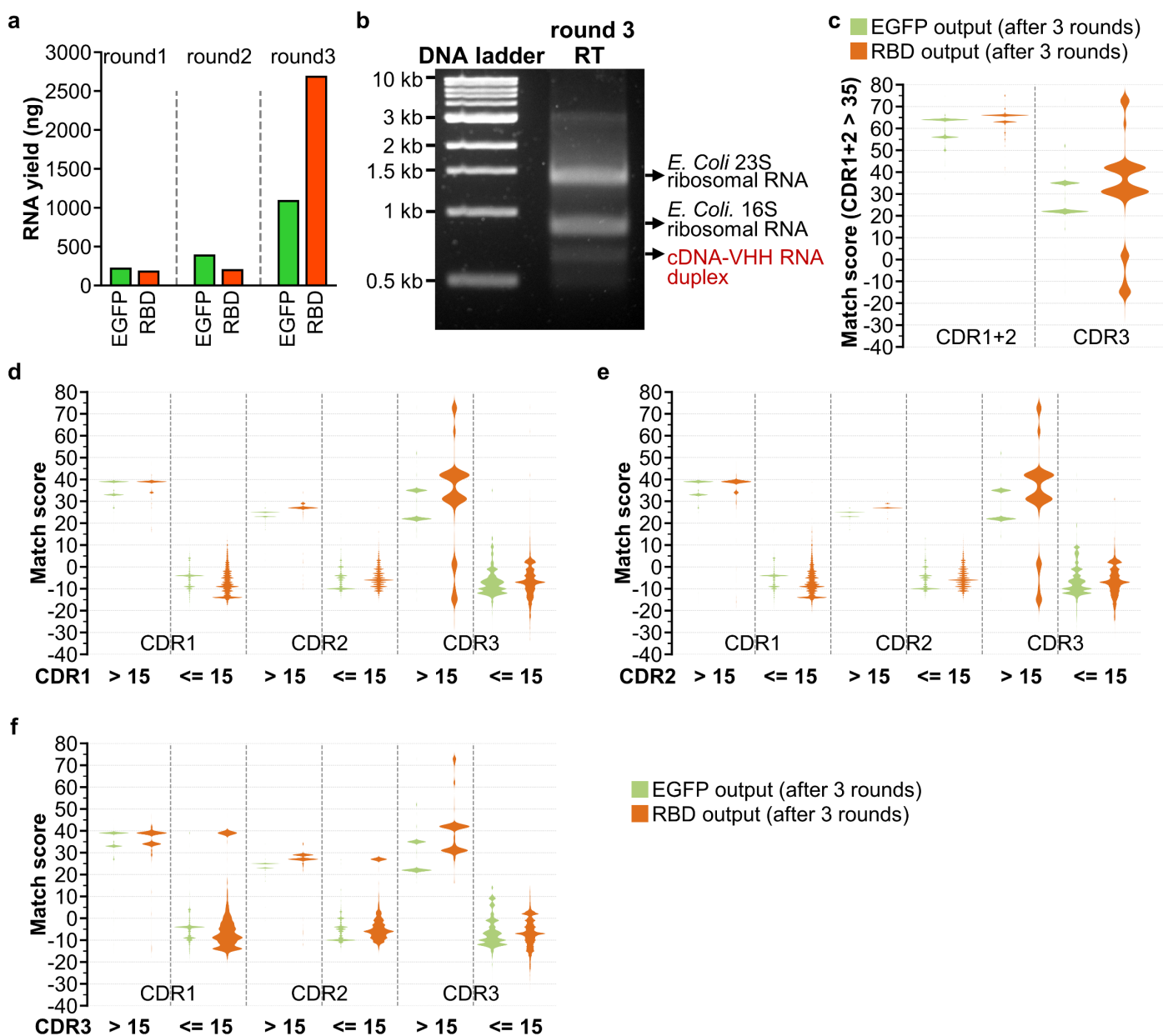
677 **Fig. S3. Working principles for orientation-controlled ligation by end blocking using**
678 **hairpin oligos.** (a) working principle for generating one end blocked DNA for orientation-
679 controlled ligation by PCR using a hairpin DNA oligo. (b) Representative orientation-controlled
680 ligation products visualized by agarose gel electrophoresis.

681

682

683

Fig S4

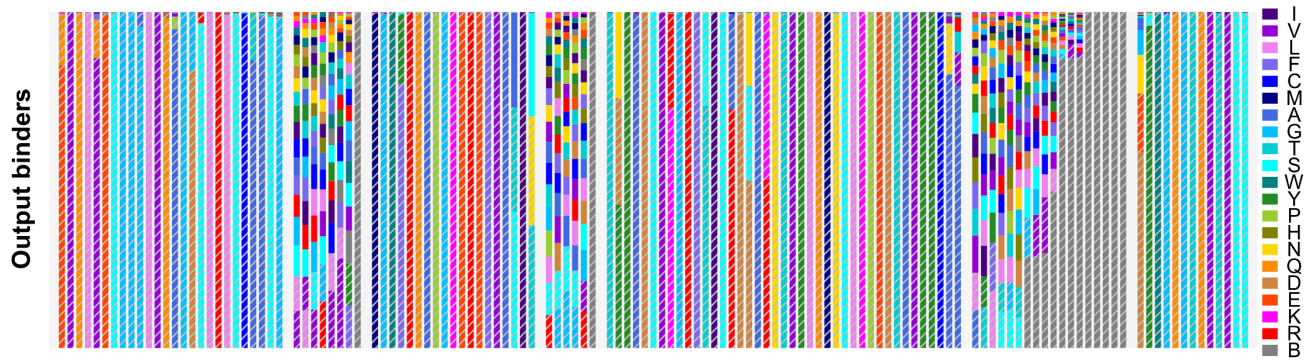


684 **Fig. S4. Evaluation of ribosome display and selection rounds.** (a) Yield of recovered RNA at
 685 each round of ribosome display and selection for EGFP or RBD targets. (b) Representative RT
 686 reaction (without heat denaturation) product for RBD selection after 3 rounds, visualized by
 687 agarose gel electrophoresis. (c) Plot of match scores of sequence pairs with a combined CDR1
 688 and CDR2 score > 35. (d) Plot of match scores of sequence pairs (from 2000 randomly sampled
 689 sequences) with indicated CDR1 scores, and (e) indicated CDR2 scores, and (f) indicated CDR3
 690 scores.

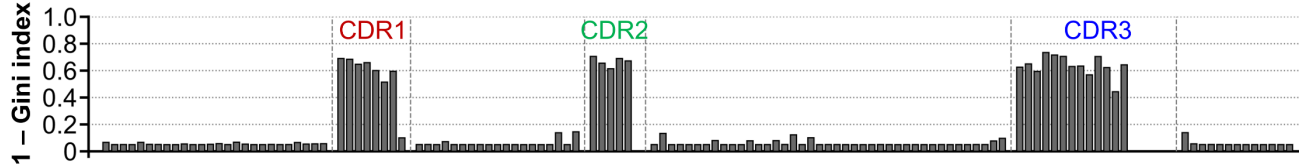
691

Fig S5

a



b



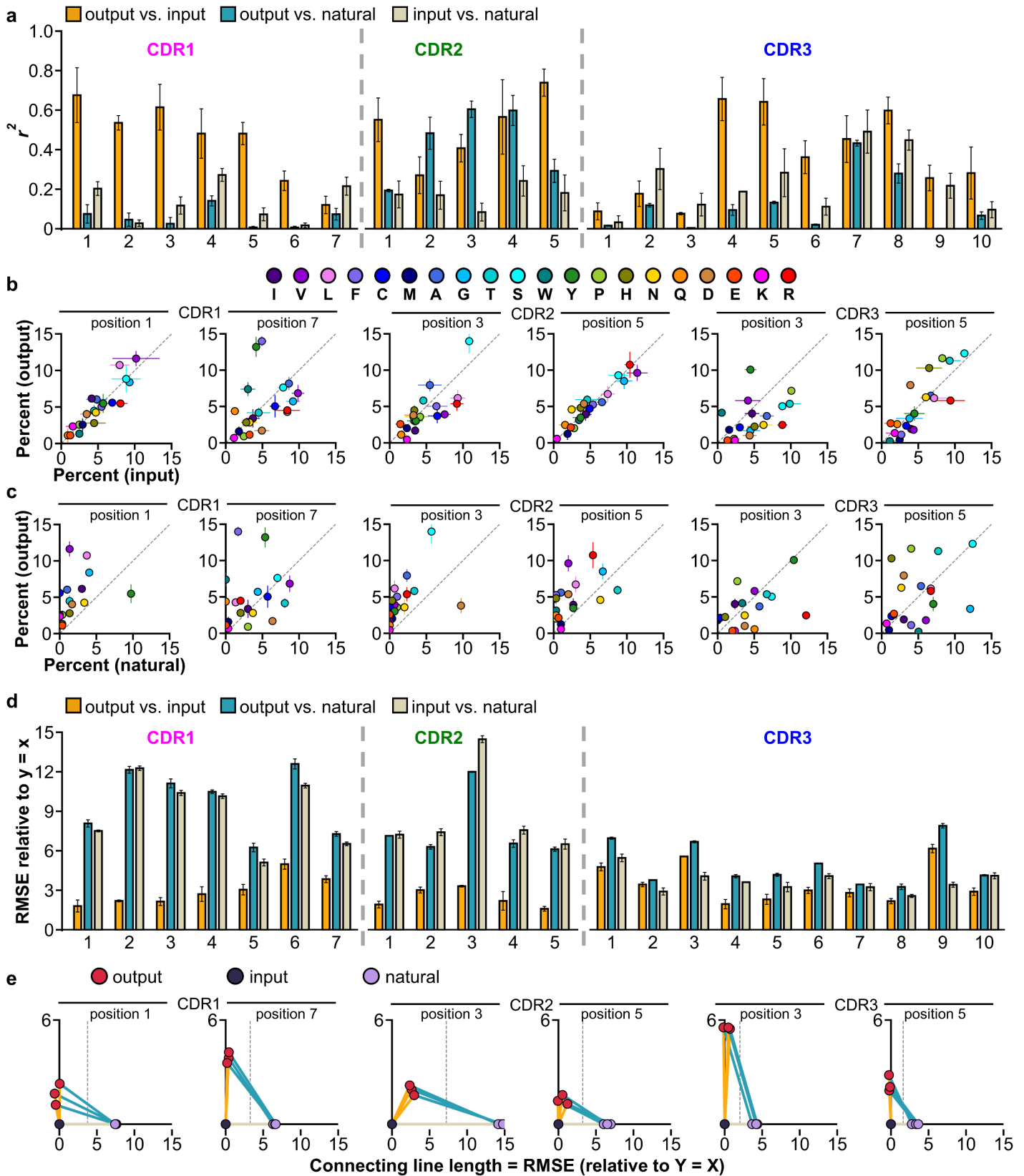
692 **Fig. S5. Amino acid profile for EGFP and RBD unique output binders.** (a) Amino acid
693 profile of representative VHH sequence for each unique cluster identified from RBD and EGFP
694 output libraries (“output binders”, 932 sequences). Plotted as described in **Fig S1a**. (b) Plot of
695 diversity index (as $1 - \text{Gini index}$) for each amino acid position of output binder VHHs.

696

697

698

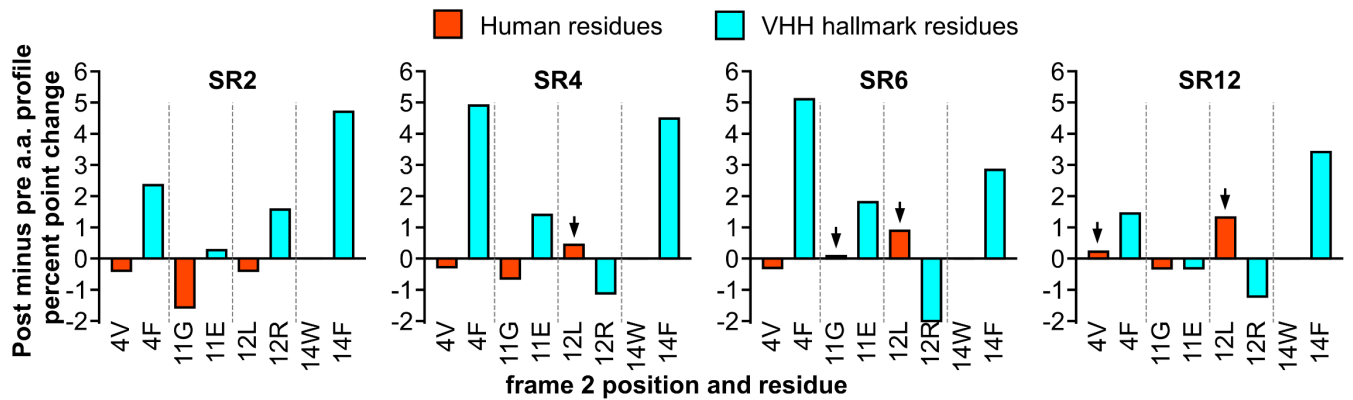
Fig S6



699 **Fig. S6. Unique output binders amino acid profiles correlate more highly with that of input**

700 **library than natural VHHs.** (a) r^2 values for the amino acid percentages in the indicated
701 sequence group pairs at each CDR position. 298 natural VHHs (natural) and 298 randomly
702 sampled sequences from input library (input) and output binders (output) were analyzed. Three
703 random sampling trials were performed to generate three r^2 for each position. (b) Scatter plots of
704 the percentage of each amino acid in the input library and the output binders and (c) that in the
705 natural VHHs and the output binders at representative CDR positions. Circles are the mean and
706 error bars are the standard deviation of data. (d) Root mean square error (RMSE, relative to $Y =$
707 X line) values for the indicated sequence group pairs at each CDR position. Using the same
708 randomly sampled sequences as (a). (e) Plot showing the similarity distances between the three
709 sequence groups, with each connecting line length between two sequence groups indicating their
710 RMSE. Vertical dashed lines indicate the middle point of the distance between output and
711 natural sequence groups
712

Fig S7



713 **Fig. S7. Affinity maturation leads to some VHH hallmark residues converting to the**
714 **corresponding human VH residues.** The post- minus pre-affinity maturation percent point
715 change of VHH hallmark residues and the corresponding human residues for each VHH. Arrows
716 indicate human residues with increased frequency as a result of affinity maturation.