Fragment-based computational design of antibodies targeting structured epitopes

Mauricio Aguilar Rangel[1,3], Alice Bedwell[1†], Elisa Costanzi[2], Stefano Ricagno[2], Judith Frydman[3], Michele Vendruscolo[1*] & Pietro Sormanni[1*]

[1]*Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.*
[2] *Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milano, Italy.*
[3] *Department of Biology, Stanford University, Stanford, California, USA.*
[†] *Current address: Chemistry Research Laboratory, University of Oxford, Oxford OX1 3TA, UK.*

[*] *Corresponding authors: ps589@cam.ac.uk (P.S.) & mv245@cam.ac.uk (M.V.)*

De novo design methods hold the promise of reducing the time and cost of antibody discovery, while enabling the facile and precise targeting of specific epitopes. Here we describe a fragment-based method for the combinatorial design of antibody binding loops and their grafting onto antibody scaffolds. We designed and tested six single-domain antibodies targeting different epitopes on three antigens, including the receptor-binding domain of the SARS-CoV-2 spike protein. Biophysical characterisation showed that all designs are highly stable, and bind their intended targets with affinities in the nanomolar range without any *in vitro* affinity maturation. We further show that a high-resolution input antigen structure is not required, as our method yields similar predictions when the input is a crystal structure or a computer-generated model. This computational procedure, which readily runs on a laptop, provides the starting point for the rapid generation of lead antibodies binding to pre-selected epitopes.

Antibodies are key tools in biomedical research, and are increasingly employed to diagnose and treat a wide range of human diseases. Currently, there are over 110 approved antibody drugs and about 90 in late-stage clinical trials (*1*). Existing antibody discovery methods have been widely successful, but still have important limitations (*2*). Extensive laboratory screenings are required to isolate those antibodies binding to the intended target, which can be time consuming and costly. Some classes of hard targets remain, including some receptors and channels, proteins within highly homologous families, aggregation-prone peptides, and disease-related short-lived protein aggregates (*3*, *4*). Most notably, it is often highly challenging to obtain antibodies targeting pre-selected epitopes. Screening procedures typically select for the tightest binders, which usually occur for immunodominant epitopes, thus disfavouring the discovery of antibodies with lower affinities but binding to functionally relevant sites (*5*). Finally, screening campaigns often yield antibodies with good binding

affinity, but otherwise poor biophysical properties, such as stability, solubility, and production yield, which may hinder their development into effective reagents.

Computational antibody design has the potential to overcome these limitations by drastically reducing time and costs of antibody discovery, and in principle allowing for a highly controlled parallel screening of multiple biophysical properties. Moreover, rational design inherently enables the targeting of specific epitopes. Most available methods for the design of binding proteins rely at least in part on the minimisation of a calculated interaction free energy, through the sampling of the mutational and conformational space (*2*, *6*, *7*). The nature of these calculations, which are based on molecular modelling, and the challenges of achieving exhaustive sampling, make design simulations rather imprecise and highly resource intensive. For these reasons, the de novo design of antibody binding has generally met low success rates, and required recursive experimental screenings and large libraries (*5*, *8–10*), which hamper its competitiveness with established laboratory-based technologies. Computational design of binding has been most successful in synergy with *in vitro* affinity maturation, and in particular when applied to mini-proteins (*11*, *12*). The small size of these mini-proteins is amenable to the high-throughput gene synthesis required to experimentally screen designed candidates on a massive scale, and their rigidity reduces the need for accurate conformational sampling. However, antibody domains are considerably larger, and bind their target using complementarity determining regions (CDRs) located within hypervariable loops on the antibody surface, which are often extended and highly flexible.

Here, we describe a novel method to design antibody CDR loops targeting epitopes of known structures, or for which a structural model is available. Designed CDRs are then grafted onto antibody scaffolds, and further optimised computationally for solubility and conformational stability. Novel antibody-antigen interactions are designed by combining together protein fragments identified as interacting with each other within known protein structures.

De novo CDR-design strategy

In order to overcome some of the limitations of molecular modelling, in particular those associated with the approximations in accounting for interatomic interactions, we exploited the availability of large structural databases to implement a fragment-based procedure to design CDRs (paratope) complementary to a target epitope. To implement this idea, we compiled from the non-redundant Protein Data Bank (PDB) a database of CDR-like fragments and corresponding antigen-like regions. CDR-like fragments are defined as linear motifs structurally compatible with an antibody CDR loop, which may be found in any protein structure in the PDB, and antigen-like regions are those found interacting with any CDR-like fragment in the structures analysed (see Methods). The structure of the input epitope is

fragmented using two different strategies: (i) a linear fragmentation, which generates fragments of at least four consecutive residues, and (ii) a surface-patch fragmentation, which takes each residue and yields the closest $n \geq 4$ solvent-exposed residues in the three-dimensional structure of the epitope. Next, each epitope fragment is compared to the antigen-like regions to identify those with compatible backbone structure and similar sequence (see Methods). This procedure yields those CDR-like fragments from the database that interact with the identified antigen-like regions. These CDR-like structures are then rotated to match the orientation of the epitope, by superimposing each antigen-like region, together with its interacting CDR-like fragment(s), to the matching part of the epitope (**Figs. 1A**, **S1** and **S2**). When possible, different CDR-like fragments whose backbones are compatible with a single longer CDR loop are joined together to yield longer interacting motifs (**Fig. 1B**, see Methods)

Some of the original interactions of each CDR-like fragment may be affected when this fragment is transferred onto the epitope, for instance if the sequence of the antigen-like region is not identical to the corresponding epitope sequence, or if the epitope side-chains are found in different conformations. Similarly, new interactions may arise when a CDR-like fragment forms contacts with parts of the epitope that were not matched onto its antigen-like region. Therefore, an optimisation procedure is carried out to replace all CDR side-chains affected with side-chains that make more favourable interactions with the epitope, by defining a local structural motif and searching for compatible hits in the PDB90 database (**Figs. 1C** and **S2**, see Methods). Next, all generated CDR-like candidates are ranked according to the number of favourable interactions, the number of interactions that could not be optimised, and a solubility score calculated with the CamSol method (*13*). Finally, top-ranking candidates are combined together to generate longer CDR loops, or multiple candidates are grafted in different CDRs of the same Fv region, and new interactions that may emerge between scaffold and antigen or between designed CDR and scaffold are optimised to yield the final designs (**Figs. 1D** and **S1**, see Methods).

To demonstrate this approach, we tested it on single-domain antibodies, because of their monomeric nature, ease of production in prokaryotic systems, and small size. Nonetheless, the computational design pipeline described here can readily be applied to other antibody fragments, including whole Fv regions, on which designed CDRs can be structurally matched and grafted in the same way (see Methods).

Description of designs and biophysical characterization

We designed six single-domain antibodies for three different antigens by exploring two grafting strategies. Two designed single-domain antibodies target the SARS-CoV-2 spike protein receptor-binding domain (RBD), three human serum albumin (HSA), and one

pancreatic bovine trypsin (**Table 1**). HSA and trypsin were selected for the initial validation. Both are available off the shelf, and binding of therapeutic proteins to HSA is a key determinant of pharmacokinetics. Therefore, single-domain antibodies targeting HSA may provide a modular tool for enhancing the half-life of biologics(*14*). Conversely, trypsin offers the opportunity of testing the design strategy on poorly accessible convex epitopes harbouring an active site. The RBD of SARS-CoV-2 exemplifies the power of targeting specific epitopes, as binding to regions overlapping with, or close to the ACE2 receptor binding site, whilst avoiding glycosylation sites, is expected to yield neutralising antibody candidates, which would sterically hinder virus binding to the human cell receptor. In this case, we used as starting point for the design the first-released cryo-EM model of the SARS-CoV-2 spike protein in the prefusion conformation(*15*) (PDB ID 6VSB). The reason for this choice was to assess how the design strategy performs with a lower resolution structure used as input. Specifically, we ran the design on the surface of the up RBD around the ACE2-binding region, which has some regions of very low resolution (~6-8 Å)(*15*), and several missing residues in the model (**Figs. 1A** and **3A**).

All designed single-domain antibodies expressed well in *E. coli*, were obtained to high purity, and showed circular dichroism (CD) spectra fully compatible with a well-folded VH domain (**Fig. S3**, see Methods). All designs were highly stable, with a melting temperature at par or better than that of immune-system-derived nanobodies (**Table 1, Fig. S3C**). The crystal structure of DesAb-HSA-P1 further confirms the correct folding of the domain and the dynamic nature of its CDR3 loop, as expected for this particular single-domain antibody scaffold(*16*) (**Fig. S4**).

Two out of the three anti-HSA single-domain antibodies, DesAb-HSA-P1 and DesAb-HSA-P2 (**Table 1**), consisted in designed CDR-like peptides grafted in place of the CDR3 of a previously characterised single-domain antibody scaffold highly amenable to CDR3 substitutions(*16*, *17*) (**Table S1**). The third design, DesAb-HSA-D3, was made by structurally matching two separate CDR-like candidates onto two CDR-loops of a nanobody scaffold identified as highly compatible with these two binding motifs (**Fig. 2A**, see Methods). The first strategy provides the opportunity to test the de novo CDR design procedure by minimising possible complications arising from the grafting, while the second is a more complex approach that allows to design multiple CDR loops onto a scaffold structurally matched to the epitope. Binding to HSA was measured in solution with micro-scale thermophoresis (MST), which yielded K_D values ranging from 140 to 800 nM (**Fig. 2B-D,F**), while a control single-domain antibody showed extremely weak signal in this assay (**Fig. S5A**). As a comparison, a nanobody isolated with yeast-display from a state-of-the-art naïve library was recently reported to bind HSA with a K_D of 430 nM(*18*), which is in the same range as those of our de novo designs. To confirm the binding, we also carried out bio-layer interferometry (BLI) with immobilised HSA,

obtaining K_D values compatible with those measured in solution, while the trypsin-targeting DesAb-Tryp employed as a negative control gave no binding signal for HSA (**Fig. 2E** and **S5B**). However, DesAb-Tryp was able to bind its intended target trypsin, while DesAb-HSA-P1 and P2 showed no binding signal and were likely partly digested by the protease during the binding assay (**Fig. S6**).

BLI competition experiments show that DesAb-HSA-D3 and DesAb-HSA-P1 compete with each other for binding to HSA, as the binding of one is hindered by the presence of the other antigen-bound DesAb. Conversely, DesAb-HSA-P2 does not compete with the other two, as its binding is not affected by the presence or absence of other antigen-bound DesAbs (**Fig. 2G**). This competition behaviour is fully compatible with the rational design, as DesAb-HSA-D3 and DesAb-HSA-P1 were designed to target partly overlapping epitopes, while DesAb-HSA-P2 targets a different epitope on the opposite side of the antigen (**Fig. 2A**).

Similar to the anti-HSA DesAbs, the two designs made to target the RBD of the spike protein showed binding in the nanomolar range. More specifically, a BLI assay with immobilised glycosylated RBD yielded K_D values of 210 and 130 nM for DesAb-RBD-C1 and DesAb-RBD-C2, respectively (**Fig. 3**). Conversely, these two anti-RBD antibodies showed no binding signal for immobilised HSA employed as a negative control and as a blocker in the assay (**Fig. S5C**, see Methods). Finally, both anti-RBD DesAbs were able to compete with the binding of the human ACE2 receptor to the viral RBD, which suggests that affinity-matured versions of these DesAbs may have neutralizing potential (**Fig. 3D**).

Applicability of the design strategy

Having established that our computational method can yield stable single-domain antibodies that bind their intended targets with good K_D values, we asked how readily and generally applicable the design strategy is. The most apparent bottleneck of the pipeline is the need for a structure of the target epitope to be used as input. As structural determination can be challenging for some targets, this aspect could limit the applicability of the method, in particular in the cases of emerging diseases or of poorly investigated areas, where novel antibodies are often most needed.

To test the dependence of the design method on the quality of the input structure, we ran our CDR design procedure on all available experimental structures from the latest Critical Assessment of Techniques for Protein Structure Prediction competition (CASP14), as well as all corresponding models generated with AlphaFold2, which was the best-performing algorithm assessed (*19*, *20*). By using all available models for each structure, we make sure to include in our analysis also lower quality models that were not top-ranking in CASP (see

Methods). Our results reveal that a large fraction of the designed CDR fragments obtained by using each model as input are identical to those obtained using the corresponding experimentally determined structure (**Fig. 4A**). More specifically, the median number of designed CDRs in common between each model and its corresponding experimental structure, expressed as a percent of the total number of designed CDRs obtained for each model, is 77%, and only 20 (10%) of the 200 models analysed have less than 50% CDRs in common with their target structures (**Fig. 4A**, **S7** and **Table S3**). These results suggest that if one were to use an AlphaFold2 model as input for our antibody design pipeline, most typically about 75% of the generated CDRs would be identical to those that would be obtained from the corresponding crystal structure, and at least 50% would be identical in 90% of the cases. Besides, we only observe a very weak correlation ($R^2 = 0.06$) between the percent of CDRs in common among model and structure, and the quality of the model itself as quantified by the global distance test total score (GDT, **Fig. 4B**). This result indicates that the aforementioned median number of designed CDRs in common among model and structure is not excessively determined by those very high-quality models (GDT $\geq$ 90) that are almost identical to their target structure. Taken together, these results imply that the CDR-design procedure is expected to yield similar results when running on computer-predicted models or on experimental structures, and that these results do not strongly depend on the quality of the model used as input, at least within the range we explored (GDT $\geq$ 40).

The target structures of the CASP assessments are selected ensuring that they represent a diverse sample of native folds, characterized by different secondary structures and overall shape (*20*). Therefore, these structures also constitute a particularly suitable test-set to explore the applicability of our design strategy, by enabling to address the question of how often our fragment-based combinatorial procedure can yield CDRs targeting a given input epitope. To this end we computed the solvent accessible surface area of each experimental structure in the presence and absence of bound designed CDRs. Our results reveal that the vast majority of the surface of each antigen is typically targetable with our strategy, with a median surface coverage of 78% (**Fig. 4C**). Furthermore, for each epitope there are typically many candidate binding loops to choose from, with a median density of 19 CDRs per nm^2 of antigen surface (**Fig. 4D**).

Conclusions

We have described a fragment-based strategy for the rational design of antibodies targeting structured epitopes. We use protein fragments of at least four residues and typically longer in order to assemble designed CDRs in a combinatorial way. The idea behind this choice is that such fragments should be large enough to contain nontrivial sequence determinants of structure and interactions (*6*, *16*, *21*).

Our experimental results demonstrate that the design pipeline that we presented can successfully yield stable single-domain antibodies binding their intended targets, with K_D values down to the nanomolar range (**Table 1**). Importantly, we have been able to achieve this outcome without the need of experimentally screening a large number of designs, but rather by pre-selecting in silico those designed CDRs that appeared most promising according to the metrics implemented, which include proxies for the predicted binding and sidechain complementarity, as well as predictions of solubility (*13*) (see Methods).

This combinatorial approach does not involve approximations to calculate interaction free energies, and is also substantially faster than approaches based on the sampling of conformational and mutational space (*2*). An intrinsic limitation of this strategy, however, is that its applicability to epitopes of interest depends on the availability of suitable CDR-like fragments in the databases used. Nonetheless, the growing number of available protein structures in public databases makes the procedure generally applicable, as for most epitopes one obtains a number of candidate CDRs to choose from (**Figs. 4C,D** and **S1**).

Our results, which are obtained with a computer code that can run on standard laptops, demonstrate that it is becoming increasingly possible to design de novo antibodies binding to pre-selected epitopes of interest. We have exploited recent advances in protein-folding predictions and ab initio structural modelling to show that our design pipeline yields similar results when running on experimental structures or on computer-generated models, even when these do not reach atomistic accuracy. We envisage that, taken together, these advances in computational biotechnology will enable in the near future to obtain lead antibodies in a matter of days from the release of a pathogen genome, or from the identification of a novel disease-relevant target.

Acknowledgements

We are grateful to Dr Paul Brear for useful advice on, and help with setting up protein crystallisation trials in Cambridge (Department of Biochemistry); to Klara Kulenkampff for an aliquot of fluorescently labelled single-domain antibody (KK5 in Fig. S5); to Dr Faidon Brotzakis for useful discussions around the structural dynamics of the SARS-CoV-2 spike protein; and to Matthias Schneider and Prof. Tuomas Knowles for useful discussion on, and help with biophysical measurements of protein interactions in solution.

P. S. is a Royal Society University Research Fellow (URF\R1\201461). S. R. and E. C. were supported by Fondazione ARISLA (project TDP-43-STRUCT). We acknowledge support from the Protein Interactions and Stability in Medicine and Genomics Challenge Programme (PRISM; NNF18OC0033950) funded by the Novo Nordisk Foundation.

Table 1. Designed single-domain antibodies (DesAbs) employed in this study.

	Target antigen	Designed CDR	Target epitope†	Scaffold (pdb)*	T_m (ºC)‡	K_D (nM)§	Buried SASA (Å^2)
DesAb-HSA-P1	HSA	IQKSLQTAESIL	575-582	6Z3X	82.5	120	247
DesAb-HSA-P2	HSA	AQAGNAEEAE	71-80	6Z3X	80	380	141
DesAb-HSA-D3	HSA	ELYALI (CDR1) KFASPDGS (CDR3)	542-546,574-580	4DKA	67.5	180	230
DesAb-Tryp	Trypsin	QSGYHF	698-702	6Z3X	78.5	1800	195
DesAb-RBD-C1	Spike RBD	GSSATEVY	449,453,492-497,500	6Z3X	77.5	210	356
DesAb-RBD-C2	Spike RBD	VVADLSV	353-359	6Z3X	80	130	366

† Residue numbering as in PDB entries 1AO6 chain B (HSA), 1S0Q chain (trypsin) and 6VSB chain A (spike RBD).

* PDB ID of scaffold in whose loop the designed CDRs are grafted, see Methods. 6Z3X is from this study.

‡ Melting temperature rounded to the closest 0.5 ºC to reflect the accuracy of the thermal-shift assay employed (see Methods and **Fig. S3C**).

§As measured with bio-layer interferometry (BLI), rounded to the closest 10 nM with exception of DesAb-Tryp, which was rounded to 100 nM (**Fig. S6**)

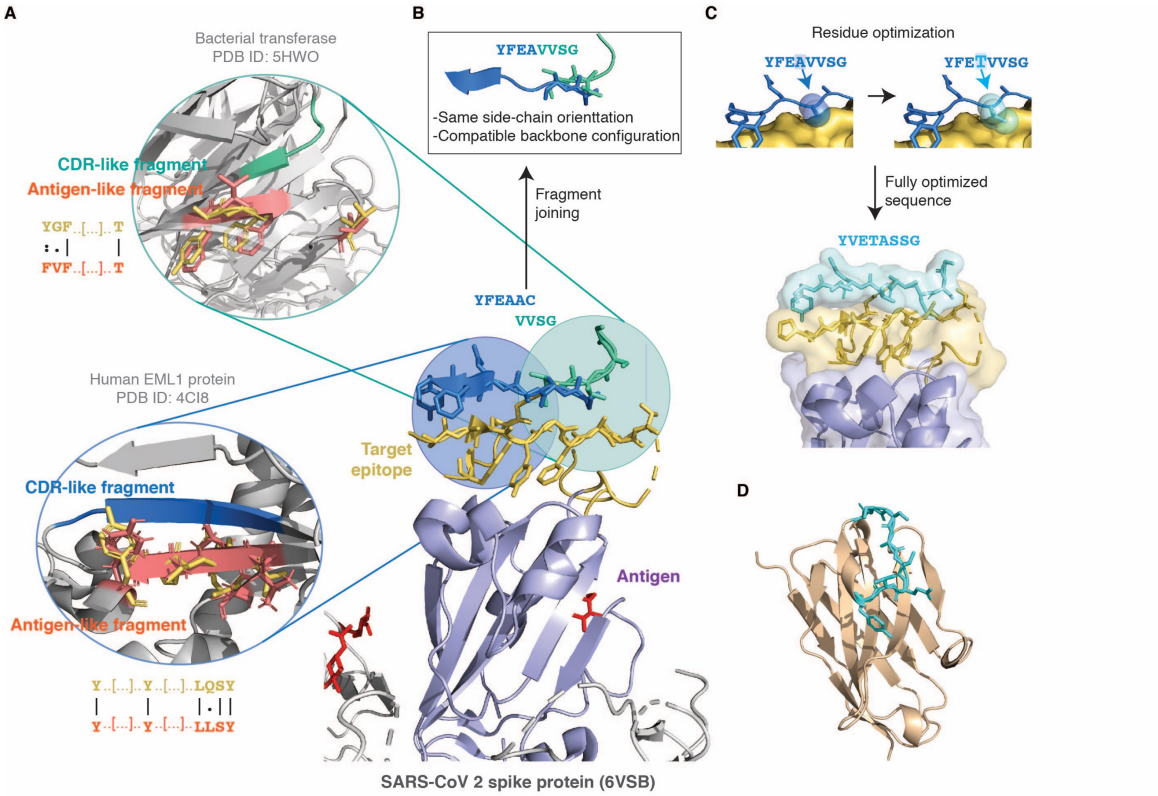
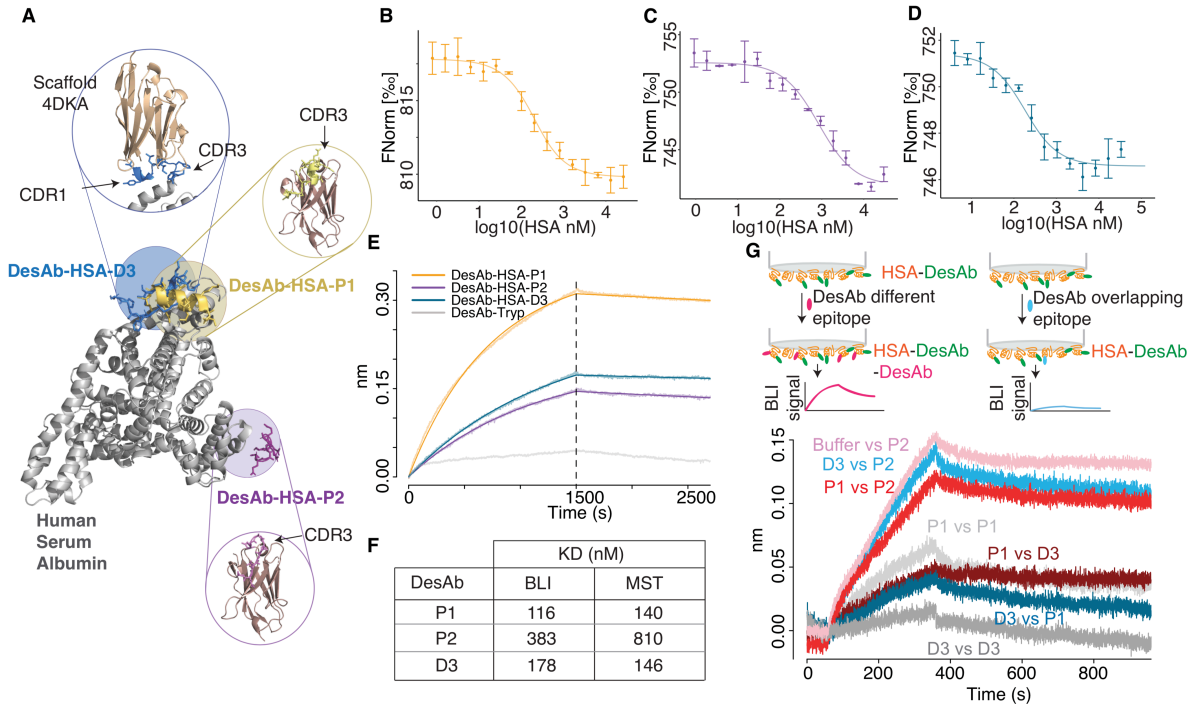


Figure 1. Schematic illustration of the combinatorial structure-based CDR design strategy introduced in this work. (**A**) The antigen is shown in purple with the glycan groups in red, and the chosen target epitope is shown in gold at the centre (the ACE2 binding site in the RBD of SARS-CoV-2, PDB ID 6VSB). The bubbles on the left show two matching antigen-like fragments (salmon) in the context of their native structure structurally superimposed to the query epitope fragments (gold) used to identify them. The CDR-like fragments (blue and green) interacting with these antigen-like fragments, and therefore predicted to bind the epitope, are also shown above it in the centre. The antigen-like fragments in the bubbles are non-contiguous as they were identified with the surface-patch fragmentation mode (see Methods) by employing as query those epitope fragments corresponding to solvent-exposed sidechains. The sequence identity between the query epitope fragment (gold) and matching antigen-like fragment (salmon) is shown as an alignment next to the bubbles. (**B**) As the two CDR-like fragments have an overlapping part with compatible backbone conformation and similar sidechain orientation they are merged together to form a single designed CDR candidate. (**C**) The sequence of the CDR candidate is further optimised (see Methods) by substituting those residues in contact with amino acids that differ between the target epitope and the matching antigen-like fragments. (**D**) The optimised CDR sequence is then grafted onto a single-domain antibody scaffold. The example presented in this figure is the CDR3 of DesAb-RBD-C1 (**Table 1**). The model in this panel was generated with the ABodyBuilder web server (*22*).



	KD (nM)	
DesAb	BLI	MST
P1	116	140
P2	383	810
D3	178	146

Figure 2. The anti-HSA DesAbs bind their target and compete for binding to overlapping epitopes. (**A**) The structure of HSA is shown in grey, and the designed CDR fragments targeting different epitopes are in different colours. The bubbles show structural models of the DesAbs with the designed CDRs grafted onto the indicated loops. DesAb-HSA-D3 contains two different designed CDRs, which are grafted onto a structurally-matched scaffold. (**B-D**) Micro-scale thermophoresis (MST) of fluorescently-labelled DesAbs (70 nM) in the presence of increasing concentrations of HSA (x-axis). The colours are the same of panel A, with DesAb-HSA-P1 in yellow (**B**), P2 in purple (**C**) and D3 in blue (**D**). Data points are mean and standard deviations of three replicates, data were fitted with a single-site binding model. (**E**) BLI binding traces (association and dissociation) obtained with APS-sensors loaded with HSA. Association was monitored in wells containing 2 µM DesAbs, and colours are like in panels A-D. 2 µM DesAb-Tryp is shown in grey and was employed as control for non-specific binding to the sensor. (**F**) Table with the dissociation constants (K_D) obtained for the three DesAbs by fitting the BLI and MST data. (**G**) Binding competition experiment at the BLI. APS-sensors were loaded with HSA, and then dipped in wells containing 5 µM of a first DesAb *X1* (see Methods), then moved in buffer wells for one minute, and then into wells containing 5 µM of a second DesAb *X2*, and finally back to buffer wells. Curves are labelled with "*X1 vs X2*" to identify the anti-HSA DesAbs employed. The plot shows the last three steps, and reference sensors monitoring the background dissociation of *X1* during these steps were subtracted from the traces shown here. The traces *P1 vs P1* and *D3 vs D3* were taken as positive controls for the competition, and the small signal observed is due to the facts that not all epitopes are

occupied by the first DesAb (*X1*), and that this is dissociating in the background. The trace *Buffer vs P2* was taken as a negative control for the competition.

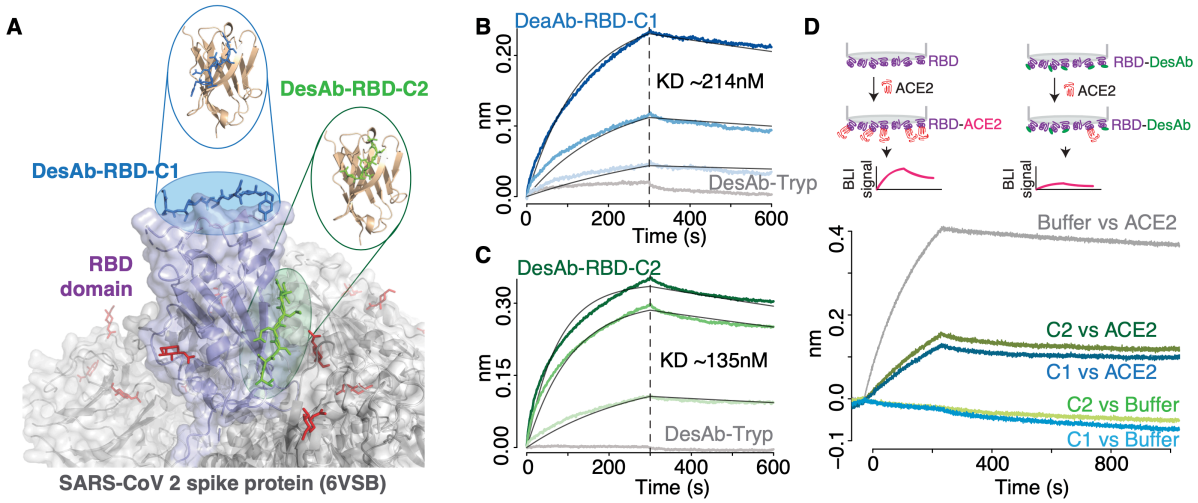


Figure 3. The anti-RBD DesAbs bind their target and compete with human ACE2. (**A**) the RBD is shown in purple, and the rest of the SARS-CoV-2 spike protein in grey with glycans in red. The designed CDRs targeting the RBD are in blue and green respectively for DesAb-RBD-C1 and C2, and corresponding structural models of the single-domain antibodies are represented in the bubbles. (**B, C**) BLI binding traces (association and dissociation) obtained with APS-sensors loaded with RBD and blocked with HSA. Grey traces are obtained with 4 µM of DesAb-Tryp used as a negative control to probe non-specific binding to the sensors. (**B**) 4 µM, 2 µM and 1 µM of DesAb-RBD-C1 (from darker to lighter blue, K_D= 214 ± 4 nM). (**C**) 4 µM, 2.5 µM and 1 µM of DesAb-RBD-C2 (from darker to lighter green, K_D= 135 ± 2 nM). Data were fitted globally to estimate the reported K_D values. (**D**) Binding competition experiment at the BLI. Sensors were loaded like in (B, C) and dipped in wells with 5 µM DesAb-RBD-C1 (blues) or DesAb-RBD-C2 (greens) or buffer (grey), then in wells containing ACE2 or buffer controls (see legend), and finally back to buffer. The plot reports the last two steps, showing that the binding of ACE2 is substantially reduced by the presence of either DesAb-RBD-C1 or DesAb-RBD-C2 bound to the RBD.

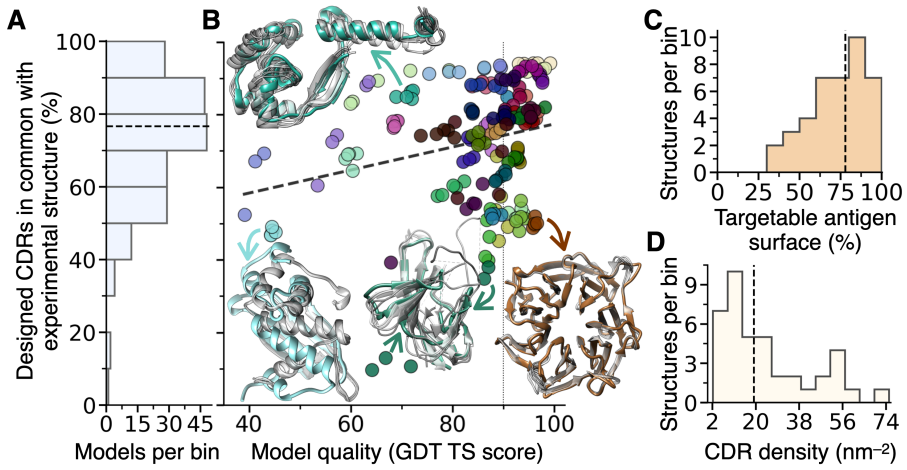


Figure 4. Performance of the CDR-design procedure on computationally predicted antigen structures. De novo modelled structures generated by the AlphaFold2 algorithm within the CASP14 competition, as well as the corresponding experimentally determined structures, were used as input for the CDR design procedure. (**A**) Histogram of the distribution of the percent of designed CDRs obtained from each model that were identical to those obtained from the corresponding structure. The horizontal dashed line is the median of the distribution at 76.6%. (**B**) Scatter plot of the same CDR percent (y-axis) as a function of the global distance test total score (GDT TS, x-axis), as reported by the CASP14 competition, which is an indicator of the model accuracy. GDT works with the percentage of α-carbons that are found within certain cut-off distances of each other. A GDT of 100 means the modelled and experimental structure have all α-carbons within 1 Å of each other, and one above 90 (vertical dotted line) is typically regarded as a good solution of the folding prediction. The dashed trendline corresponds to a weak correlation ($R^2 = 0.06$). Data point are coloured according to the target experimental structure of each model (see **Table S3** and **Fig. S7**). Four example structures are drawn in the same colour as their model data points, which are pointed by the arrows. Their models are overlaid to the structures and shown in grey. (**C**) Histogram of the distribution of the percent of targetable surface area for each antigen (experimental structure from CASP14). Targetable surface is defined as the solvent-accessible surface area of the antigen made inaccessible by at least one designed CDR. The dashed line is the median at 78%. (**D**) Histogram of the distribution of the CDR density for each antigen, expressed as the mean number of different designed CDRs per nm^2 of antigen surface. The dashed line is the median at 19.2 CDRs per nm^2.

References

1. H. Kaplon, M. Muralidharan, Z. Schneider, J. M. Reichert, Antibodies to watch in 2020. *mAbs*. **12**, 1703531 (2020).

2. P. Sormanni, F. A. Aprile, M. Vendruscolo, Third generation antibody discovery methods: in silico rational design. *Chem. Soc. Rev.* **47**, 9137–9157 (2018).

3. F. A. Aprile, P. Sormanni, M. Podpolny, S. Chhangur, L.-M. Needham, F. S. Ruggeri, M. Perni, R. Limbocker, G. T. Heller, T. Sneideris, T. Scheidt, B. Mannini, J. Habchi, S. F. Lee, P. C. Salinas, T. P. J. Knowles, C. M. Dobson, M. Vendruscolo, Rational design of a conformation-specific antibody for the quantification of Aβ oligomers. *Proc. Natl. Acad. Sci.* **117**, 13509–13518 (2020).

4. C. J. Hutchings, P. Colussi, T. G. Clark, Ion channels as therapeutic antibody targets. *mAbs*. **11**, 265–296 (2019).

5. S. Fischman, Y. Ofran, Computational design of antibodies. *Curr. Opin. Struct. Biol.* **51**, 156–162 (2018).

6. J. Zhou, A. E. Panaitiu, G. Grigoryan, A general-purpose protein design framework based on mining sequence–structure relationships in known protein structures. *Proc. Natl. Acad. Sci.* **117**, 1059–1068 (2020).

7. P.-S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature*. **537**, 320–327 (2016).

8. G. Nimrod, S. Fischman, M. Austin, A. Herman, F. Keyes, O. Leiderman, D. Hargreaves, M. Strajbl, J. Breed, S. Klompus, K. Minton, J. Spooner, A. Buchanan, T. J. Vaughan, Y. Ofran, Computational design of epitope-specific functional antibodies. *Cell Rep.* **25**, 2121-2131.e5 (2018).

9. D. Baran, M. G. Pszolla, G. D. Lapidoth, C. Norn, O. Dym, T. Unger, S. Albeck, M. D. Tyka, S. J. Fleishman, Principles for computational design of binding antibodies. *Proc. Natl. Acad. Sci.* **114**, 10900–10905 (2017).

10. K. C. Entzminger, J. Hyun, R. J. Pantazes, A. C. Patterson-Orazem, A. N. Qerqez, Z. P. Frye, R. A. Hughes, A. D. Ellington, R. L. Lieberman, C. D. Maranas, J. A. Maynard, De novo design of antibody complementarity determining regions binding a FLAG tetra-peptide. *Sci. Rep.* **7**, 10295 (2017).

11. A. Chevalier, D.-A. Silva, G. J. Rocklin, D. R. Hicks, R. Vergara, P. Murapa, S. M. Bernard, L. Zhang, K.-H. Lam, G. Yao, C. D. Bahl, S.-I. Miyashita, I. Goreshnik, J. T. Fuller, M. T. Koday, C. M. Jenkins, T. Colvin, L. Carter, A. Bohn, C. M. Bryan, D. A. Fernandez Velasco, L. Stewart, M. Dong, X. Huang, R. Jin, I. A. Wilson, D. H. Fuller, D. Baker, Massively parallel de novo protein design for targeted therapeutics. *Nature*. **550**, 74–79 (2017).

12. L. Cao, I. Goreshnik, B. Coventry, J. B. Case, L. Miller, L. Kozodoy, R. E. Chen, L. Carter, A. C. Walls, Y.-J. Park, E.-M. Strauch, L. Stewart, M. S. Diamond, D. Veesler, D. Baker, De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science*, eabd9909 (2020).

13. P. Sormanni, F. A. Aprile, M. Vendruscolo, The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* **427**, 478–490 (2015).

14. B. M. Tijink, T. Laeremans, M. Budde, M. Stigter-van Walsum, T. Dreier, H. J. de Haard, C. R. Leemans, G. A. M. S. van Dongen, Improved tumor targeting of anti-epidermal growth factor receptor Nanobodies through albumin binding: taking advantage of modular Nanobody technology. *Mol. Cancer Ther.* **7**, 2288–2297 (2008).

15. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. **367**, 1260–1263 (2020).

16. P. Sormanni, F. A. Aprile, M. Vendruscolo, Rational design of antibodies targeting specific epitopes within intrinsically disordered proteins. *Proc. Natl. Acad. Sci.* **112**, 9902–9907 (2015).

17. F. A. Aprile, P. Sormanni, M. Perni, P. Arosio, S. Linse, T. P. J. Knowles, C. M. Dobson, M. Vendruscolo, Selective targeting of primary and secondary nucleation pathways in Aβ42 aggregation using a rational antibody scanning method. *Sci. Adv.* **3**, e1700488 (2017).

18. C. McMahon, A. S. Baier, R. Pascolutti, M. Wegrecki, S. Zheng, J. X. Ong, S. C. Erlandson, D. Hilger, S. G. F. Rasmussen, A. M. Ring, A. Manglik, A. C. Kruse, Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Biol.* **25**, 289–296 (2018).

19. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, K. Tunyasuvunakool, O. Ronneberger, R. Bates, A. Žídek, A. Bridgland, C. Meyer, S. A. A. Kohl, A. Potapenko, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, M. Steinegger, M. Pacholska, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, High Accuracy Protein Structure Prediction Using Deep Learning. *Fourteenth Crit. Assess. Tech. Protein Struct. Predict. Abstr. Book* (2020).

20. A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, J. Moult, Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins Struct. Funct. Bioinforma.* **87**, 1011–1020 (2019).

21. V. Frappier, J. M. Jenson, J. Zhou, G. Grigoryan, A. E. Keating, Tertiary Structural Motif Sequence Statistics Enable Facile Prediction and Design of Peptides that Bind Anti-apoptotic Bfl-1 and Mcl-1. *Structure*. **27**, 606-617.e5 (2019).

22. J. Dunbar, K. Krawczyk, J. Leem, C. Marks, J. Nowak, C. Regep, G. Georges, S. Kelm, B. Popovic, C. M. Deane, SAbPred: a structure-based antibody prediction server. *Nucleic Acids Res.* **44**, W474-8 (2016).