1 **Strategies to improve scFvs as crystallization chaperones suggested by analysis of a**

2 **complex with the human PHD-bromodomain SP140**

3 Michael Fairhead[1], Charlotta Preger[2], Edvard Wigren[2], Claire Strain-Damerell[3], Elena

4 Ossipova[2], Mingda Ye[1], Mpho Makola[1], Nicola A. Burgess-Brown[1], Helena Persson[4,5], Frank

5 von Delft[1], Susanne Gräslund[2].*

6

7 [1]Structural Genomics Consortium, University of Oxford, Old Road Campus Research

8 Building, Roosevelt Drive, Oxford, OX3 7DQ, UK

9 [2]Structural Genomics Consortium, Rheumatology Unit, Department of Medicine Solna,

10 Karolinska Institutet, Stockholm, 171 76, Sweden

11 [3]Diamond Light Source Ltd., The Research Complex at Harwell, Harwell Science and

12 Innovation Campus, Didcot, Oxfordshire, OX11 0FA, UK

13 [4]Science for Life Laboratory, Drug Discovery and Development, Stockholm, 171 65, Sweden

14 [5]School of Engineering Sciences in Chemistry, Biotechnology and Health, Royal Institute of

15 Technology (KTH), Stockholm, 114 28, Sweden

16

17 *Corresponding author: susanne.graslund@ki.se

18

19 **Short title:** Towards better scFv crystallization chaperones

20

21

22

23

24

25

1 Abstract

2 Antibody fragments have great potential as crystallization chaperones for structural biology

3 due to their ability to either stabilise targets, trap certain conformations and/or promote crystal

4 packing. Here we present an example of using a single-chain variable fragment (scFv) to

5 determine the previously unsolved structure of the multidomain protein SP140. This nuclear

6 leukocyte-specific protein contains domains related to chromatin-mediated gene expression

7 and has been implicated in various disease states. The structure of two of the domains (PHD-

8 bromodomain) was solved by crystallizing them as a complex with a scFv generated by phage

9 display technology. SP140 maintains a similar overall fold to previous PHD-bromodomains

10 and the scFv CDR loops predominately interact with the PHD, while the framework regions of

11 the scFv makes numerous interactions with the bromodomain. Analysis of our and other

12 complex structures suggest various protein engineering strategies that might be employed to

13 improve the usefulness of scFvs as crystallization chaperones.

14

15 Introduction

16 Despite extensive efforts in construct design (1, 2) and screening of chemical space (3), many

17 targets routinely fail to yield the protein crystals necessary for structural studies *via* x-ray

18 diffraction. This failure may be due to poor protein stability (4), multiple conformational states

19 leading to sample heterogeneity or it even being the wrong time of year (5)! Antibody

20 fragments have proven useful in overcoming some of these issues by acting as crystallization

21 chaperones (6). They can stabilise and/or trap specific conformations of proteins as well as

22 directly mediate crystal contacts. Mimetic binders based on non-immunoglobulin scaffolds

23 have also been engineered and proven useful in this application (7).

1   Examples of target:binder complexes in the PDB are numerous and steadily growing, with

2   several different binder types having been used, such as Fabs, scFvs, nanobodies and Darpins

3   (see Table 1). Currently there is no clear consensus which binder format is best and most likely

4   this will vary from target to target. For structural studies binders should have a stable and rigid

5   scaffold and regions (such as loops) that tolerate high sequence variability that generates

6   specificity for a target. Ideally binders should also be produced in a way that gives rise to a

7   homogenous preparation (*i.e.* without glycosylations and other modifications).

8

9   **Table 1. Number of binder and binder:complex structures* in the PDB (21st June 2019)**

Binder class	Number of structures[1]	Number of complexes[2]
Affibody	13	10
Affimer	9	2
Darpin	141	70
Fab	2502	1646
Nanobody	293	207
scFv (+Fv)[4]	165 (255)	118 (158)

10 *The search outcomes were not specifically curated beyond the basic search criteria. [1]The type of
11 "binder class" was used to search the full text of the mmCIF file deposited with the PDB. [2]The type of
12 "binder class" & the term "complex" was used to search the PDB. Searches were made using the
13 advanced search page on the PDB (http://www.rcsb.org/pdb/search/advSearch.do).[4] Although we
14 searched for scFv complexes in the PDB several examples of Fv complexes were found. These have
15 been included in the analysis to increase the number of examples studied. Observations with these
16 binder types should be translatable to the scFv arrangement.

17

18   A single chain fusion of the variable heavy (V_H) and light (V_L) domains (connected via a

19   peptide linker) of an immunoglobulin (Ig) (8, 9) can fulfil many of the above requirements.

20   The complementary determining regions (CDR) on the surface of the V_H and V_L domains

21   together generate the antigen-binding site, with the potential of providing high specificity and

22   affinity to most proteins and other types of antigens (10-13). ScFvs with new specificities can

23   be readily generated and selected via techniques such as phage display (14) and if necessary

24   they can be further engineered for increased performance, for example by increasing its

25   stability (15) or affinity (16, 17). We chose the epigenetic target, SP140, as a case study to

1   help develop a pipeline that has the ambition of routinely generating scFvs that recognize

2   natively folded proteins and which can be used as crystallization chaperones.

3   SP140 is a nuclear leukocyte specific protein containing chromatin reader modules (Figure

4   1A). So far only the NMR structure of one of these modules (plant homeodomain, PHD) (18)

5   and the PHD-bromodomain x-ray crystal structure of the related SP100C (19) and SP100A

6   (4PTB) proteins have been reported. Although the Structural Genomics Consortium (SGC) has

7   historically been successful in solving the structures of many other bromodomains and related

8   chromatin reader modules (20) the structure of SP140 had remained unsolved. Using an scFv,

9   generated by phage display, we solved the structure of the PHD-Bromodomain of SP140

10   (6G8R). As well as generating a novel structure, we analysed this complex and those reported

11   by others to investigate some of the features of scFvs, which may aid crystallization and points

12   the way to possible improvements to this scaffold as a crystallization chaperone.

13

14   **Figure 1: Isolation, crystallization and structure of the scFv:SP140 complex.** (A) Overview of

15   domains in human SP140. (B) Size exclusion chromatograms of (i) scFv:SP140A mix, (ii) scFv and

16   (iii) SP140 respectively. (C) Crystals of the scFv:SP140 complex obtained in 1.2 M sodium malonate,

17   0.5 % jeffamine ED-2003, 0.1 M HEPES pH 7. (D) Cartoon representation (Pymol) of 6G8R with the

18   PHD-bromodomain of SP140, the scFv V_H and V_L and their respective CDR loops coloured as

19   indicated.

20

21   **Results and Discussion**

22

23   **scFv generation, expression and purification.**

4

1. ScFv binders against the PHD-bromodomain of SP140 (Supplementary Method 1) were

2. generated by phage display technology using a human synthetic library denoted SciLifeLib

3. with biotinylated recombinant SP140 protein as target protein. The top candidate, G-SP140-8,

4. with a measured affinity of approximately 6 nM (Supplementary Figure S1) was selected and

5. used in co-crystallization experiments with SP140.

6. Although scFvs are useful binder scaffolds, we have found that production in *Escherichia coli*

7. routinely gives low yields of less than 1 mg/L of cells for some if not most constructs. As

8. crystallography can require several milligrams of protein, we evaluated different strategies to

9. improve yield and purity. The most successful strategy was to produce the G-SP140-8 scFv as

10. an MBP fusion and express it in the oxidising cytoplasm of the mutant *E. coli* strain Shuffle

11. (NEB) (21). Isolation of the MBP-scFv fusion with Ni-Sepharose and Protein A (22) resins

12. yielded highly pure protein. Subsequent MBP removal by His_6-tagged TEV protease followed

13. by reverse chromatography with Ni-Sepharose and amylose resins (to remove MBP, uncleaved

14. protein and his tagged TEV protease) routinely generated more than 5 mg of purified scFv per

15. litre.

16.

17. **Complex isolation and crystallization**

18. After mixing an excess of SP140 with scFv, the resulting complex was readily isolated via size

19. exclusion chromatography (SEC) (Figure 1B i-iii). A clear shift in molecular weight can be

20. observed for the complex as compared to the individual components, (Figure 1B i-iii).

21. Curiously the isolated scFv was always found to run lower than its predicted MW (Figure 1B

22. ii). This may indicate some dissociation of the two domains during the run (V_H and V_L), *i.e.*

23. the protein overall behaves more like a long linear molecule rather than a single larger globular

24. protein. After isolation by SEC, the complex of scFv:SP140 was put directly into crystallization

1   trials and crystals with a bipyramid morphology were found in many conditions after 1-2 days

2   at room temperature. Unfortunately, most of these crystals were found to diffract to a low

3   resolution (3-5 Å). Only crystals grown in a mixture of malonate/jeffamine (Figure 1C.) were

4   found to diffract to a resolution of 2.7 Å and allow the generation of a structure with reasonable

5   refinement statistics, Table 2.

1   **Table 2. Data collection and refinement statistics.**

	6G8R
Wavelength	0.96861 Å
Resolution range	51.32 - 2.74 (2.84 - 2.74) Å
Space group	P 61 2 2
Unit cell	89.2x89.2x343.35 Å 90x90x120 °
Total reflections	421847 (42464)
Unique reflections	22245 (2164)
Multiplicity	19.0 (19.6)
Completeness (%)	99.46 (99.36)
Mean I/sigma(I)	11.89 (1.02)
Wilson B-factor	72.26
R-merge	0.2313 (3.242)
R-meas	0.2376 (3.327)
R-pim	0.05388 (0.7413)
CC1/2	0.998 (0.652)
CC*	1 (0.889)
Reflections used in refinement	22226 (2160)
Reflections used for R-free	1139 (107)
R-work	0.2243 (0.3616)
R-free	0.2665 (0.3482)
CC(work)	0.918 (0.756)
CC(free)	0.905 (0.679)
Number of non-hydrogen atoms	3057
macromolecules	3033
ligands	10
solvent	14
Protein residues	393
RMS(bonds)	0.003 Å
RMS(angles)	0.9 °
Ramachandran favoured (%)	93.67
Ramachandran allowed (%)	6.33
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.93
Clashscore	2.54
Average B-factor	74.49
macromolecules	74.55
ligands	72.99
solvent	61.43

2   Statistics for the highest-resolution shell are shown in parentheses.

3

Structure of SP140 PHD:Bromodomain

5   The packing observed in the crystal structure consists of a crystallographic dimer of SP140

6   (Chain B) surrounded by four copies of the scFv (Chain A) (Supplementary Figure S5).

7   Specifically, the structure of the SP140 PHD:Bromodomain is very similar to that observed for

8   the related SP100 (4PTB) and SP100C (5FB1) proteins with less than a 0.7-0.8 Å RMSD

1 between the structures (Supplementary Figure S6). This small difference is perhaps

2 unsurprising giving the relatively high sequence identity of 61 % (Supplementary sequence

3 S3).

4 It has previously been demonstrated that the isolated PHD cannot bind the substrate histone

5 tail peptide (18) in the absence of the bromodomain (19). This suggests that the bromodomain

6 is necessary to stabilise the conformation of the PHD and that this stabilization is required to

7 allow binding (19). The structure of the SP140 PHD:Bromodomain presented here confirms

8 this observation as the PHD adopts a conformation similar to that observed in two related

9 PHD:Bromodomain structures, the substrate peptide bound SP100C (5FB1), (19) and the apo

10 SP100 (4PTB). This is in contrast to the isolated PHD NMR structures which have distinct cis

11 or trans conformations (18), neither of which resemble the SP140 PHD:Bromodomain structure

12 (Supplementary Figure S5).

13

14 **Interaction of scFv with SP140**

15 As would be expected for an antibody-based binder, residues from the CDR-loops on the

16 surface of scFv form numerous interactions with SP140 (Figure 1D and Figure 2A),

17 predominantly through the PHD. PISA (23) analysis shows binding of the scFv to SP140 results

18 in an extensive interface covering a buried surface area of 842 Å^2 and of the six CDR-loops,

19 CDR-H3 (395 Å^2) dominates the interface followed by CDR-H2 (118 Å^2) and CDR-L3 (86

20 Å^2) (Supplementary Figure S7). None of the CDR loops occupy the peptide-binding site of the

21 PHD domain (Figure 2D) suggesting that the scFv would not inhibit protein activity.

22 More unexpected is the importance of the interactions of residues on the side of the scFv

23 opposite to the CDR-loops, which participate in crystal contacts with neighbouring copies of

1   SP140 (Figure 2B). Crystal packing is also observed between neighbouring scFv molecules

2   (Figure 2C).

3   The interaction of the scFv with both the bromodomain and PHD and its probable stabilising

4   effect may be important given the above noted fact that the PHD in isolation appears to be non-

5   functional (18). This might explain our failure to crystallize the isolated protein, despite success

6   with many similar classes of protein (20).

7

8   **Figure 2: Binding and packing features of the scFv:SP140 complex.** Cartoon representation (Pymol)

9   of (A) CDR loop interactions of the scFv with surface residues of SP140. (B) Interactions of the

10   backside of the scFv with a symmetry copy of SP140 (B*). (C) Interactions of the C-terminal tail of the

11   scFv with a symmetry copy (A*). (D) overlay of 6G8R coloured as in Figure 1D with peptide bound

12   SP100C structure (5FB1) with bound peptide in stick representation (hot pink) and the PHD and

13   bromodomains coloured hot pink and pale pink respectively.

14

15   **scFvs as crystallization chaperones**

16   Of the binder:complex structures reported in the PDB (Table 1) sixty-one of these contain an

17   scFv or Fv in complex with another protein, the remainder being small molecules, peptides,

18   apo or anomalies (see Table S1 for details). As can be seen from Table S2, these binders have

19   been used in isolation and with other antibody fragments such as Fabs to solve a wide variety

20   of proteins from small toxins (24) to large virus assemblies (25, 26) and a range of proteins

21   in between, demonstrating their utility as crystallization chaperones. While studies have shown

22   that the crystallization properties of the scFv can be modified to produce new space groups,

23   until recently this has been limited to scFvs binding small molecules (27) or peptides (28).

24   However, a recent example demonstrates the potential benefits of engineering the scFv (as well

1   as a Fab) for improved crystal packing, allowing a structural study of small molecules bound

2   to the gp120 subunit of the HIV-1 envelope trimer (29). We reasoned that an analysis of our

3   and others protein:scFv/Fv complexes would yield suggestions on how to generate scFvs with

4   improved crystallization properties.

5   Table S3 shows a summary of information on the scFv/Fv and their amino acid sequences in

6   these complexes. It can be seen from Table S3 that different investigators have arranged the

7   V_L and V_H domains from different species in a variety of orders (*i.e.* V_L or V_H can be first in

8   the scFv) with different linkers and tags. It can also be seen that in some cases the V_L and V_H

9   domains were sufficiently stable as a complex to not require a linker (Fv rather than scFv).

10   A simple analysis of the complexes (Table S3) shows that many regions of the scFv constructs

11   are not visible in the final crystal structure; these include long C-terminal purification tags (*e.g.*

12   His_6, Avi tags). Usually such flexible tags are removed from targets when setting up crystal

13   trials as they are typically disordered and rarely visible in the final structure. The increased

14   entropic cost associated with this disorder reduces the likelihood of crystallization (30). It

15   would be best if the N- and C-termini of the scFv were trimmed to their approximate Fv

16   boundaries.

17   Another observation is that the V_H-V_L linker is rarely visible in structures (Table S3). If this

18   linker could be redesigned to be more rigid for the above reasons of entropic cost, this may

19   increase crystallization success rates. One approach may be to simply increase the serine

20   content, which has been reported to increase stiffness (31). Alternatively a computational

21   approach to the design could be taken (32). Care has to be taken as linker modifications can

22   have unintended adverse effects such as reducing affinity, stability and oligomerization (33,

23   34), any of which which may be disadvantageous for crystallization and target binding.

In Table S4 an analysis of the regions commonly found in Fv crystal contacts (not involving the primary Fv:target complex) shows that many crystal packing arrangements of Fv molecules are possible. These multiple possible packing arrangements may partly explain the usefulness of these binders as crystallization chaperones.

A summary of the various Fv crystal contacts and an alignment of the different V_L and V_H domains (IMGT numbering done using the ANARCI program (35)) is shown in Tables S5 and S6. Although there are regions which appear to be quite frequently observed at crystal contacts (*e.g.* around residues 5-15 in V_L), a limitation is that the number of examples of complexes analysed is relatively small (61 in total) and that the Fv domains are from a number of species (human, mouse, rhesus, rabbit & rat), meaning a high degree of sequence variability. Even with these limitations it is perhaps surprisingly how frequently the same regions are found in crystal packing points with other Fv (Tables S5 and S6). Targeting of these regions with mutagenesis is likely a good strategy for improving scFv as crystallization chaperones.

Analysis of the alignments in Tables S5 and S6 also shows that the Fv:Fv packing frequently involve a glutamine, lysine or glutamate. These residues are associated with high entropic cost and mutating these to alanine or a residue with lower entropy (arginine, aspartate or asparagine) has been shown in other proteins to be a successful strategy to improve crystallization success (36). Such a surface entropy reduction (SER) strategy has been done to improve the crystallization properties of an EE epitope (EYMPME) binding scFv (28), although without a target protein bound. Similarly a SER approach has been used on a Fab scaffold to improve crystallization when in complex with RNA (37).

Conclusions

1   The structure of SP140 in complex with scFv presented here shows that the PHD adopts the

2   expected fold as observed in previous PHD:Bromodomain structures such as SP100 (4PTB)

3   and SP100C (5FB1) The success of this crystallization attempt is likely a result of stabilization

4   of the PHD due to the presence of interactions with both the bromodomain and the scFv.

5   Analysis of the scFv in this and previously reported structures suggests that the utility of this

6   class of binders as routine crystallization chaperones would be greatly improved through

7   certain modifications, summarized in Figure 3.

8

9   **Figure 3: Summary of potential strategies to improve scFv as crystallization chaperones.** (A)

10   Trimming of scFv to standard Fv boundaries prior to setting up crystallization trials, since tags used for

11   purification and/or assays are rarely observed in crystal structures. (B) Alternative linkers should be

12   explored for connecting the V_H and V_L domains, since the widely used flexible glycine serine linkers

13   are rarely observed in crystal structures. (C) Mutations made to surface residues (non-CDR) to promote

14   crystal packing, targeting either surface entropy reduction (38) or introduction of crystal contacts found

15   in other deposited structures (39). (D) Mutations made to improve solubility and/or stability, which

16   may mean adding a penta peptide tag (40) or a more extensive series of mutations such as that

17   implemented in the PROSS approach (41).

18

19   **Methods**

20   **SP140 protein expression and purification**

21   The DNA sequence corresponding to amino acids 688-862 of the SP140 protein (UniProtKB

22   – Q13342) was sub-cloned into the expression vector pNIC-Bio2 adding an N-terminal His_6

23   tag and a C-terminal Avi tag, creating the SP140A-c055 construct, Supplementary sequence

1   S1. Expression and purification were performed essentially as described (42), see

2   supplementary methods S1 for a detailed protocol.

3

4   **Antibody phage selection and ELISA screening of selected scFv**

5   The phage selection procedure was carried out basically as described (43) but the entire

6   process, including the steps of antigen-phage incubation to trypsin elution, was carried out in

7   ordinary 1.5 mL tubes on a rotator with no automation. The number of washing steps was

8   modified and increased with succeeding selection rounds; six in round one and 12 in round

9   four. Also, the recovered phages were propagated in XL1-Blue *E. coli* between the selection

10 rounds. Re-cloning of the selected material in pool followed by transformation into TOP10 *E.*

11 *coli*, small-scale expression of 94 randomly picked scFv and subsequent ELISA and

12 sequencing experiments were performed analogues to previously reported (43). The affinity

13 was measured using SPR, Supplementary Method S3 and Figure S1.

14

15 **scFv protein expression and purification**

16 The scFv gene was sub-cloned into pNIC-MBP2-LIC, adding an N-terminal TEV cleavable

17 MBP-His$_6$ tag, creating the YYSP140A-c009 construct (Supplementary sequence S2). This

18 plasmid was used along with pGro7 (Takara) to transform Shuffle T7 Express cells (NEB).

19 Expression and purification was done using a similar method to that described previously (1,

20 2), see Supplementary methods S2 and Supplementary Figure S2 for further details.

21

22 **Complex isolation**

13

1   To isolate the scFv:SP140 complex, 3 mg of SP140 were mixed with 1 mg of scFv. The mixture

2   was then injected onto a 104 mL Yarra SEC-2000 column (Phenomenex) connected to a NGC

3   system (Bio-Rad) using 10 mM HEPES, 500 mM sodium chloride, 5 % glycerol, 0.5 mM

4   TCEP, pH 7.5 as the mobile phase, (Figure 1B and supplementary figures S2, S3 and S4).

5   Fractions corresponding to the complex were pooled and concentrated to 7.7 mg /mL.

6

7   **Crystallization, data collection and structure determination**

8   Crystallization trials were setup at 20 °C in JCSG-plus (Molecular Dimensions) and Crystal

9   Screen HT (Hampton Research) in Swissci 3 well sitting drop plates (Molecular Dimensions).

10   Crystals appeared in several conditions over 4-7 days. Crystals were mounted in MiTeGen

11   loops and flash frozen in liquid nitrogen using 30 % ethylene glycol as cryoprotectant. After

12   screening the best diffracting crystals were found in a condition containing 1.2 M sodium

13   malonate, 0.5 % jeffamine ED-2003, 0.1 M HEPES pH 7. A data set from a single crystal

14   obtained in this condition was collected at the I24 beamline at the Diamond Light Source. The

15   Xia2 auto-processed images (44) were used to perform molecular replacement using Phaser

16   (45). Further refinement of the structure was done using the Phenix suite of programs (46) in

17   combination with Coot (47) and Molprobity (48), see Table 2. The structure factors and co-

18   ordinates were deposited with the protein data bank, PDB code: 6G8R. Protein structure figures

19   were made using either PyMOL (Schrödinger, LLC.) or CCP4mg (49).

20

References

1. Gileadi O, Burgess-Brown NA, Colebrook SM, Berridge G, Savitsky P, Smee CE, et al. High throughput production of recombinant human proteins for crystallography. Methods Mol Biol. 2008;426:221-46.

2. Savitsky P, Bray J, Cooper CDO, Marsden BD, Mahajan P, Burgess-Brown NA, et al. High-throughput production of human proteins for crystallization: The SGC experience. Journal of Structural Biology. 2010;172(1):3-13.

3. Luft JR, Snell EH, Detitta GT. Lessons from high-throughput protein crystallization screening: 10 years of practical experience. Expert opinion on drug discovery. 2011;6(5):465-80.

4. Deller MC, Kong L, Rupp B. Protein stability: a crystallographer's perspective. Acta Crystallogr F Struct Biol Commun. 2016;72(Pt 2):72-95.

5. Cheng QD, Chen RQ, He J, Li DW, Yang F, Liu YM, et al. Effect of the weather conditions during solution preparation on lysozyme crystallization. Journal of Applied Crystallography. 2017;50:1341-51.

6. Griffin L, Lawson A. Antibody fragments as tools in crystallography. Clin Exp Immunol. 2011;165(3):285-91.

7. Batyuk A, Wu Y, Honegger A, Heberling MM, Pluckthun A. DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography. J Mol Biol. 2016;428(8):1574-88.

8. Skerra A, Pluckthun A. Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. Science. 1988;240(4855):1038-41.

9. Glockshuber R, Malia M, Pfitzinger I, Pluckthun A. A comparison of strategies to stabilize immunoglobulin Fv-fragments. Biochemistry. 1990;29(6):1362-7.

10. Polonskaya Z, Savage PB, Finn MG, Teyton L. High-affinity anti-glycan antibodies: challenges and strategies. Curr Opin Immunol. 2019;59:65-71.

11. Diamant E, Torgeman A, Ozeri E, Zichel R. Monoclonal Antibody Combinations that Present Synergistic Neutralizing Activity: A Platform for Next-Generation Anti-Toxin Drugs. Toxins (Basel). 2015;7(6):1854-81.

12. MacCallum RM, Martin AC, Thornton JM. Antibody-antigen interactions: contact analysis and binding site topography. J Mol Biol. 1996;262(5):732-45.

13. Persson H, Lantto J, Ohlin M. A focused antibody library for improved hapten recognition. J Mol Biol. 2006;357(2):607-20.

14. Griffiths AD, Malmqvist M, Marks JD, Bye JM, Embleton MJ, McCafferty J, et al. Human anti-self antibodies with high specificity from phage display libraries. EMBO J. 1993;12(2):725-34.

15. Chowdhury PS, Vasmatzis G. Engineering scFvs for improved stability. Methods Mol Biol. 2003;207:237-54.

16. Baran D, Pszolla MG, Lapidoth GD, Norn C, Dym O, Unger T, et al. Principles for computational design of binding antibodies. P Natl Acad Sci USA. 2017;114(41):10900-5.

17. Boder ET, Midelfort KS, Wittrup KD. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc Natl Acad Sci U S A. 2000;97(20):10701-5.

18. Zucchelli C, Tamburri S, Quilici G, Palagano E, Berardi A, Saare M, et al. Structure of human Sp140 PHD finger: an atypical fold interacting with Pin1. The FEBS journal. 2014;281(1):216-31.

19. Zhang X, Zhao D, Xiong X, He Z, Li H. Multifaceted Histone H3 Methylation and Phosphorylation Readout by the Plant Homeodomain Finger of Human Nuclear Antigen Sp100C. The Journal of biological chemistry. 2016;291(24):12786-98.

20. Filippakopoulos P, Picaud S, Mangos M, Keates T, Lambert JP, Barsyte-Lovejoy D, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. Cell. 2012;149(1):214-31.

21. Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact. 2012;11:56.

22. Akerstrom B, Nilson BH, Hoogenboom HR, Bjorck L. On the interaction between single chain Fv antibodies and bacterial immunoglobulin-binding proteins. J Immunol Methods. 1994;177(1-2):151-63.

23. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372(3):774-97.

24. Canul-Tec JC, Riano-Umbarila L, Rudino-Pinera E, Becerril B, Possani LD, Torres-Larios A. Structural basis of neutralization of the major toxic component from the scorpion Centruroides noxius Hoffmann by a human-derived single-chain antibody fragment. J Biol Chem. 2011;286(23):20892-900.

25. Schmid M, Ernst P, Honegger A, Suomalainen M, Zimmermann M, Braun L, et al. Adenoviral vector with shield and adapter increases tumor specificity and escapes liver and immune control. Nat Commun. 2018;9(1):450.

26. Jardine JG, Sok D, Julien JP, Briney B, Sarkar A, Liang CH, et al. Minimally Mutated HIV-1 Broadly Neutralizing Antibodies to Guide Reductionist Vaccine Design. PLoS Pathog. 2016;12(8):e1005815.

27. Honegger A, Spinelli S, Cambillau C, Pluckthun A. A mutation designed to alter crystal packing permits structural analysis of a tight-binding fluorescein-scFv complex. Protein Sci. 2005;14(10):2537-49.

28. Kalyoncu S, Hyun J, Pai JC, Johnson JL, Entzminger K, Jain A, et al. Effects of protein engineering and rational mutagenesis on crystal lattice of single chain antibody fragments. Proteins. 2014;82(9):1884-95.

29. Lai YT, Wang T, O'Dell S, Louder MK, Schon A, Cheung CSF, et al. Lattice engineering enables definition of molecular features allowing for potent small-molecule inhibition of HIV-1 entry. Nat Commun. 2019;10(1):47.

30. Cooper CDO, Marsden BD. N- and C-Terminal Truncations to Enhance Protein Solubility and Crystallization: Predicting Protein Domain Boundaries with Bioinformatics Tools. Methods Mol Biol. 2017;1586:11-31.

31. van Rosmalen M, Krom M, Merkx M. Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. Biochemistry. 2017;56(50):6565-74.

32. Kuhlman B, Jacobs T, Linskey T. Computational Design of Protein Linkers. Methods Mol Biol. 2016;1414:341-51.

33. Arndt KM, Muller KM, Pluckthun A. Factors influencing the dimer to monomer transition of an antibody single-chain Fv fragment. Biochemistry. 1998;37(37):12918-26.

34. Long NE, Sullivan BJ, Ding H, Doll S, Ryan MA, Hitchcock CL, et al. Linker engineering in anti-TAG-72 antibody fragments optimizes biophysical properties, serum half-life, and high-specificity tumor imaging. J Biol Chem. 2018;293(23):9030-40.

35. Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. Bioinformatics. 2016;32(2):298-300.

36. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Sci. 2007;16(8):1569-76.

37. Ravindran PP, Heroux A, Ye JD. Improvement of the crystallizability and expression of an RNA crystallization chaperone. Journal of Biochemistry. 2011;150(5):535-43.

38. Newman J, Egan D, Walter TS, Meged R, Berry I, Ben Jelloul M, et al. Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. Acta Crystallogr D Biol Crystallogr. 2005;61(Pt 10):1426-31.

39. Price WN, 2nd, Chen Y, Handelman SK, Neely H, Manor P, Karlin R, et al. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol. 2009;27(1):51-7.

40. Islam MM, Nakamura S, Noguchi K, Yohda M, Kidokoro S, Kuroda Y. Analysis and Control of Protein Crystallization Using Short Peptide Tags That Change Solubility without Affecting Structure, Thermal Stability, and Function. Crystal Growth & Design. 2015;15(6):2703-11.

1   41.   Goldenzweig A, Goldsmith M, Hill SE, Gertman O, Laurino P, Ashani Y, et al. Automated
2   Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. Mol Cell.
3   2016;63(2):337-46.
4   42.   Keates T, Cooper CD, Savitsky P, Allerston CK, Phillips C, Hammarstrom M, et al. Expressing
5   the human proteome for affinity proteomics: optimising expression of soluble protein domains and in
6   vivo biotinylation. N Biotechnol. 2012;29(5):515-25.
7   43.   Sall A, Walle M, Wingren C, Muller S, Nyman T, Vala A, et al. Generation and analyses of human
8   synthetic antibody libraries and their application for protein microarrays. Protein engineering, design
9   & selection : PEDS. 2016;29(10):427-37.
10   44.   Winter G, Lobley CM, Prince SM. Decision making in xia2. Acta Crystallogr D Biol Crystallogr.
11   2013;69(Pt 7):1260-73.
12   45.   McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser
13   crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-74.
14   46.   Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a
15   comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol
16   Crystallogr. 2010;66(Pt 2):213-21.
17   47.   Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol
18   Crystallogr. 2004;60(Pt 12 Pt 1):2126-32.
19   48.   Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More
20   and better reference data for improved all-atom structure validation. Protein Sci. 2018;27(1):293-315.
21   49.   McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: the CCP4mg
22   molecular-graphics software. Acta Crystallographica Section D-Structural Biology. 2011;67:386-94.
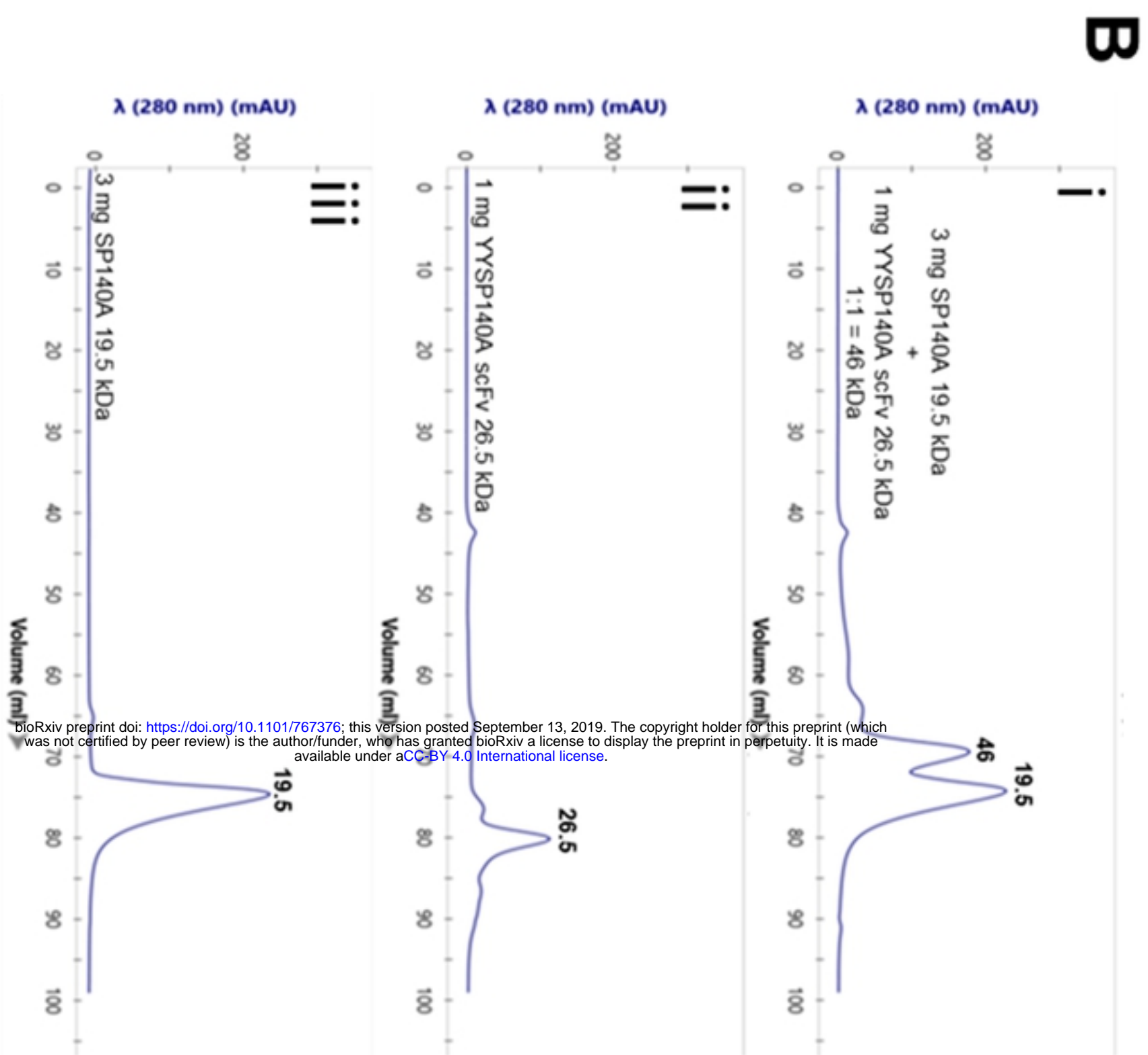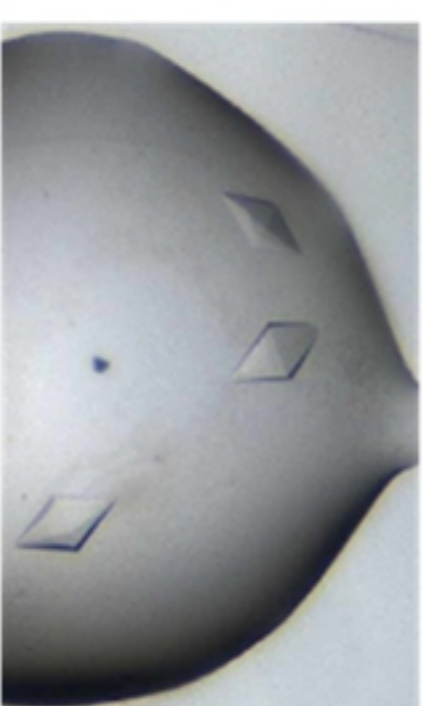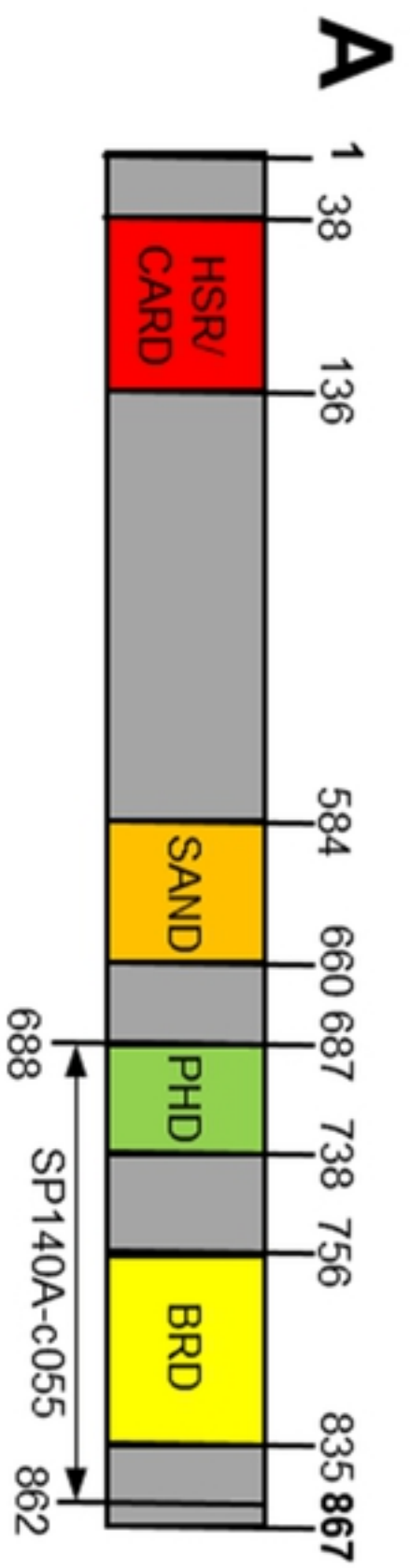
23
24

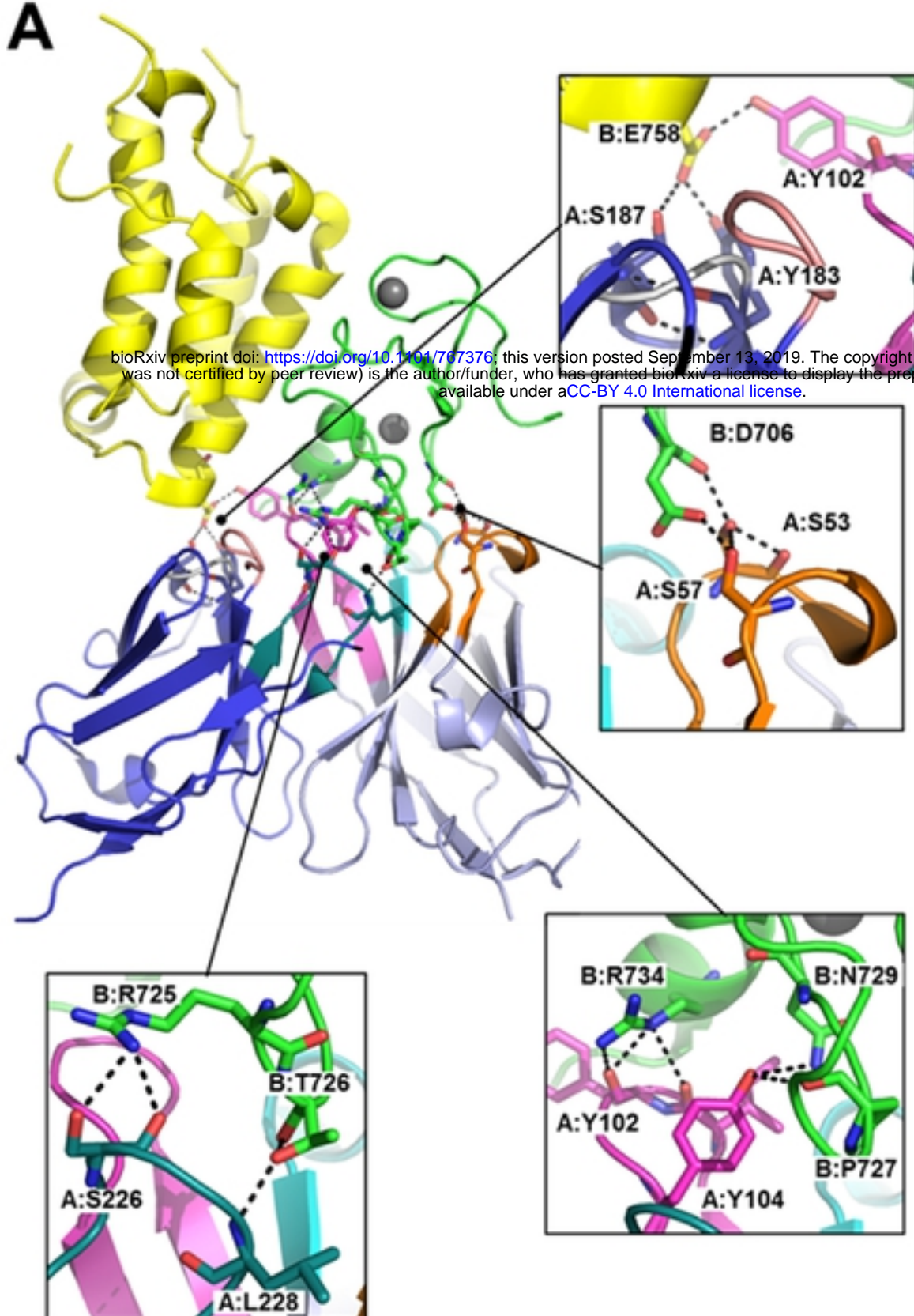**A**

1
38
HSR/CARD
136
584
SAND
660 687
688
PHD
738 756
BRD
835 867
862
SP140A-c055

B

i:
λ (280 nm) (mAU)
200
0
3 mg SP140A 19.5 kDa
+
1 mg YYSP140A scFv 26.5 kDa
1:1 = 46 kDa
Volume (ml)
0 10 20 30 40 50 60 70 80 90 100
46
19.5

ii:
λ (280 nm) (mAU)
200
0
1 mg YYSP140A scFv 26.5 kDa
Volume (ml)
0 10 20 30 40 50 60 70 80 90 100
26.5

iii:
λ (280 nm) (mAU)
200
0
3 mg SP140A 19.5 kDa
Volume (ml)
0 10 20 30 40 50 60 70 80 90 100
19.5

C

D

BRD
PHD

V_L
V_H

CDRL1
CDRL2
CDRL3

CDRH1
CDRH2
CDRH3

Figure1

Figure2

A Tag removal or minimization to make scFv more compact

B Linker modification to reduce conformational freedom

C Mutations to improve crystal packing

D Mutations to increase solubility and/or stability

V_H

V_L

Figure3