Gluebodies improve crystal reliability and diversity through transferable nanobody mutations that introduce constitutive crystal contacts

Mingda Ye[1,2], Mpho Makola[1,2], Joseph A. Newman[1,2], Michael Fairhead[1,2], Elizabeth Maclean[1,2,8], Nathan D. Wright[1,2], Lizbé Koekemoer[1,2], Andrew Thompson[1,2], Gustavo A. Bezerra[1,2,14], Gangshun Yi[8], Huanyu Li[1,2], Victor L. Rangel[10,13], Dimitrios Mamalis[5,6], Hazel Aitkenhead[1,2], Benjamin G. Davis[5,6,7], Robert J.C. Gilbert[8,9], Katharina Duerr[1,2,11], Opher Gileadi[1,2,12] and Frank von Delft[1,2,3,4]

1. Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK
2. Structural Genomics Consortium, University of Oxford, Old Road Campus, Roosevelt Drive, Headington, OX3 7DQ, UK
3. Diamond Light Source, Harwell Science and Innovation Campus, Didcot, United Kingdom
4. Department of Biochemistry, University of Johannesburg, Auckland Park, 2006, South Africa
5. Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford, UK
6. The Rosalind Franklin Institute, Oxfordshire, Oxford, UK
7. Department of Pharmacology, University of Oxford, Oxford, UK
8. Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.
9. Calleva Research Centre for Evolution and Human Sciences, Magdalen College, University of Oxford, Oxford OX1 4AU, UK
10. Laboratory of Protein Crystallography, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo, Brazil
11. Present address: OMass Therapeutics, Ltd., Oxford, UK
12. Present address: Exscientia Ltd, Oxford, UK
13. Present address: Evotec Ltd, Oxford, UK
14. Present address: Bicycle Therapeutics Plc, Cambridge, UK

Abstract

Obtaining protein crystals suitable for characterizing ligands and understanding different protein conformations remains a bottleneck for structural studies. Using nanobodies as crystallization chaperones is one strategy to address the problem, but its reliability is uncharacterized and in this study, we observed it to have a poor success rate. We therefore set out to engineer robust crystallization behaviour into the nanobody scaffold by exploring the nanobody-nanobody interface predominant in the PDB using >200 combinations of surface mutations. This yielded multiple polymorphs, all mediated by this interface, with one providing far superior resolution and reliability of diffraction. We therefore refer to the modified nanobody scaffolds as “Gluebodies” and propose at least 2 variants that should be routinely generated for chaperone experiments. We furthermore show that Gluedodies cannot rescue intrinsically non-crystallizing proteins, but are a powerful approach to improve the packing and

resolution limit of poorly diffracting crystals.

Introduction

Protein X-ray crystallography has long ago become a routinely used method to elucidate the near-atomic structure of proteins of interest. However, getting the protein to crystallize remains highly troublesome, with no substantial methodological breakthroughs for many decades. The mechanistic specifics remain poorly understood, so that there is no systematic route to coax a protein to pack in an ordered lattice[1-5]. Nevertheless, numerous techniques have emerged throughout the past decade to aid crystallization and they generally follow one of three general strategies.

The first is to vary the protein environment by exploring crystallization solution, temperature or physical format[6,7]. This approach has been thoroughly commercialized for over two decades, with a large repertoire of crystallization screens purchasable from vendors, and many laboratories equipped with various flavours of automation.

The second strategy is to modify the protein itself to favor crystallization[8-10], by introducing either major changes to the protein through varying the expression construct[10], or by subtle changes on the protein surface, such as by surface entropy reduction[11] or the "crystal epitope" approach[8].

The third approach is to introduce "chaperone" proteins that bind to the target of interest and discover additional routes for forming crystal contacts. Those include fusion tags (such as lysozyme[12] and BRIL[13]) and protein binders (such as Fabs[14], scFvs[15], nanobodies[16] and DARPins[17]). Since adding fusion tags might introduce experimental problems, like decreasing purified protein yields[10], chaperone-assisted protein crystallization is especially effective on difficult proteins like membrane proteins[9]. In addition, with the improvement and diversification of binder selection methods, generating chaperones to assist structural studies is now routinely attempted in many labs[18-20].

Other strategies have been reported, but have either not yet borne fruit (e.g. embedding the protein of interest into a highly porous crystal lattice[21-23]), or have not achieved widespread usage and thus validation (e.g. imprinted polymers[24], gravity[7]).

The third strategy above, binder-assisted crystallization, holds the promise to address several key crystallization challenges beyond obtaining the first crystal structure[9]. Firstly, they can lock the target protein into conformations that cannot be otherwise isolated[25,26]. Secondly, it can help find the considerably more robust crystallization systems required for structure-based lead discovery, including crystal-based fragment screening, since the introduction of new or additional crystal-crystal contacts mediated by the chaperone can systematically increase polymorphism[27,28]. Thirdly, by engineering binder scaffolds, symmetry can be introduced to the binder:target complex to facilitate crystallization processes[29,30].

Nevertheless, the approach does seem to retain significant opportunities for improvement. For instance, a recent study showed that the shortening of the FG loop on human Fabs by only two amino acids can significantly improve crystallization of human Fabs and Fab:antigen complexes[31]. Another more recent study on diabody scaffold engineering also showed that with an additional disulfide between the VH-VH interface on the diabody scaffold, the diabody:HIV-1 reverse transcriptase (RT) complex was able to crystallize in significantly more conditions than the wild type complex and the RT alone[32].

In the current study, we set about optimizing the crystallization behavior of nanobodies, as these are far simpler to produce than Fabs or diabodies. We first show that an ineffective nanobody, which does not help the DNA helicase RECQL5 crystallize, can be made effective by a limited number of

mutations on the nanobody scaffold. Secondly, through mutational screening of more than 200 constructs designed through data mining, we identify 4 key mutations on the RECQL5 nanobody scaffold far from the CDR surfaces, which lead to extensive crystal polymorphism and robust diffraction. Finally, we explored the ability of these engineered nanobodies to help non-crystallizing proteins form crystals.

Results

Naïve nanobodies are unsatisfactory crystallization chaperones

We selected 17 nanobody-target pairs across 5 different target proteins (Supplementary Figure 1 and Supplementary Table 1) and performed standard crystallization trials of the 10 protein complexes against two commercially available coarse screens (192 conditions, 2 temperatures) for each protein complex. This yielded crystals for only one protein complex (MAGEB1), which diffracted well and could be readily solved (PDB ID 6R7T). This implies the effectiveness of nanobodies as crystallization chaperones is less than 10%, demonstrating the need for dramatic improvement of the approach.

A single nanobody surface mutation yields an effective chaperone for RECQL5

The strategy of Crystal Epitope[8] entails identifying short sequence motifs (3-6 residues) that frequently appear in crystal contacts in the Protein Data Bank. We have previously successfully used this strategy to obtain crystals of several different target proteins (e.g., BRD1A: 5AMF, GALT: 6GQD, Glycogenin-2 catalytic domain: 4UEG), and thus applied it to the non-CDRs of a specific nanobody targeting RECQL5.

74 crystal epitope mutations (variant class G0 and variant G1-001) were designed through 3 iterations on the scaffold (non-CDR, non-variable region) of a RECQL5 nanobody (Supplementary Figure 2 and 3A, Supplementary Table 2). Purified nanobodies all co-migrated with RECQL5 on size exclusion chromatography, indicating adequate affinity of the nanobodies. The purified protein complexes were put into crystallization trials using the Hampton Index coarse screen, and 5 of the 74 yielded crystals (Supplementary Figure 3B). These variants shared one common mutation, D69Y, the single mutation of variant G1-001 (Supplementary Figure 2 and Supplementary Table 4) and crystals of the G1-001 complex diffracted to the highest resolution (2.76 Å). In the mutant structure, Y69 on the nanobody forms a hydrogen bond with D400 of RECQL5, which most likely stabilizes the crystal formation (Figure 5A).

Thus, non-CDR surface mutations can increase crystallization propensity while preserving nanobody affinity for its target protein. However, the effective mutations were not mediating nanobody-nanobody crystal contacts, and thus were only useful for their specific target RECQL5.

One nanobody-nanobody interface predominates in PDB structures

We instead devised a strategy that would be applicable to all nanobody-protein complexes, by engineering only nanobody-nanobody interactions. To understand how nanobodies commonly pack with each other in crystals, we examined the 335 nanobody-containing crystal structures in the PDB. We defined two molecules as forming crystal contact if their nearest atoms are below 4 Å apart[33], and thus found 273 structures with nanobody-nanobody contacts. Contacts with more than 5 interacting residues seem most suitable for engineering, and therefore only the 356 large contacts across 119 structures were further analyzed.

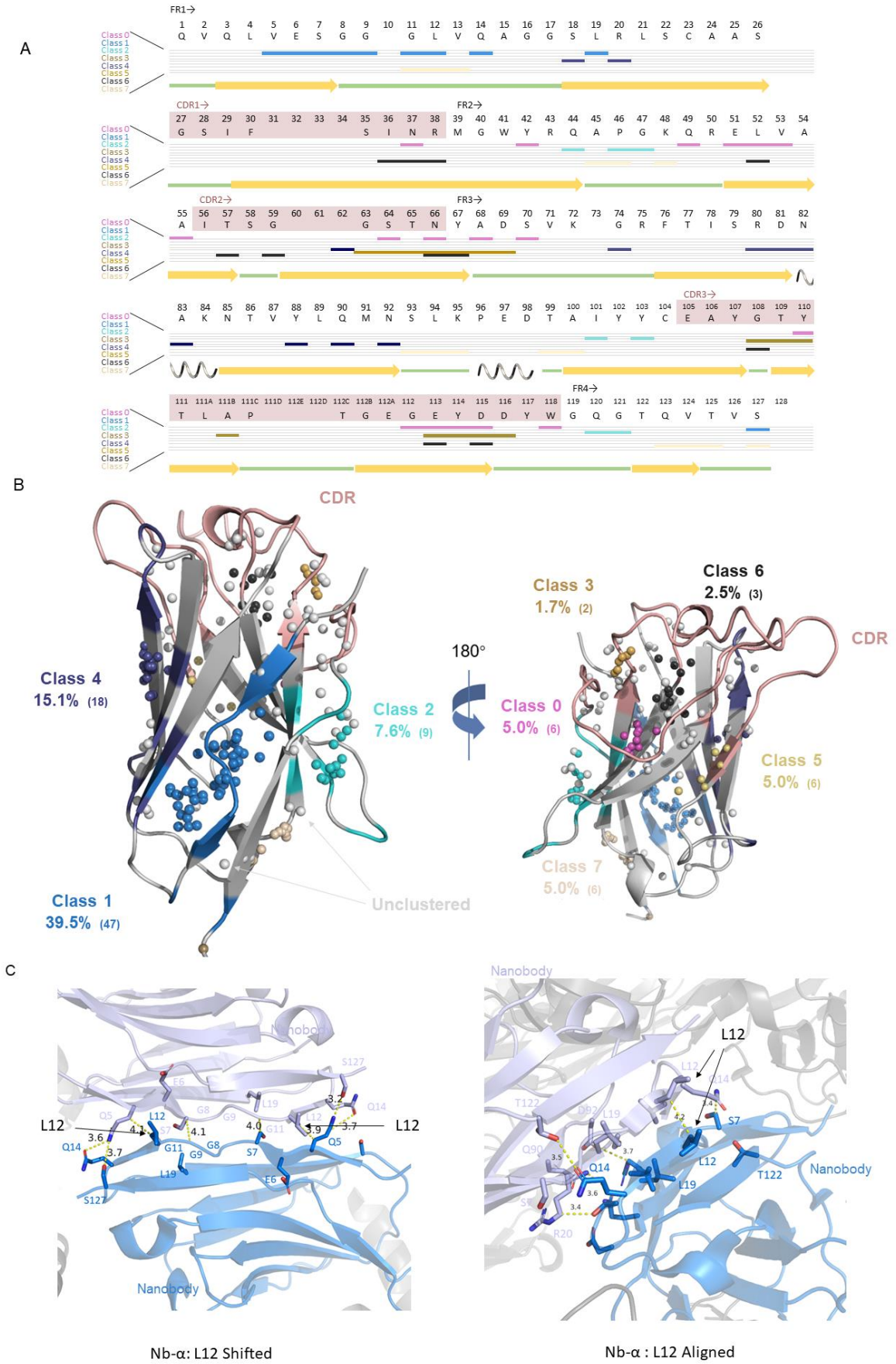


Figure 1 Nanobody-nanobody crystal contacts fall into limited classes. (*to be continued*)

Figure 1 (Continued) A) Standard nanobody residue numbering of RECQL5 nanobody Ctag:WT using IMGT scheme provided by ANARCI. CDRs are in light brown background and non-CDRs are in white background. Participating residues of each category are indicated by a bar below the one-letter code. B) DBSCAN clustering result: global view of major nanobody-nanobody crystal interfaces on a model drawn from chain E of 5O0W in PDB. Each colored sphere on the structure represents the average location of participating residues of an interface analyzed. Grey spheres represent ungrouped nanobody interfaces in the analysis. The three largest interface patterns are colored on the cartoon of the nanobody structure with respective class colors. Colours are consistent in A and B. C) Typical close views of Class 1 (Nb-α) Nanobody-nanobody crystal contacts. L12 Shifted type is shown in the left panel, and L12 Aligned type is shown in the right panel. Structure cartoons in marine and light blue represent two nanobody molecules participating in the interface. Residues represented as sticks are interacting residues of the current nanobody molecules. Residues represented as pink lines or sticks are interacting residues of the reference nanobody molecules. (See Supplementary Figure 4 and Supplementary Table 2 for detailed information of each class of nanobody-nanobody crystal contacts)

By using the Pymol package in Python 2.7, we were able to extract all interface information in the 335 crystal structures and focused on the interfaces that only involved nanobodies. The nanobody:nanobody interfaces were represented by index arrays of the residues involved. The index array was then renumbered by the IMGT scheme on the online server ANARCI[34] (Figure 1A), enabling comparisons across the nanobodies appearing in different deposited structures. The renumbered arrays were further converted into coordinate matrices, which then immediately shrank into a 4-dimensional array (n^*, μ_x, μ_y, μ_z) (Supplementary Figure 4). The four items in the array represent 1/3 of the number of residues in the interface, and the mean values of the X, Y, Z coordinates of the residues in the interface, respectively. Selected nanobody interfaces were then mapped onto a single nanobody structure (PDB id 5O0W) using their coordinate information (Supplementary Figure 4). As numeric representations of the interfaces, 4-dimensional vectors were used and we further performed density-based spatial clustering (DBSCAN) on the 356 vectors (Supplementary Figure 4).

The clustering result shows 8 classes of nanobody-nanobody crystal contacts, 4 of them away from the CDRs, accounting for 39.5%, 7.6%, 15.1% and 5.0% of the total (Figure 1B). The most prevalent 'Nb-α' (39.5%) is an edge-to-edge interaction of the two beta sheets at the N terminus (Figure 1C). The Nb-α typically has 10 residues interacting in this type of contact, creating a plethora of options for engineering an interface that will more likely crystallize.

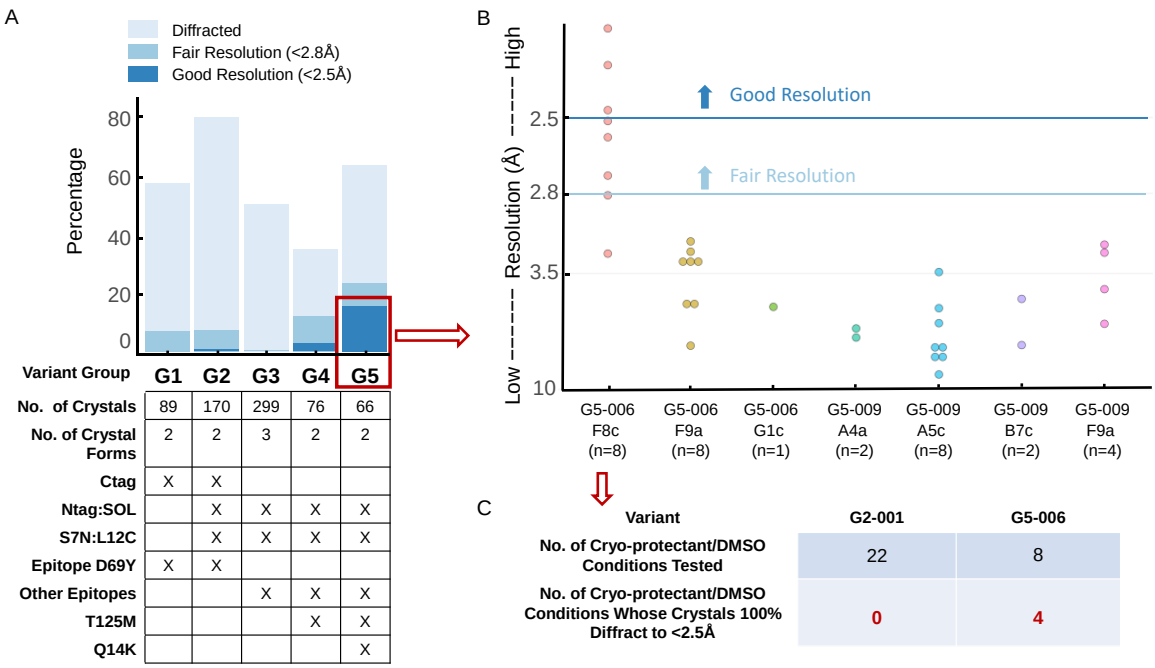
Iterations of nanobody scaffold mutagenesis yield consistently diffracting crystals

Based on the data mining result and within our RECQL5:nanobody system, we designed 5 generations of nanobody scaffolds (G1-G5) of more than 200 constructs with different combinations of Nb-α mutations and crystal epitope mutations (Supplementary Figure 2 and Figure 2A). The rationale of the mutational design around Nb-α was to introduce more interactions to enhance the interface, e.g. salt bridges, hydrogen bonds and potential disulfide bonds.

G1 and G2 were mutants based on the original RECQL5 nanobody scaffold with a C-terminal 6XHis tag, which was cleaved during the purification process and resulted in a 6-residue tail (ENLYFQ) at the C terminus. G2 crystals with the S7N:L12C mutations had higher diffraction rate, however the high-resolution fraction did not significantly increase. In G3, we introduced some solubility mutations (G40T:Q49E:L52W:I101V, abbreviated as SOL), changed the C-terminal 6XHis tag to the N-terminal MBP-6XHis tag which did not yield the 6-residue C-terminal tail, and meanwhile used other crystal epitope

mutations rather than D69Y. Although G3 crystals had an even worse diffraction quality, two more mutations (T125M and Q14K) based on G3 scaffolds resulted in significantly improved diffraction quality in G5.

We then further explored the crystals from the G5 category. There were several crystallization conditions yielding quick crystals within 2 days, which were selected. 10-12 crystals from each variant-condition pair were harvested for diffraction analysis. The pair G5-006-F8c had significantly larger fraction of crystals diffracting better than 2.5 Å (Figure 2B), and it went into the further optimization on cryo-protectant and DMSO concentrations (fragment soaking conditions). Out of the 8 conditions we tested, 100% crystals from 4 conditions diffracted better than 2.5 Å (Figure 2C and Supplementary Figure 8), which satisfied crystallographic fragment screening prerequisites. Comparing G5-006 with G2-001 (best variant in G2), diffraction quality of G5-006 crystals was significantly improved.



Variant Group	G1	G2	G3	G4	G5
No. of Crystals	89	170	299	76	66
No. of Crystal Forms	2	2	3	2	2
Ctag	X	X			
Ntag:SOL		X	X	X	X
S7N:L12C		X	X	X	X
Epitope D69Y	X	X			
Other Epitopes			X	X	X
T125M				X	X
Q14K					X

Variant	G2-001	G5-006
No. of Cryo-protectant/DMSO Conditions Tested	22	8
No. of Cryo-protectant/DMSO Conditions Whose Crystals 100% Diffract to <2.5Å	0	4

Figure 2 Consistently well-diffracting crystal emerge from several iterations of mutagenesis around Nb-α. A) Diffraction quality of different groups of variants measured by the percentage of crystals which diffract to high resolution. The numbers of crystals and mutations present in each group of variants are indicated in the table below each bar. B) Diffraction quality of different nanobody variant and condition pairs of RECQL5:nanobody variant complexes presented as a dotplot. C) A table summary of results of the fragment soaking condition test of G2-001 and G5-006. The test involves different concentrations of cryo-protectant and DMSO.

Constitutive nanobody interfaces were present in 6 different crystal forms

In the 5 generations of nanobody scaffolds, we aimed to find constitutive nanobody-nanobody interfaces which were independent of the target protein (RECQL5 in our case).

The interface we started with was Nb-α in G1. Upon combining S7N:L12C with G1, the crystallization propensity of RECQL5:nanobody complex increased dramatically for both N-terminal and C-terminal tagged variants (Figure 3A and 3B). Comparing the structures of G1-001 and G2-001, there was an additional hydrogen bond in Nb-α introduced by S7N, between the side chain N7 and the carbonyl oxygen of A15 (Figure 3C). Therefore, the S7N:L12C mutations were retained in the following iterations of mutagenesis.

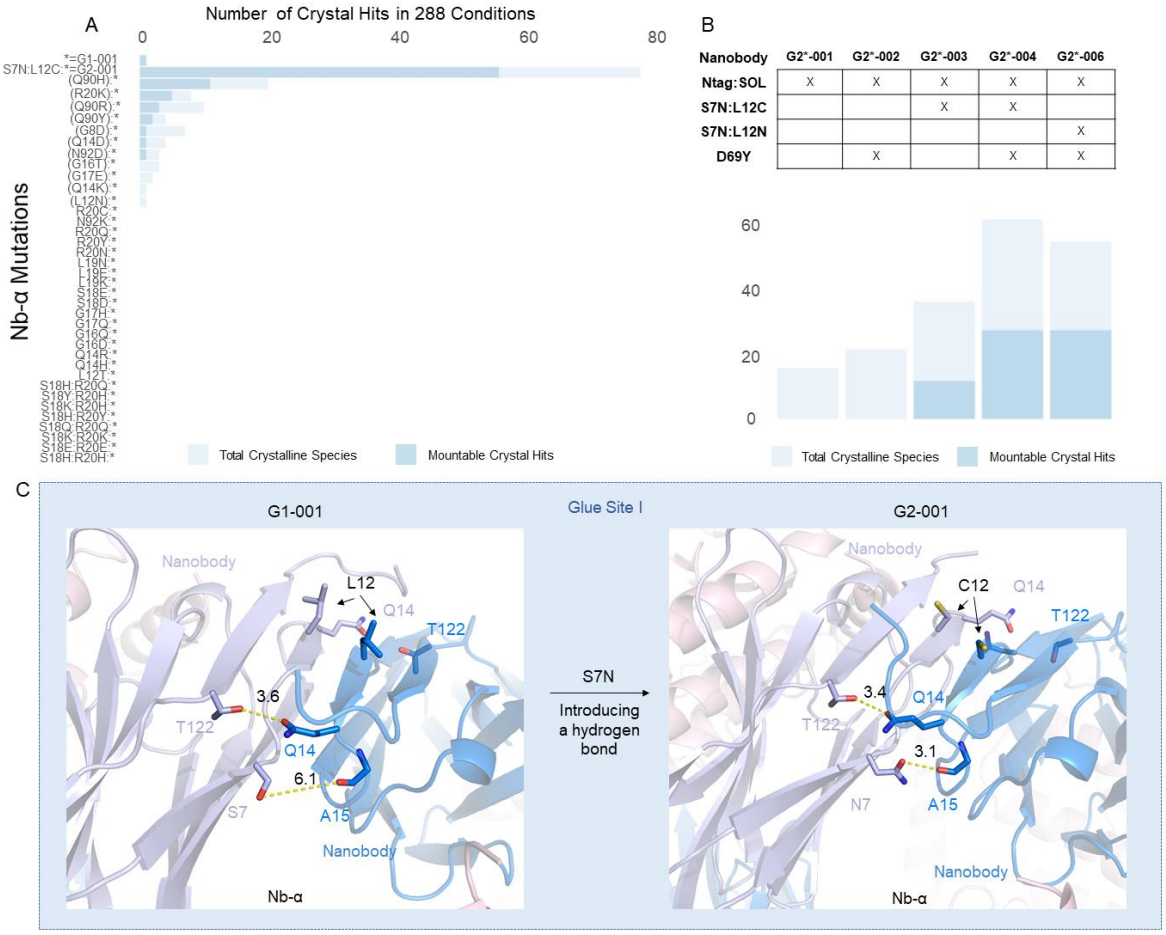


Figure 3 S7N enhances Nb-α by introducing an additional pair of hydrogen bonds. A) Hampton Index 3 (HIN3) triplicate drop coarse crystallization screen experiments show that S7N:L12C stands out among Nb-α mutations in significantly increasing the crystallization propensity of the RECQL5:nanobody complex. The asterisk represents the variant complex G1-001. B) HIN3 triplicate drop coarse crystallization screen experiments show that S7N:L12C significantly increases the crystallization propensity of the Ntag:SOL variant of RECQL5:nanobody complex with or without D69Y as the crystal epitope mutation. C) Close views of the Glue Site I before and after the mutation. S7N (stick representation) introduces an additional hydrogen bond with the carbonyl oxygen of A15 (stick representation) on the opposite nanobody molecule. Oxygen atoms are colored red and Nitrogen atoms are colored blue. The distances between the atoms connected by dash yellow lines are shown in Å.

Different combinations of crystal epitope mutations and Nb-α mutations were further explored. We replaced D69Y with 9 other crystal epitope mutations in G3, and in the meantime, changed the tag to the N-terminal MBP-6XHis tag. Surprisingly, the crystallization pattern changed dramatically with most G3 crystals in high salt conditions whilst most G2 crystals in PEG conditions (Figure 4A). It was noticed that Residue Cys12 was essential in the crystallization of G3, since introduction of the C12N mutation decreased the average number of observed hits significantly (Figure 4A). We further mutated the residue Cys12 to a selection of other residues, and found even good crystallizers would immediately stop crystallizing (Figure 4B). In addition, when Cys12 was chemically converted to dehydroalanine (Dha)[35] (Supplementary Figure 11), the good crystallizer G5-007 in complex with RECQL5 also gave no crystals.

Since Cys12 was demonstrated important in improving crystallization, mutagenesis of the neighbouring residues of Cys12 were further explored, especially Thr125. When T125M was introduced

in G4 and G5, again the crystallization pattern dramatically changed, expanding hits in both high salt and PEG conditions (Figure 4A). Furthermore, with the Nb-α mutation Q14K in G5, the crystallization propensity was further improved (Figure 4C) and the crystals grew large. However, due to the ambiguity of the electron densities captured from the K14 side chains, the rationale of this improvement remains unknown.

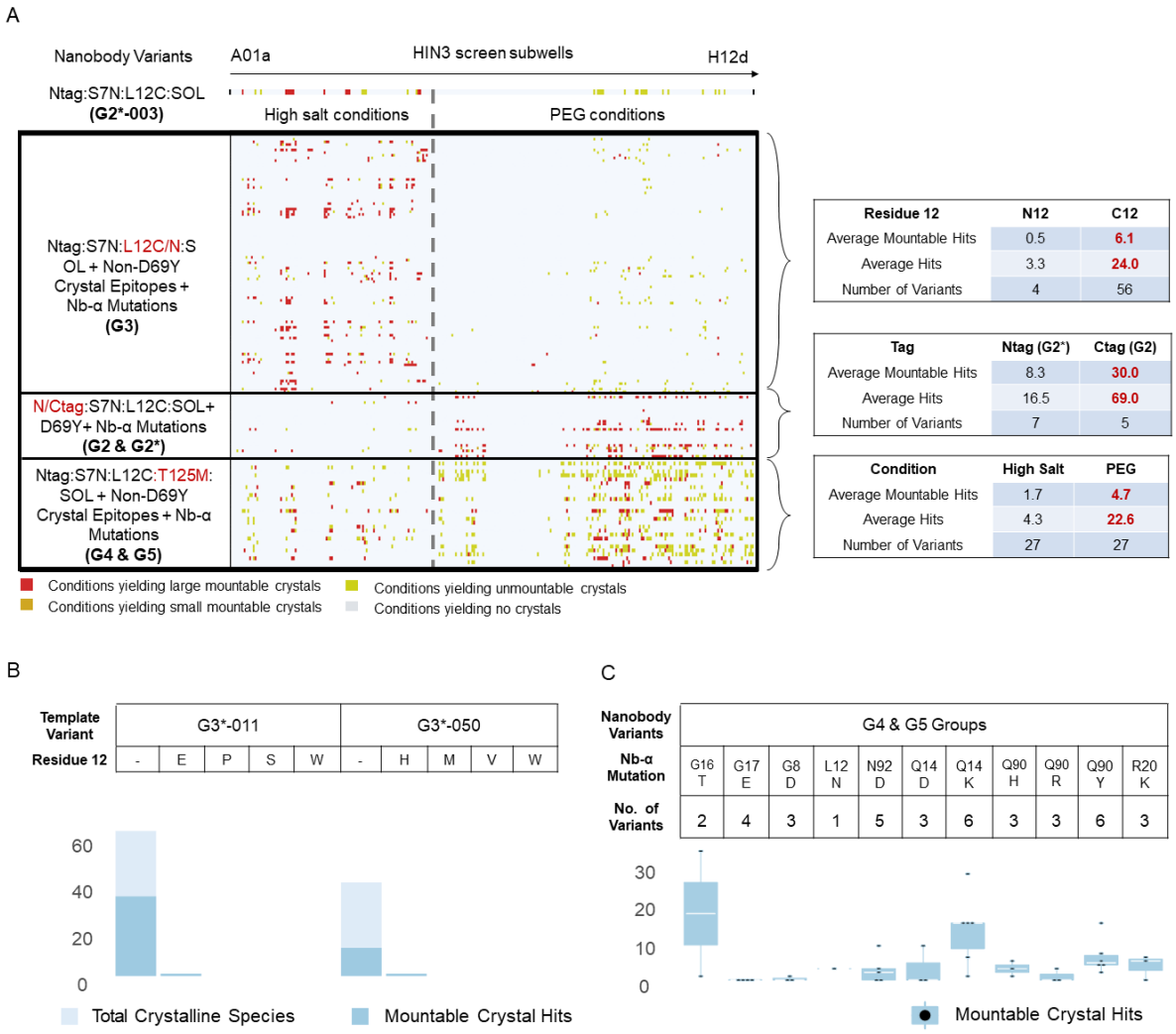


Figure 4 G2, G3 and G4&5 have different crystallization pattern in the Hampton Index coarse screen. A) Hampton Index 3 coarse screen triplicate crystal trials of RECQL5:nanobody complex with various mutations on the nanobody scaffold. Each row represents a nanobody construct, and each column in the chart represents a crystallization condition in the Hampton Index 3 coarse screen. The conditions are linearized sequentially from A01a on the left to H12d on the right. To the left of the grey dashed line are high salt conditions and to the right are PEG conditions. The color of each cell in the row represents whether there are crystals growing in that condition and the quality of the crystals. Some statistics of each group of nanobody variants are generalized on the right of the main plot. See Supplementary Table 5 for detailed information of what each row and column represents. B) The RECQL5 nanobody loses crystallization propensity when C12 is mutated to a selection of other residues. C) HIN3 triplicate drop coarse crystallization screen experiments show the effect of different Nb-α mutations to G4 &5 nanobody variants.

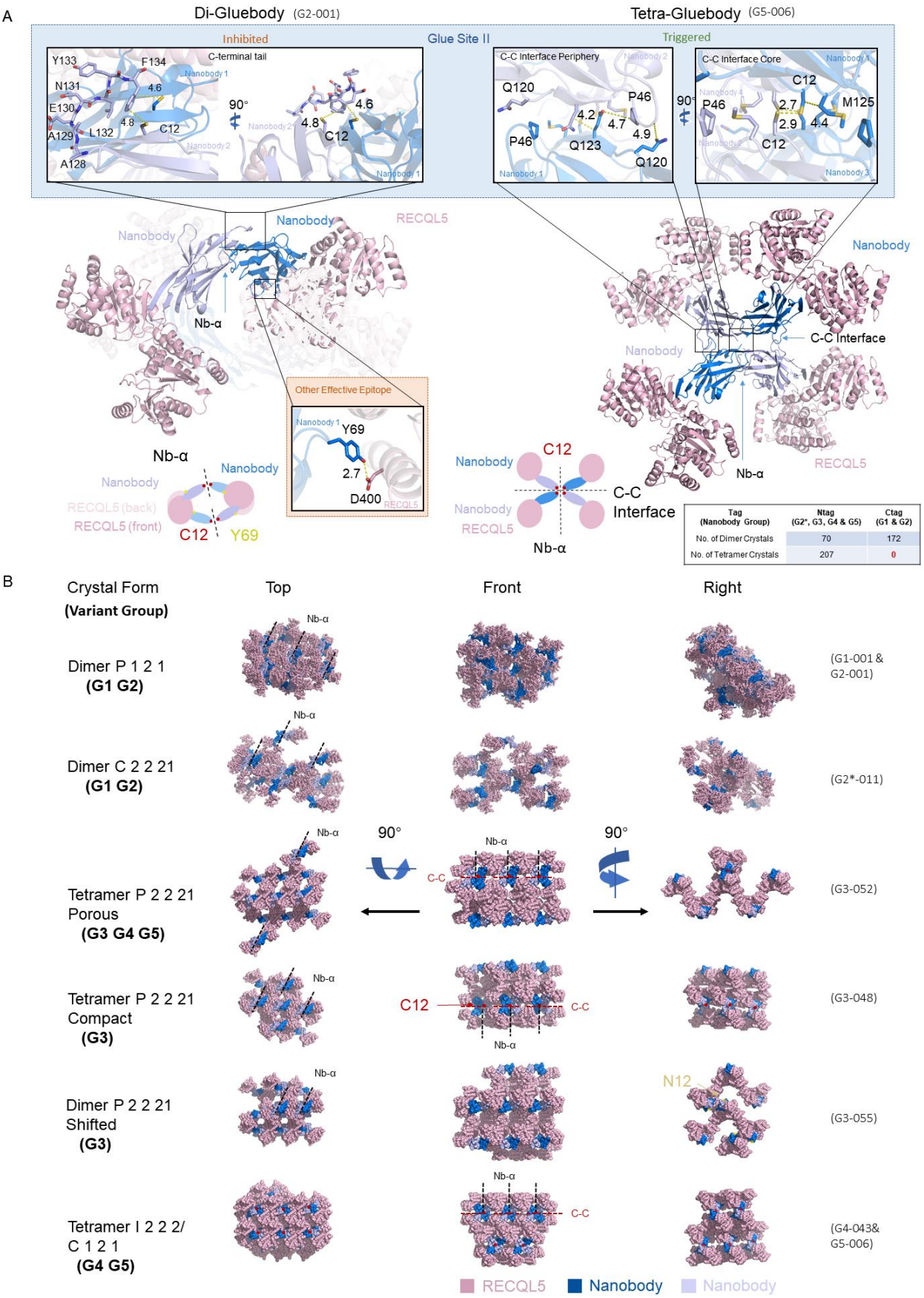


Figure 5 Two constitutive nanobody interfaces were present in 6 different crystal forms. (*to be continued*)

Figure 5 (Continued) A) Dimeric (Di-Gluebody) and tetrameric (Tetra-Gluebody) crystallization pattern of RECQL5:nanobody with close views of Glue Site I & II and the crystal epitope site of Y69. The left panel shows RECQL5:nanobody complex crystallizes in a dimeric pattern when mutation D69Y and the C-terminal tail exist on the nanobody scaffold. The cartoon structure with high transparency is a symmetry mate of the dimer. The schematic of the dimer pattern is shown on the bottom left. The right panel shows the tetrameric crystallization. The M125 has a longer side chain than the T125 and occupies the cavity around the cysteine core within the C-C interface. The schematic of the tetramer pattern is shown on the bottom right. Oxygen atoms are colored red and Nitrogen atoms are colored blue. The distances between the atoms connected by dash yellow lines are shown in Å. B) Six different crystal forms of the RECQL5:nanobody complex with respective mutations on the nanobody scaffold are presented in three perspectives (top, front and right views). RECQL5 molecules are in light-pink, and nanobody molecules are in marine and light-blue. C12 residues are presented as red dots on the structures and N12 residues are presented as yellow dots. Nb-α and C-C interfaces are indicated by dashed lines in the lattices. The crystal contacts of last four crystal forms are presented in Supplementary Figure 7.

The reason why there was a significant crystallization pattern shift from G2 to G3 might be that a totally different nanobody arrangement formed. Instead of a dimer nanobody core (Figure 5A left panel) mediated by Nb-α in G1 and G2, we observed a tetrameric nanobody core (Figure 5A right panel) in G3. In the tetrameric nanobody, the Cys12 residues were in the core of the interface in contact with each other, and thus we named the new interface perpendicular to Nb-α as 'C-C interface'. Due to the geometric arrangement of the four nanobody molecules, the tetrameric core would not form with the 6-residue C-terminal tail ENLYFQ, which we observed in the G2-001 structure hanging over the Cys12 residues (Figure 5A left panel). This could explain why there were no tetrameric nanobodies observed in C-terminal tagged variants (Figure 5A). The Cys12 itself was also essential in forming the tetrameric nanobody arrangement. When we mutated Cys12 to Asn12, the C-C interface dissociated and only the dimeric nanobody arrangement was observed (Figure 5B).

From G1-G3 we already observed 5 different crystal forms (Figure 5B). Dimer P 1 2 1 and Dimer C 2 2 21 existed mostly in PEG conditions. Tetramer P 2 2 21 Porous/Compact and Dimer P 2 2 21 shifted mostly existed in high salt condition. In G4 and G5 we observed crystals in both high salt and PEG conditions. The high salt crystals of G4&5 were Tetramer P 2 2 21 Porous form, which we already observed in G3. The reason why G4&5 crystallized extensively in PEG conditions might be that it unlocked an additional crystal form Tetramer I 2 2 2/C1 2 1, and this crystal form led to our best-diffracting data (Supplementary Figure 5), indicating that the crystal packing was also improved by the mutation T125M.

Among all 6 crystal forms, the Nb-α was present, and among all tetramer crystal forms, the C-C interface was present (Figure 5B). This indicated that these two interfaces were favoured. We therefore termed the nanobody scaffolds with our modifications to enhance the two interfaces as 'Gluebodies'. We also termed interacting residues in Nb-α as 'Glue site I' (Figure 3C and 5A) and interacting residues in the C-C interface as 'Glue site II' (Figure 5A).

Gluebodies can improve crystallizing proteins but do not rescue non-crystallizing proteins

The sets of mutations around the Nb-α on the Gluebody are demonstrated to improve the crystallization propensity of the nanobody with the model protein RECQL5, by enabling various crystal forms via the dimeric or tetrameric Gluebody arrangements. However, the improvement on the nanobody alone is not in our hands sufficient to generate crystals for non-crystallizing proteins. This was attempted on 5 other proteins (Figure 6 and Supplementary Figure 6) yielding crystals for only two, both

weakly diffracting (TUT4 and PRNP), work still ongoing. For a Major Facilitator Superfamily (MFS) protein, the crystals of the Gluebody complex did not diffract as well as the original nanobody scaffold. In such case, the protein itself will likely need improving, or in the case of Lipid Cubic Phase (LCP), multiple combinations of the kinds of detergents and LCP lipids need exploring as well.

Target	Number of Glubodies Tried	Number of Crystal Plates set up with Gluebodies	Lipid Cubic Phase?	Crystals with Gluebody?	Crystals with original nanobody scaffold?	Diffraction	Comment
TUT4	5	20	No	Yes	No	6-8Å	
WRN	8	18	No	Yes	No	2Å	Unsolvable dataset
PRNP	1	2	No	Yes	No	6-8Å	
MFS protein	2	12	Yes	Yes	Yes	Do not diffract	Crystals with original nanobody scaffold diffract to 7-10Å
MAGEC2	3	6	No	No	No	/	

Figure 6 Experimental details of the crystallization trials of several targets in complex with respective Gluebodies.

Discussion

Gluebodies are a route to new crystal polymorphs

In this study, we identified multiple crystal polymorphs in the RECQL5:nanobody model system, just by modifying the nanobody scaffold. The key to uncovering more crystal forms might be triggering the tetrameric nanobody pattern, which is the rationale behind the 'Gluebody'.

The tetrameric nanobody pattern is more difficult to form with the naïve nanobody scaffold. Until recently known, there are only two cases of a tetramer crystal form in PDB (PDB id 6OS1 and 6OS2). The nanobody pattern of 6OS1 and 6OS2 is similar to the tetramer crystal form in our study with Nb-α and the C-C interface perpendicular to each other, but with fewer interacting residues.

Gluebodies can form the tetrameric pattern because they do not have C terminal tails, which is the prerequisite of forming the tetrameric nanobody pattern. However, C-terminal tags are more frequently used in nanobodies than their N-terminal counterparts. In the nanobody interface analysis of 335 crystal structures containing nanobodies, 211 of them (63%) have nanobodies with more than 5 residues as C-terminal tails. In our analysis of the dimer and tetramer crystal forms, C-terminal tails were a strong inhibitor of the formation of the tetramer crystal form, as they hung over the C12 surface, preventing further crystal contacts forming, including the C-C interface.

The four mutations are truly refinement of intra-nanobody interfaces. In our study we confirmed that the four key mutations of the Gluebody would be helpful in forming the tetramer crystal form more easily and because those mutated residues only contacted nanobody residues, they would not depend on widely ranging properties of the target proteins.

Multiple Gluebody variants are needed to fully explore the chaperoning ability of nanobodies

In this work, it is demonstrated that different crystal polymorphs could be obtained just by modifying the nanobody scaffold, and that the diffraction characteristics of different crystal polymorphs are different (Supplementary Figure 5). To obtain well-diffracting crystals, it is imperative to discover crystallization conditions that yield a suitable crystal polymorph. With different Gluebody scaffolds, new crystal polymorphs can be obtained, and therefore this allows us to suggest 3 strategies:

1) C-terminal tailed naïve nanobody. There is no reason why the original scaffold should be abandoned completely. When the original nanobody is at hand, using it straight away should also provide some chance for successful co-crystallization of the target protein.
2) Di-Gluebody (Figure 7). By introducing Glue Site I mutations (Figure 3C, 5A and 7), (i.e. adding the mutation S7N), in addition to truncating the C-terminal tail, Di-Gluebodies are made. The Nb-α is enhanced in this case, but the C-C interfaces are not necessarily aligned with the Cys residues. There is a case in the PDB (id 6QGY) where two dimer nanobodies pack back-to-back in an angle with an interface perpendicular to Nb-α (Supplementary Figure 9), indicating there is sometimes no need for an anchored C-C interface to assist crystallization.
3) Tetra-Gluebody (Figure 7). By introducing Q14K and Glue Site II mutations (Figure 5A and 7) in addition to Di-Gluebody, Tetra-Gluebodies are made. The 4 key mutations are installed to try to trigger the tetrameric crystallization polymorphs with an anchored C-C interface (Figure 5B).

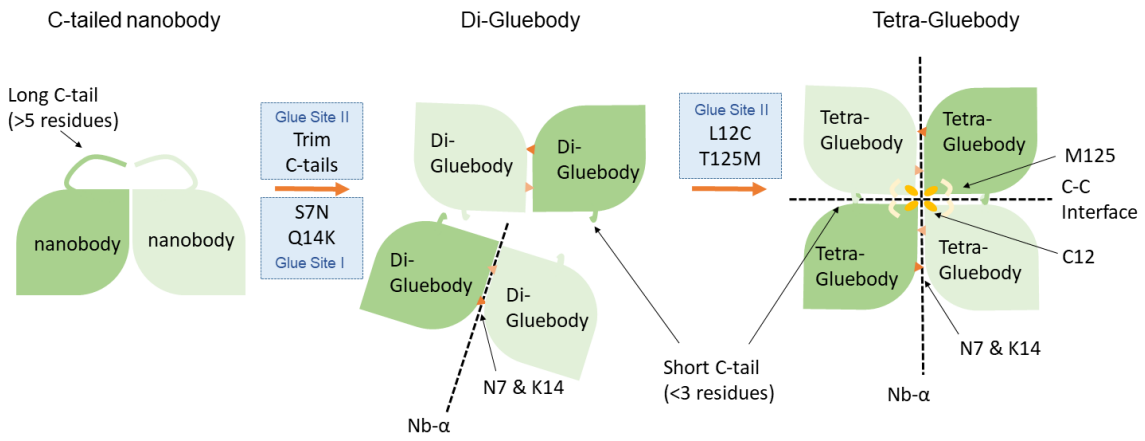


Figure 7 Key modifications on the nanobody scaffold in the evolution of the Gluebody. The putative crystallization patterns from the three scaffolds are presented from left to right with the scaffold type at the top of the sketches. The leaf shapes colored in green and light green represent nanobody molecules. The long or short overhangs on the leaf shapes represent the C-terminal tails of the nanobody molecules. The shapes colored in yellow, orange and pale represent the key mutations on the nanobody scaffold. (C12, N7 & K14 and M125, respectively) The dashed lines represent nanobody-nanobody interfaces occurring in the putative crystallization patterns.

The tetrameric nanobody assembly can be applied in both crystallography and cryo-EM

In this work, it is demonstrated that the Gluebody has facilitated good crystal packing for RECQL5. Furthermore, there is also good chance that some adaptors could make the Gluebody a homotetrameric protein complex in solution. If the tetra-Gluebody complex is rigid enough, it could be used as an imaging scaffold in cryo-EM to visualize small soluble proteins more easily with a D2 symmetry.

There are a few imaging scaffolds demonstrated useful for the cryo-EM structure solution of small proteins, such as the DARPin cage as a multivalent imaging scaffold[36], and the Megabody[37] and the Legobody[38] as monovalent imaging scaffolds. The advantage for the latter is that they are designed base on the nanobody scaffold, which is more frequently used than DARPins. However, the advantage for the DARPin cage is that its symmetry is more useful when further map refinement of the target protein is necessary. The tetra-Gluebody complex might be a tool which has both advantages and become a popular tool.

To make the tetra-Gluebody a rigid complex in solution, the chemistry of its interfaces might be interesting. In the tetrameric crystallization pattern, we observed 4 Cysteine residues in the C-C interface core. There was clearly electron density between the two sulfur atoms in each pair of Cys residues (Supplementary Figure 10), and the distances between two sulfur atoms were around 2.8 Å. This was longer than a disulfide bond, which is around 2.05 Å[39]. We have no evidence supporting there exist chemical bonds among the four cysteines, and chances are there is still unoccupied space in the tetrameric nanobody core. Therefore, further metal coordination or organic adaptor molecules could be explored to make the tetrameric nanobody pattern rigid in solution.

Methods

Crystal contact analysis of structures containing nanobodies in PDB

Crystal structures were fetched and analyzed using the open-source PYMOL package in Python[40]. The workflow of analysis is described in results and shown in detail in Supplementary Figure 4. Finally, a few interfaces that fell into the noise category were manually inspected and added to respective categories resulting in the final classification.

High-throughput cloning, expression and purification of nanobody variants

The nanobody constructs of G0-002 to G0-074 and G1-001 were generated by site-directed mutagenesis method based on G0-001 using pNIC-CTH0 as the vector plasmid. The nanobody construct of G2-001 was generated by mutagenesis method based on G1-001 using pNIC-CTH0 as the vector plasmid. The nanobody constructs of G1-002 to G1-091 were generated by mutagenesis method based on G1-001 using pNIC-CTBH as the vector plasmid. The nanobody constructs of G2*-005 to G2*-016 were generated by mutagenesis method based on G2*-004 using pNIC-MBP2-LIC as the vector plasmid and G2-002 to G2-013 using pNIC-CTBH as the vector plasmid with the same method and template DNA. G3*-001 to G3*-019, G3*-020 to G3*-038, G3*-039 to G3*-057, G3*-058 to G3*-076 and G3*-077 to G3*-095 were generated using mutagenesis based on templates G3-048, G3-055, G3-052, G3*-011 and G3*-050 respectively. G3-001 to G3-096 and G2*-001 to G2*-004 were directly synthesized as DNA fragments and subcloned into pNIC-MBP2-LIC vectors. G4-001 to G4-087 and G5-001 to G5-009 were also based on the synthesized DNA fragments using a different set of primers in the PCR step. All cloning procedures followed ligation independent cloning described previously[10,41].

Constructs of nanobody variants were transformed into Single step (KRX) E.coli strain (Promega) at end of molecular cloning. Single colony was inoculated into 1 ml of LB media (Merck) and incubated overnight. Starter culture of each nanobody was then inoculated into 100ml autoinduction media (Formedia) containing 0.1% Rhamnose, 0.01% antifoam 204 and antibiotics, at 37°C for 5.5 hours followed by 40-44 hours at 18°C. The base buffer we used for purification contained 5% glycerol, 10 mM HEPES (pH 7.5), 500 mM NaCl, 0.5 mM TCEP. Bacteria for each nanobody were harvested at 4000g and resuspended in 10ml base buffer + 30 mM imidazole, 1% Triton, 0.5 mg/ml lysozyme, 10 ug/ml homemade benzonase, followed by storage in -80°C freezer overnight for complete cell lysis. The purification started next day with thawing the frozen pallets in room temperature water bath, followed by centrifugation at 5000g for 1 h to obtain clear supernatant. The supernatant for each nanobody was then applied to one 1 ml Ni-NTA pre-packed column (GE healthcare) pre-equilibrated with base buffer + 30 mM imidazole. After thoroughly washing the Ni-NTA columns with base buffer, nanobody was eluted using 2.5 ml of base buffer + 500 mM imidazole directly into base buffer equilibrated PD-10 columns (GE healthcare). 3.5 ml of base buffer was then applied to PD10 columns to elute the nanobody protein in 0 imidazole buffer. Then TEV protease was added to protein solution with a 1:10 mass ratio for overnight incubation, and 10 mM imidazole was also added in the solution. In the next day, two Ni-NTA columns were pre-equilibrated for each nanobody with base buffer + 10 mM imidazole and the nanobody solution with TEV was applied to the columns to get rid of nanobody with uncleaved tags, TEV protease and contaminants. Flow-through fractions were collected and 2 ml of base buffer was applied to the Ni-NTA columns to flush all nanobody through. The collected fractions were ready for complex preparation.

Expression and Purification of RECQL5

The truncated RECQL5 protein (11-453) was subcloned into pNIC-Bsa4 vector with TEV cleavable 6xHIS tag at the N terminus. The protein was expressed using BL21-DE3-pRARE strain in auto-induction TB media (Formedia) containing antibiotics and 0.01% antifoam 204 at 37°C for 5.5 hours followed by 40-44 hours at 18°C. The base buffer we used for purification contained 5% glycerol, 10 mM HEPES (pH 7.5), 500 mM NaCl, 0.5mM TCEP. Bacteria were harvested by centrifugation at 4000Xg and re-suspended in 3 times volume of base buffer + 30 mM Imidazole, 1% Triton, 0.5 mg/ml lysozyme, 10 ug/ml homemade benzonase, followed by storage in -80°C freezer overnight for complete cell lysis. The purification started next day with thawing the frozen pallets in room temperature water bath, followed by centrifugation at 5000 g for 1h to obtain clear supernatant. The supernatant was then applied to Ni-NTA pre-packed columns (GE healthcare) pre-equilibrated with base buffer + 30 mM imidazole. After thoroughly washing the Ni-NTA columns with base buffer, protein was eluted using 2.5 ml of base buffer + 500 mM imidazole directly into base buffer equilibrated PD-10 columns (GE healthcare). 3.5ml of base buffer was then applied to PD10 columns to elute the RECQL5 protein in base buffer. Then TEV protease was added to protein solution with a 1:10 mass ratio for overnight incubation, 20 mM imidazole was also added in the solution. In the next day, twice quantity of Ni-NTA columns were pre-equilibrated with base buffer + 20 mM imidazole and the RECQL5 solution with TEV was applied to the columns to get rid of RECQL5 with uncleaved 6xHIS tag, TEV protease and contaminants. Flow-through fractions were collected, and flash frozen for making nanobody complexes.

Crystallization of RECQL5:nanobody complex variants and X-ray diffraction

Purified RECQL5 and nanobody were mixed at a ratio of 1:1.5 and concentrated using 10kDa concentrator (Amicon and Corning), followed by Sepax SRT SEC-300 HPLC or Superdex-200 SEC. The peak fractions containing both RECQL5 and nanobody were pooled and concentrated to certain extent (supplementary table 3) for crystallization trials. Crystal plates were set up by Mosquito using Swiss-CI 3 drop plates with precipitants from Hampton Index screen (HIN3 HT-96, Molecular dimensions). Plates were sealed and sent into Formulatrix imager and imaged at 12 hours, 1 day, 4 days, 7 days, 14 days, 28 days and 56 days. Crystals were harvested using Shifter[42] with 17% glycerol for high salt conditions or ethylene glycol for PEG conditions added as cryo-protectant, then snap frozen and shipped to multiple beamlines at Diamond Light Source for X-ray diffraction.

Structure determination of RECQL5:nanobody complex variants

Diffraction data were processed by multiple auto-processing (including Xia2 and Autoproc) pipelines on ISPYB[43-45]. The high-resolution cutoff was determined by CC ½ above 0.3. The resolution values used in the analysis were the results of the Xia2 processing pipeline. Datasets were truncated using Aimless in CCP4-i suite[46]. Phasing of RECQL5 was done by Phaser[47], using the D1 and D2 domains of RECQCL5(PDB id 5LB8) as search models. The search model for phasing the nanobody was chain A of 2X1O. Further refinement was done by COOT and Refmac5 in CCP4-i suite[48,49].

Data availability

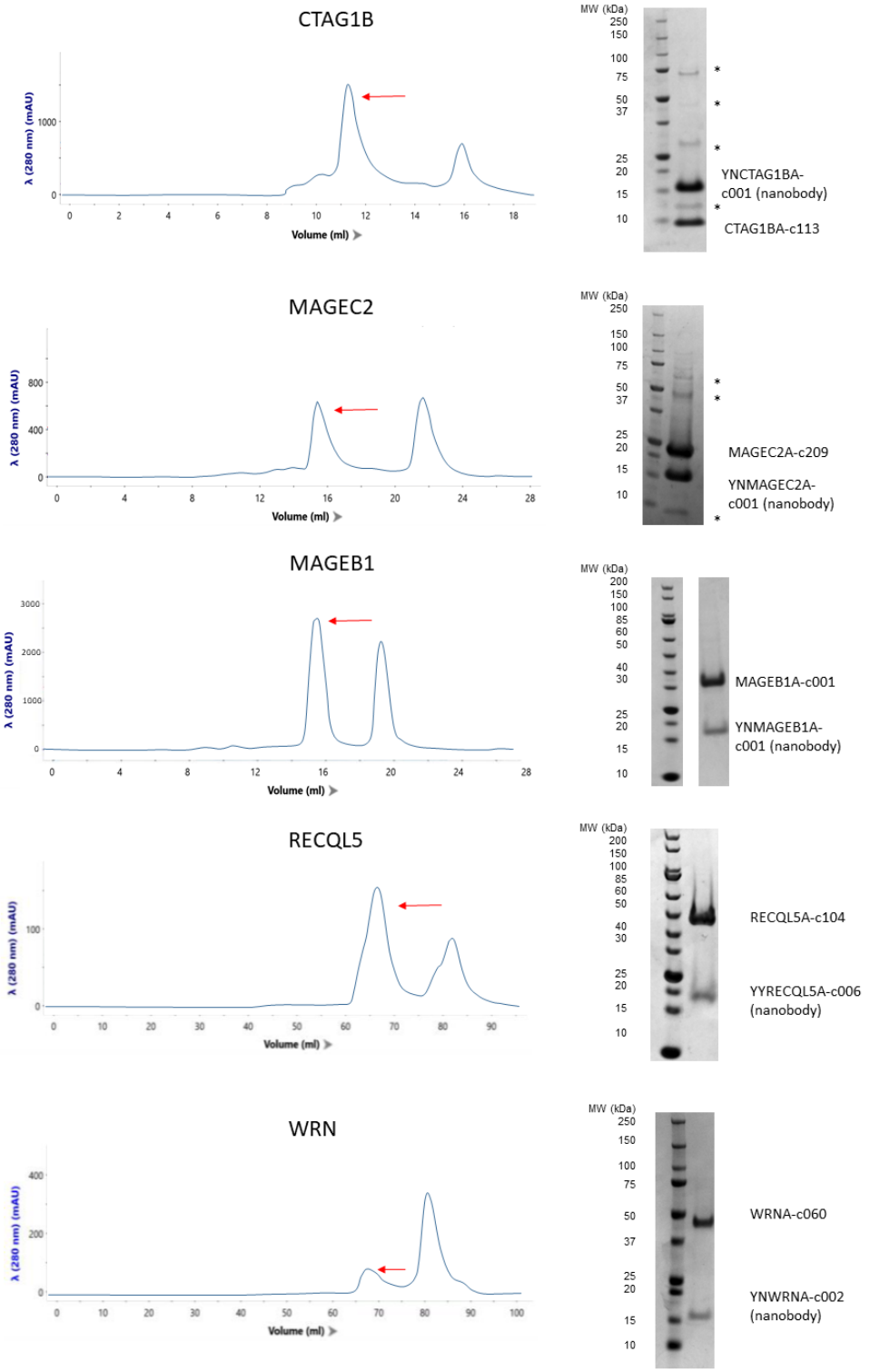
The structures of RECQL5 in complex with engineered nanobodies are deposited in the Protein Data Bank (PDB), with PDB IDs listed as follows.

Nanobody involved	PDB ID
G1-001	7ZML
G2-001	7ZMM
G3-048	7ZMN
G3-052	7ZMO
G3-055	7ZMP
G2*-006	7ZMQ
G2*-011	7ZMR
G4-043	7ZMS
G5-006	7ZMT
G5-006	7ZMV

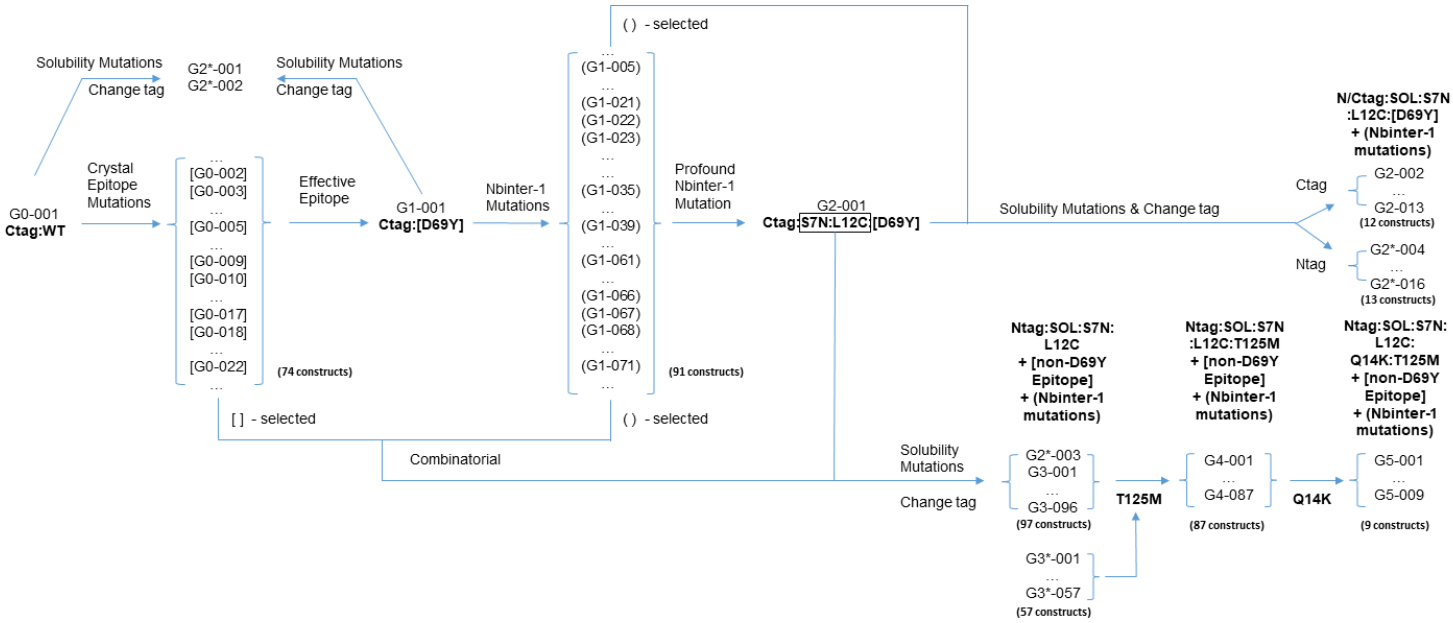
The PDB IDs of structures used in the nanobody interface analysis are listed as follows - 2X1O, 2X1P, 2X1Q, 2X6M, 2X89, 2XT1, 2XV6, 2XXC, 2XXM, 3CFI, 3DWT, 3EAK, 3EBA, 3EZJ, 3G9A, 3K1K, 3K74, 3OGO, 3P0G, 3SN6, 3STB, 3ZKQ, 3ZKS, 4AQ1, 4BEL, 4BFB, 4C57, 4C58, 4C59, 4CDG, 4EIG, 4EIZ, 4EJ1, 4FHB, 4GFT, 4GRW, 4HEM, 4HEP, 4I0C, 4I13, 4I1N, 4IOS, 4JVP, 4KML, 4KRL, 4KRM, 4KRN, 4KRO, 4KRP, 4LDE, 4LDL, 4LDO, 4LGP, 4LGR, 4LGS, 4LHJ, 4LHQ, 4MQS, 4MQT, 4N9O, 4OCL, 4OCM, 4OCN, 4P2C, 4QGY, 4QKX, 4QLR, 4QO1, 4S10, 4S11, 4W6W, 4W6X, 4W6Y, 4WEM, 4WEN, 4WEU, 4WGV, 4WGW, 4X7C, 4X7D, 4X7E, 4X7F, 4XT1, 4Y8D, 4Z9K, 4ZG1, 5BOP, 5BOZ, 5C1M, 5C2U, 5C3L, 5DA0, 5DA4, 5DFZ, 5DXW, 5000, 5E0Q, 5E5M, 5E7B, 5E7F, 5F1K, 5F1O, 5F21, 5F7K, 5F7L, 5F7M, 5F7N, 5F7W, 5F7Y, 5F8Q, 5F8R, 5F93, 5F97, 5F9A, 5F9D, 5FWO, 5G5R, 5G5X, 5GXB, 5H8D, 5H8O, 5HDO, 5HGG, 5HVF, 5HVG, 5HVH, 5IMK, 5IML, 5IMM, 5IMO, 5IP4, 5IVN, 5IVO, 5JA8, 5JA9, 5JDS, 5JQH, 5LHN, 5LHP, 5LHQ, 5LHR, 5LMJ, 5LMW, 5LZ0, 5M13, 5M14, 5M15, 5M2M, 5M2W, 5M30, 5M7Q, 5M94, 5M95, 5MJE, 5MP2, 5MWN, 5MY6, 5MZV, 5NBD, 5NBL, 5NBM, 5NLU, 5NLW, 5NM0, 5NML, 5NQW, 5O02, 5O03, 5O04, 5O05, 5O0W, 5O2U, 5O8F, 5OCL, 5OJM, 5OMM, 5OMN, 5OVW, 5TD8, 5TJW, 5TOJ, 5TOK, 5TP3, 5U64, 5U65, 5UK4, 5UKB, 5USF, 5VAK, 5VAN, 5VL2, 5VLV, 5VM0, 5VM4, 5VM6, 5VNV, 5VNW, 5WB1, 5WB2, 5Y7Z, 5Y80, 6B20, 6B73, 6C5W, 6C9W, 6DBA, 6DBD, 6DBE, 6DBF, 6DBG, 6EHG, 6EQI, 6EY0, 6EY6, 6EZW, 6F0D, 6FE4, 6FUZ, 6FV0, 6GZP, 6H02, 6H15, 6H16, 6H1F, 6H6Y, 6H6Z, 6H70, 6H71, 6H72, 6H7J, 6H7L, 6H7M, 6H7N, 6H7O, 6I6J, 6IBL, 6MXT, 6QTL, 6R7T, 6O3C, 6OS0, 6OS1, 6OS2, 6OYH, 6OYZ, 6OZ6, 6Q6Z, 6QD6, 6QGW, 6QGX, 6QGY, 6QPG, 6QUZ, 6QV0, 6QV1, 6QV2, 6QX4, 6RNK, 6RTW, 6RTY, 6RU3, 6RU5, 6RUL, 6RUM, 6RUV, 6RVC, 6S0Y, 6SGE, 6SSI, 6SSP, 6TEJ, 6TYL, 6SWR, 6U12, 6U14, 6U50, 6U51, 6U52, 6U53, 6U54, 6U55, 6VI4, 6WAQ, 6WAR, 4DK3, 4DK6, 4DKA, 4KDT, 4KSD, 4PIR, 5J1S, 5J1T, 5M2I, 5M2J, 6GJQ, 6GJS, 6GJU, 6GK4, 6GKD, 6GWN, 6GWP, 6GWQ, 6HD8, 6HD9, 6HDA, 6HDB, 6HDC, 6HEQ, 6HER, 6HHD, 6HHU, 6HJX, 6HJY, 6I2G, 6I8G, 6I8H, 6IBB, 6ITP, 6ITQ, 6JB2, 6JB5, 6JB8, 6JB9, 6DO1, 6DYX, 6F2G, 6F2W, 6FPV, 6GCI, 6GS1, 6GS4, 6GS7, 6N4Y, 6N50, 6NFJ.

Supplementary figures

Supplementary Figure 1 The size exlucsion chromatography profiles (left) and SDS-PAGE images (right) for each target protein. The red arrows show the protein peak fractions collected for crystallization trials. '*' indicates contaminants in the protein samples.



Supplementary Figure 2 The road map of the several iterations of mutagenesis on RECQL5 nanobody scaffold. Variants in '[]' are selected crystal epitope mutants and in '()' are selected Nb-α mutants for subsequent iterations of mutagenesis.



Supplementary Figure 3 The crystal epitope mutation D69Y on the scaffold rescues the RECQL5 nanobody to be effective again as a crystal chaperone. A) The combinatorial mutation designs. 22 single site mutations were chosen from the list of mutations provided by the crystal epitope server. 11 of the single site mutations were then selected to generate double-site and triple-site mutations thereafter, which resulted in 74 nanobody variants for crystallization trials. B) The information of the crystal epitope variants that yielded crystals in the Hampton Index 3 Screen. All those effective nanobody variants shared the D69Y mutation, highlighted in red. C) The G1-001 crystal. D) G0-057 crystals.

A

Single Site (22 constructs)

	Nanobody Mutation ID	Protein ID
	Ctag:[D69Y]	G1-001
Site 1	Ctag:[A50F:A51T]	G0-002
	Ctag:[E6Q]	G0-003
	Ctag:[L12W]	G0-004
Site 2	Ctag:[T86A:V87L:Y88S]	G0-005
	Ctag:[T78S]	G0-006
	Ctag:[K84S]	G0-007
Site 3	Ctag:[Q90E]	G0-008
	Ctag:[Y88H]	G0-009
	Ctag:[N82S:N85G]	G0-010
	Ctag:[K95E]	G0-011
	Ctag:[N85G]	G0-012
	Ctag:[N82S]	G0-013
	Ctag:[N85S]	G0-014
	Ctag:[N82K:N85G]	G0-015
	Ctag:[N82G:N85G]	G0-016
	Ctag:[S79Q]	G0-017
	Ctag:[R20K:L21V]	G0-018
	Ctag:[A15P:G17D]	G0-019
	Ctag:[A15P:G17A]	G0-020
	Ctag:[A45V]	G0-021
	Ctag:[Q44R]	G0-022

Double Site (31 constructs)

Site 1 X Site 2 → 3 (G0-023 – 25)

Site 1 X Site 3 → 7 (G0-026 – 32)

Site 2 X Site 3 → 21 (G0-033 – 53)

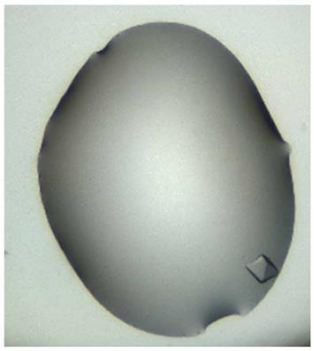
Triple Site (21 constructs)

Site 1 X Site 2 X Site 3 → 21 (G0-054 – 74)

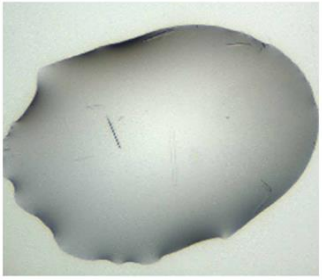
B

Nanobody mutation ID	Protein ID	Concentration (mg/ml)	No. of Drops with Mountable crystals	No. of Drops with Crystalline species
Ctag:[D69Y]	G1-001	11.46	1	1
Ctag:L12W:[D69Y:K84S]	G0-054	10.43	6	6
Ctag:L12W:[D69Y:Y88H]	G0-056	10.15	2	19
Ctag:L12W:[D69Y:N82S:N85G]	G0-057	14.05	4	10
Ctag:L12W:[D69Y:N82G:N85G]	G0-060	16.70	0	4

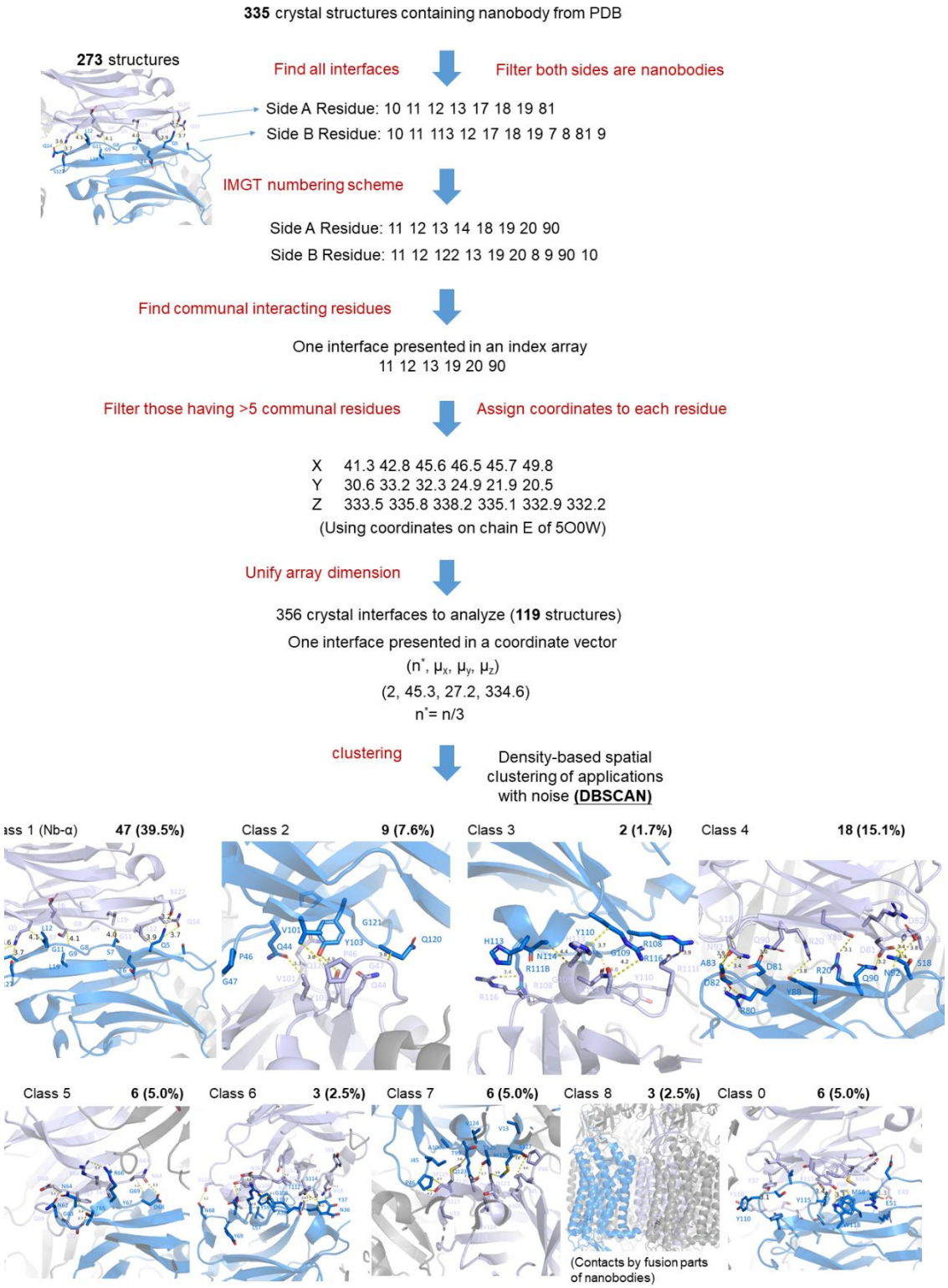
C



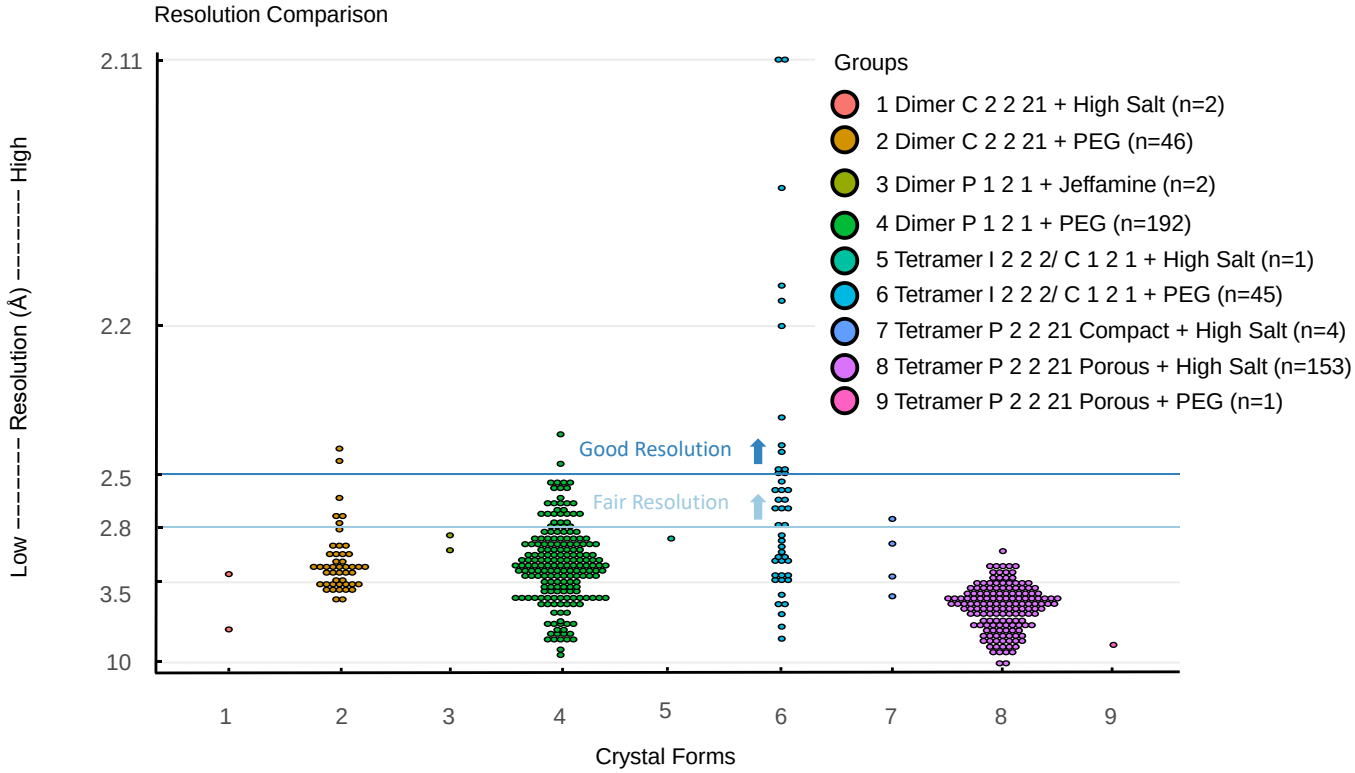
D



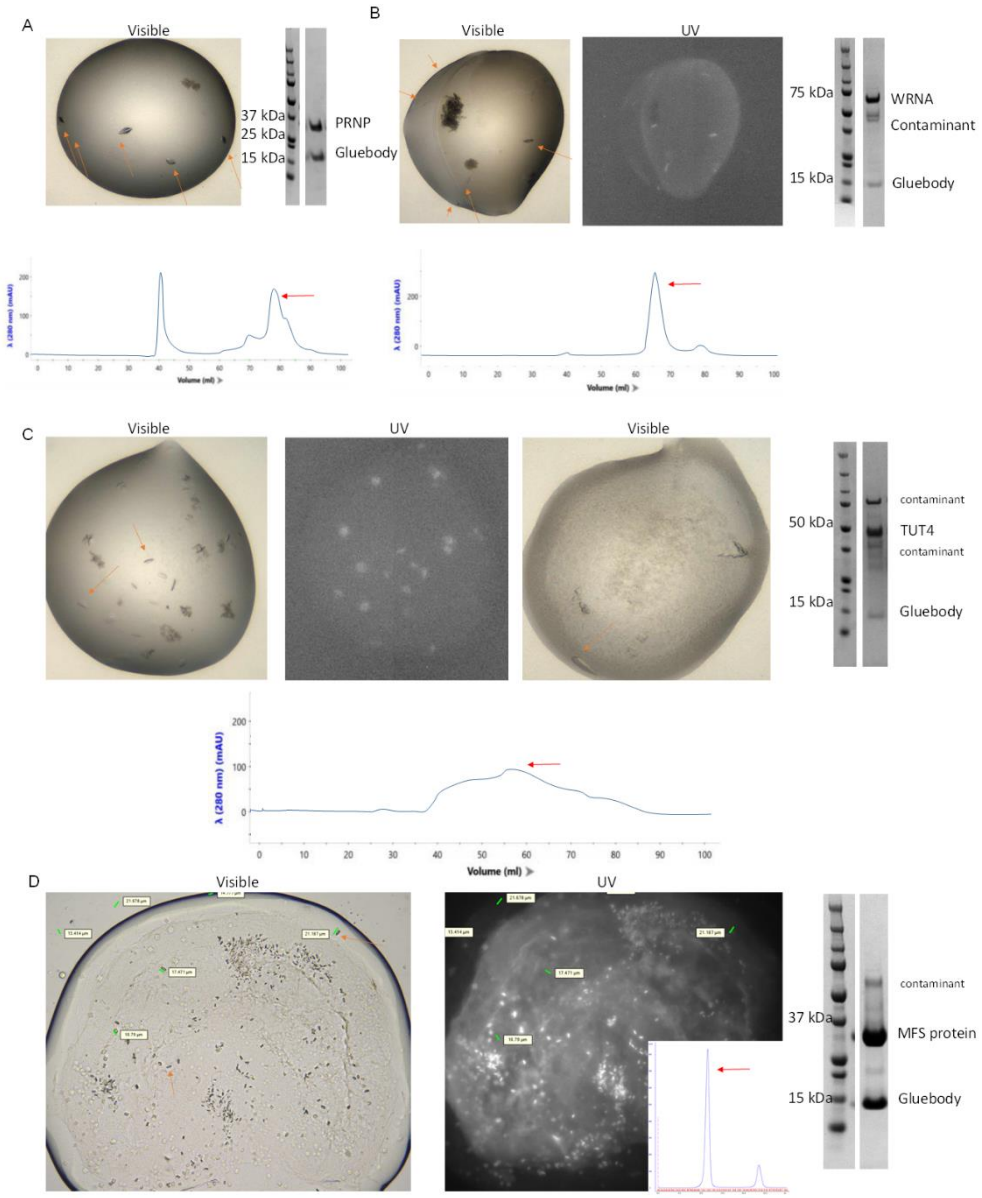
Supplementary Figure 4 The computational workflow of nanobody-nanobody crystal contact analysis and classification results. The texts in red represent actions taken in each step. Structure cartoons in marine and light blue represent two nanobody molecules participating in the interface. Residues represented as sticks are interacting residues of the current nanobody molecules. Residues represented as pink lines or sticks are interacting residues of the reference nanobody molecules. Class 8 interfaces are between fusion parts of the nanobody and therefore are not counted as nanobody-nanobody interfaces.



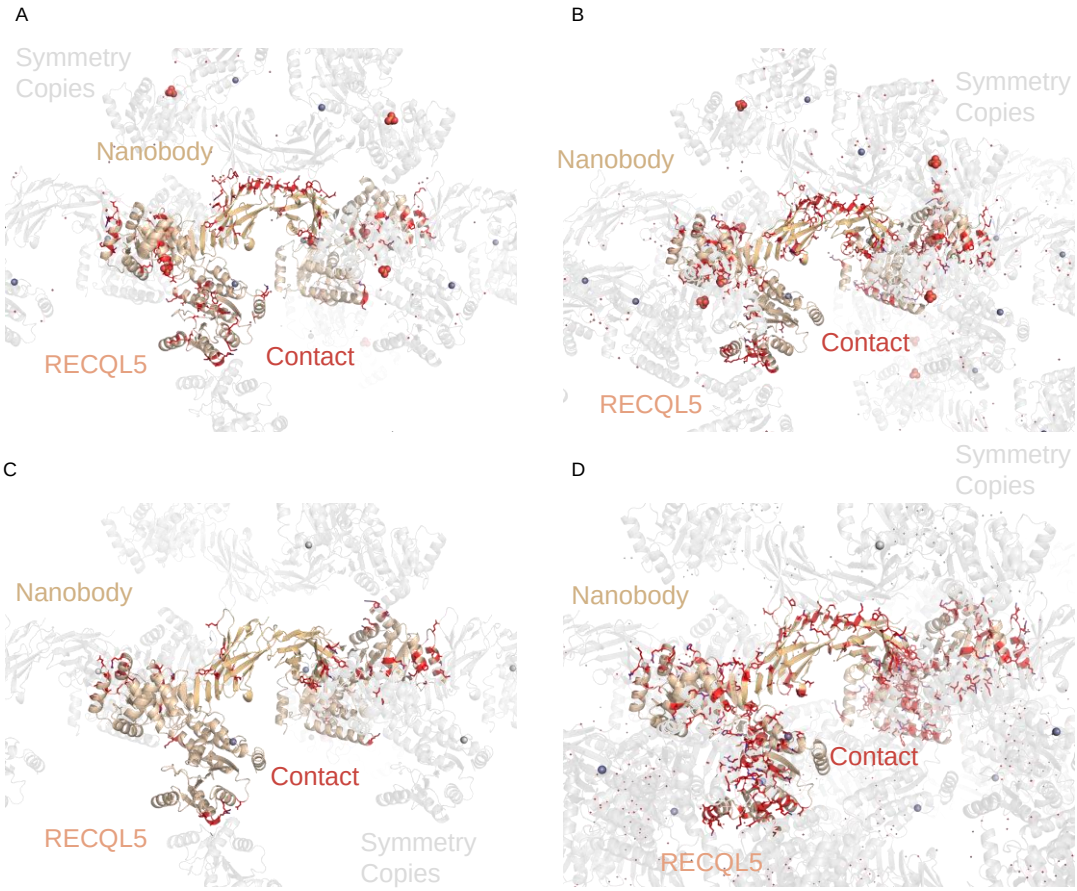
Supplementary Figure 5 Resolution comparison of different crystal forms from RECQL5:nanobody complexes. Each combination of crystal forms and crystallization condition categories are grouped and indicated in different colors.



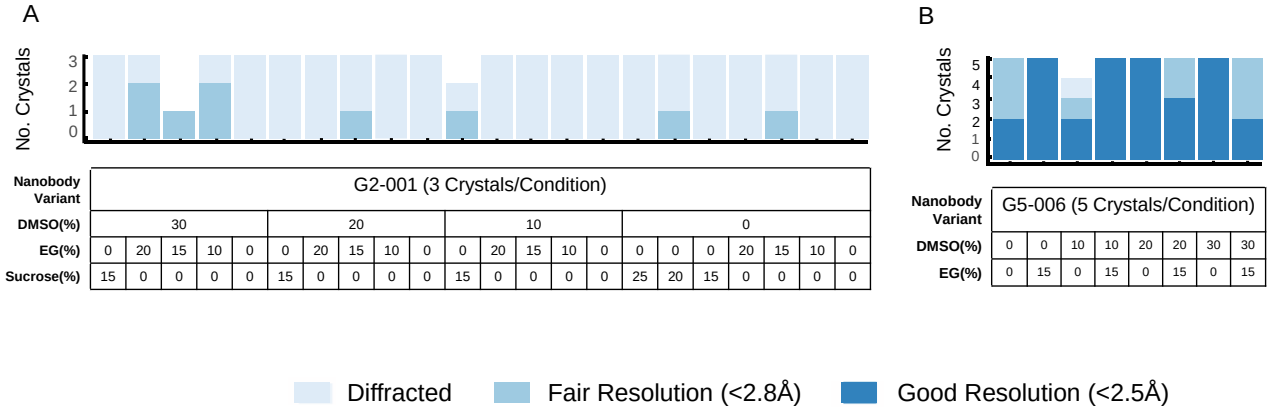
Supplementary Figure 6 Crystals from PRNP, WRN, TUT4 and MFS protein in complex with respective Tetra-Gluebodies. There are visible and UV images for crystals and SDS-PAGE, SEC profiles showing the quality of the protein. The protein peaks in SEC collected for crystallization trials are indicated by the red arrows. Crystals are indicated by orange arrows. A) PRNP:Gluebody complex crystals. Crystallization happened at 14 mg/ml, 20 °C in the condition of 0.2M Magnesium chloride hexahydrate, 0.1M tris pH 8.5 25% w/v PEG3350, with a protein:precipitant ratio of 1:2. B) WRN:Gluebody complex crystals. Crystallization happed at 27.3 mg/ml, 20 °C in the condition of 0.15 M Potassium bromide, 30% w/v Polyethylene glycol monomethyl ether 2,000, with a protein:precipitant ratio of 1:2. C) TUT4:Gluebody complex crystals. There are several conditions yielding crystals. In the visible image on the left, the condition is 0.2 M Magnesium chloride hexahydrate, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350, with a protein:precipitant ratio of 1:2. In the visible image on the right, the condition is 0.1 M HEPES pH 7.5, 2.0 M Ammonium sulfate, with a protein:precipitant ratio of 1:2. All crystallization happened at 14 mg/ml, 20 °C. D) The LCP crystals formed in the condition of 0.1M NaCl 0.1M Li_2SO_4 40%v/v PEG200 0.1M MES pH6 at 40 mg/ml, 20 °C.



<u>Supplementary Figure 7</u> Detailed crystal contacts of the RECQL5:nanobody complex from the A) Tetramer P 2 2 21 porous, B) Tetramer P 2 2 21 compact, C) Dimer P 2 2 21 shifted and D) Tetramer C 1 2 1 crystal forms. Crystal contacts are shown in red.

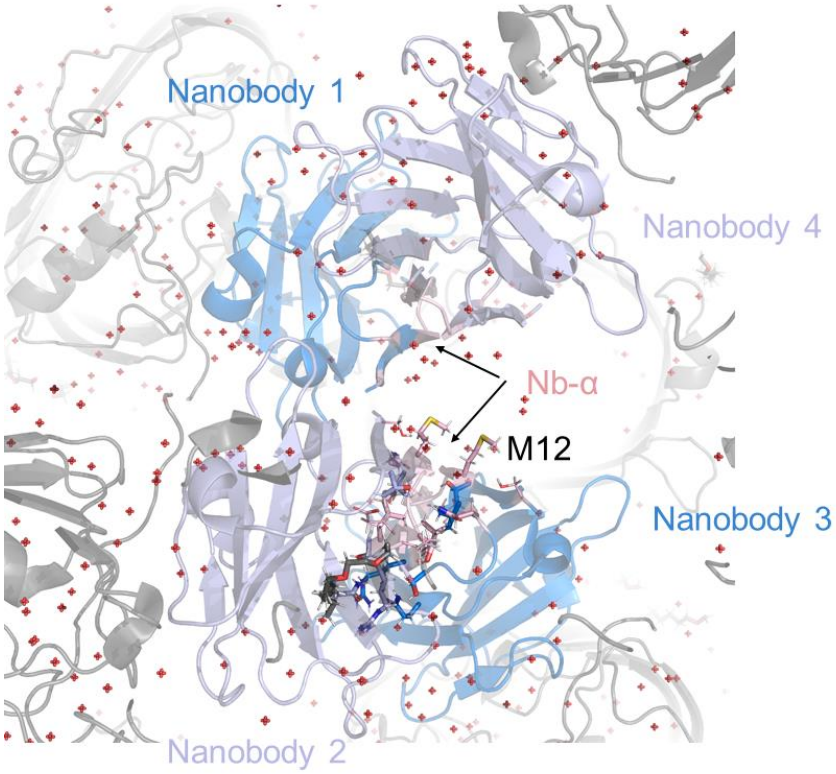


Supplementary Figure 8 Diffraction quality of G5-006 (A) and G5-009 (B) in different soaking conditions measured by the percentage of crystals which diffract to high resolution. In the barplot, deep blue fractions represent good resolution (<2.5 Å) crystal percentage, blue bar heights represent fair resolution (<2.8 Å) crystal percentage and the total heights of the bars represent percentage of crystals that diffracted. The numbers of crystals, nanobody variants and soaking conditions in each group are indicated in the table below each bar.

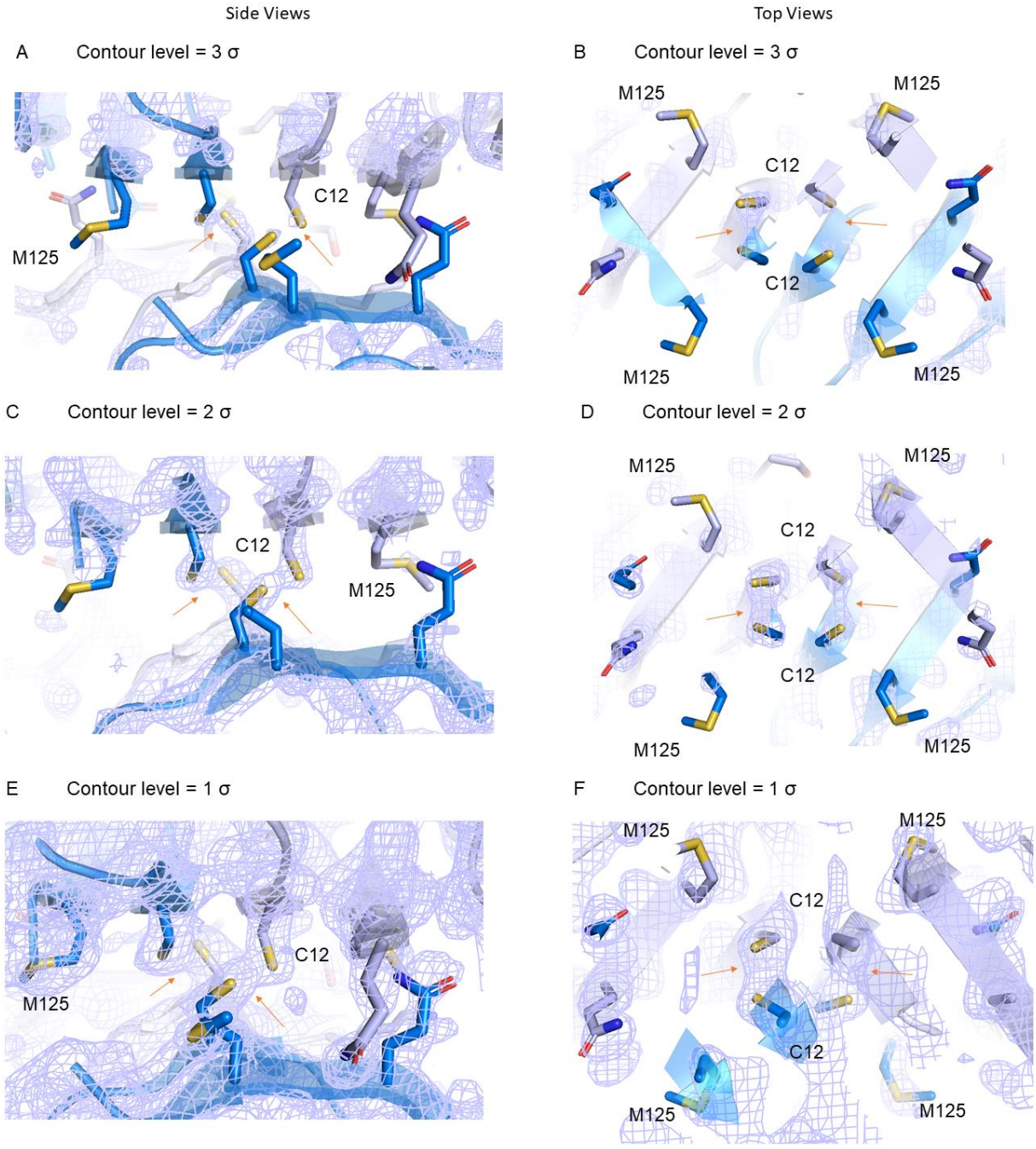


Diffracted Fair Resolution (<2.8Å) Good Resolution (<2.5Å)

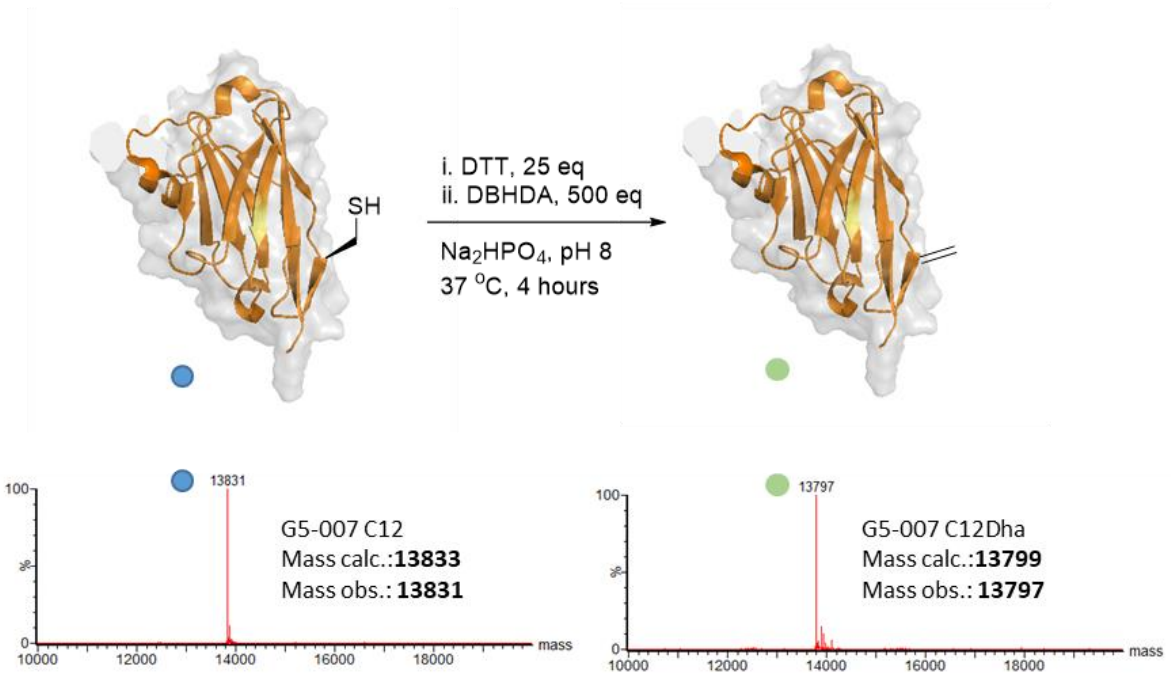
Supplementary Figure 9 Tilted back-to-back packing of nanobody dimers from PDB id 6QGY. Molecules colored in marine and light blue are nanobody molecules. Nb-α is colored pink. The target protein molecules of the nanobodies are colored grey.



Supplementary Figure 10 Electron density connecting the sulfur atoms in the C-C interface core viewed from the side and the top. C12 and M125 are presented in stick and the sulfur atoms are colored yellow. The electron density around C12 is indicated in light blue mesh with the contour levels at 1σ, 2σ and 3 σ.



Supplementary Figure 11 Chemical conversion of the Cys12 on G5-007 nanobody to dehydroalanine (Dha). To a solution of G5-007 nanobody (1 mL, 1 mg/mL, 72 µM) in Na_2HPO_4 buffer (50 mM, pH 8.0), 0.28 mg (25 eq.) of DTT were added. The solution was shaken at room temperature for 20 min and then the protein was treated with 22 µL of 0.5M DBHDA (2,5-Dibromohexanediamide) in DMSO (500 eq.) and heated to 37 °C for 4 h, at which point analysis by mass spectrometry showed reaction completion. The protein was purified using a PD Miditrap™ G-25 column (Cytiva #28918008) pre-equilibrated with Na_2HPO_4 buffer (50 mM, pH 8.0).



References

1 Chung, S., Shin, S.-H., Bertozzi, C. R. & De Yoreo, J. J. Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proceedings of the National Academy of Sciences* **107**, 16536-16541 (2010).

2 Sauter, A. *et al.* Real-time observation of nonclassical protein crystallization kinetics. *Journal of the American Chemical Society* **137**, 1485-1491 (2015).

3 Sleutel, M., Lutsko, J., Van Driessche, A. E., Durán-Olivencia, M. A. & Maes, D. Observing classical nucleation theory at work by monitoring phase transitions with molecular precision. *Nature communications* **5**, 1-8 (2014).

4 Van Driessche, A. E. *et al.* Molecular nucleation mechanisms and control strategies for crystal polymorph selection. *Nature* **556**, 89-94 (2018).

5 Yau, S.-T. & Vekilov, P. G. Quasi-planar nucleus structure in apoferritin crystallization. *Nature* **406**, 494-497 (2000).

6 Newman, J. *et al.* On the need for an international effort to capture, share and use crystallization screening data. *Acta Crystallographica Section F* **68**, 253-258, doi:doi:10.1107/S1744309112002618 (2012).

7 Strelov, V., Kuranova, I., Zakharov, B. & Voloshin, A. Crystallization in space: Results and prospects. *Crystallography Reports* **59**, 781-806 (2014).

8 Price Ii, W. N. *et al.* Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nature biotechnology* **27**, 51-57 (2009).

9 Bukowska, M. A. & Grütter, M. G. New concepts and aids to facilitate crystallization. *Current opinion in structural biology* **23**, 409-416 (2013).

10 Savitsky, P. *et al.* High-throughput production of human proteins for crystallization: the SGC experience. *Journal of structural biology* **172**, 3-13 (2010).

11 Derewenda, Z. S. Rational protein crystallization by mutational surface engineering. *Structure* **12**, 529-535 (2004).

12 Thorsen, T. S., Matt, R., Weis, W. I. & Kobilka, B. K. Modified T4 lysozyme fusion proteins facilitate G protein-coupled receptor crystallogenesis. *Structure* **22**, 1657-1664 (2014).

13 Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20**, 967-976 (2012).

14 Day, P. W. *et al.* A monoclonal antibody for G protein–coupled receptor crystallography. *Nature methods* **4**, 927-929 (2007).

15 Fairhead, M. *et al.* Strategies to improve scFvs as crystallization chaperones suggested by analysis of a complex with the human PHD-bromodomain SP140. *BioRxiv*, 767376 (2019).

16 Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β 2 adrenoceptor. *Nature* **469**, 175-180 (2011).

17 Batyuk, A., Wu, Y., Honegger, A., Heberling, M. M. & Plückthun, A. DARPin-based crystallization chaperones exploit molecular geometry as a screening dimension in protein crystallography. *Journal of molecular biology* **428**, 1574-1588 (2016).

18 Boder, E. T., Raeeszadeh-Sarmazdeh, M. & Price, J. V. Engineering antibodies by yeast display. *Archives of biochemistry and biophysics* **526**, 99-106 (2012).

19 Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making antibodies by phage display technology. *Annual review of immunology* **12**, 433-455 (1994).
20 Zimmermann, I. *et al.* Generation of synthetic nanobodies against delicate proteins. *Nature protocols* **15**, 1707-1741 (2020).
21 Ward, A. R. & Snow, C. D. Porous crystals as scaffolds for structural biology. *Current opinion in structural biology* **60**, 85-92 (2020).
22 Maita, N. Crystal structure determination of ubiquitin by fusion to a protein that forms a highly porous crystal lattice. *Journal of the American Chemical Society* **140**, 13546-13549 (2018).
23 Zacharchenko, T. & Wright, S. Functionalization of the BCL6 BTB domain into a noncovalent crystallization chaperone. *IUCrJ* **8** (2021).
24 Saridakis, E. & Chayen, N. E. Imprinted polymers assisting protein crystallization. *Trends in biotechnology* **31**, 515-520 (2013).
25 Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature* **469**, 175-180, doi:10.1038/nature09648 (2011).
26 Steyaert, J. & Kobilka, B. K. Nanobody stabilization of G protein-coupled receptor conformational states. *Current opinion in structural biology* **21**, 567-572 (2011).
27 Collins, P. M. *et al.* Achieving a good crystal system for crystallographic X-ray fragment screening. *Methods in enzymology* **610**, 251-264 (2018).
28 Krojer, T., Fraser, J. S. & von Delft, F. Discovery of allosteric binding sites by crystallographic fragment screening. *Current Opinion in Structural Biology* **65**, 209-216 (2020).
29 Wukovitz, S. W. & Yeates, T. O. Why protein crystals favour some space-groups over others. *Nature Structural Biology* **2**, 1062-1067, doi:10.1038/nsb1295-1062 (1995).
30 Banatao, D. R. *et al.* An approach to crystallizing proteins by synthetic symmetrization. *Proc Natl Acad Sci U S A* **103**, 16230-16235, doi:10.1073/pnas.0607674103 (2006).
31 Lieu, R. *et al.* Rapid and robust antibody Fab fragment crystallization utilizing edge-to-edge beta-sheet packing. *PloS one* **15**, e0232311 (2020).
32 Chesterman, C. & Arnold, E. Co-crystallization with diabodies: a case study for the introduction of synthetic symmetry. *Structure* (2021).
33 Carugo, O. & Djinović-Carugo, K. How many packing contacts are observed in protein crystals? *Journal of structural biology* **180**, 96-100 (2012).
34 Dunbar, J. & Deane, C. M. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* **32**, 298-300 (2016).
35 Josephson, B. *et al.* Light-driven post-translational installation of reactive protein side chains. *Nature* **585**, 530-537 (2020).
36 Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nature communications* **10**, 1-7 (2019).
37 Uchański, T. *et al.* Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nature Methods* **18**, 60-68, doi:10.1038/s41592-020-01001-6 (2021).
38 Wu, X. & Rapoport, T. A. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proceedings of the National Academy of Sciences* **118**, e2115001118, doi:10.1073/pnas.2115001118 (2021).
39 Higashi, L. S., Lundeen, M. & Seff, K. Empirical relations between disulfide bond

lengths,(nitrogen or carbon)-carbon-sulfur-sulfur torsion angles, and substituents in aromatic disulfides. Crystal and molecular structure of 3, 3'-dihydroxydi-2-pyridyl disulfide. *Journal of the American Chemical Society* **100**, 8101-8106 (1978).

40 DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography* **40**, 82-92 (2002).

41 Gileadi, O. *et al.* in *Structural Proteomics* 221-246 (Springer, 2008).

42 Wright, N. D. *et al.* The low-cost Shifter microscope stage transforms the speed and robustness of protein crystal harvesting. *Acta Crystallographica Section D: Structural Biology* **77** (2021).

43 Delagenière, S. *et al.* ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics* **27**, 3186-3192 (2011).

44 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **43**, 186-190 (2010).

45 Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D: Biological Crystallography* **67**, 293-302 (2011).

46 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235-242 (2011).

47 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007).

48 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **67**, 355-367 (2011).

49 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica section D: biological crystallography* **60**, 2126-2132 (2004).

Supplementary tables

Supplementary Table 1 The crystallization experiment on a selection of target:nanobody pairs with naïve nanobody scaffolds.

Supplementary Table 2 Detailed DBSCAN result with full list of PDB codes of each nanobody-nanobody interface category. In the 'Pattern name' column, first four letters represent PDB and the digit after '_' represents the No. of the pattern in the structure. In the 'Pattern' column, each number represents the IMGT normalized index of a residue of the nanobody scaffold involved in the pattern.

Supplementary Table 3 Full list and the crystallization result of nanobody variants tested in the RECQL5:nanobody system

Supplementary Table 4 Full list of diffraction data collected from crystals of all nanobody variants tested in the RECQL5:nanobody system

Supplementary Table 5 Detailed information with numeric values of each cell in Figure 3A