Supporting Information for "Contribution of Light Chain Residues to High Affinity Binding in an HIV-1 Antibody Explored by Combinatorial Scanning Mutagenesis"

Gustavo F. Da Silva, Joseph S. Harrison, and Jonathan R. Lai*

Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461

Mechanism of HIV-1 Viral Membrane Fusion. The HIV-1 resting state envelope spike consists of three molecules of glycoprotein gp41 (anchored to the virus), and three external molecules of glycoprotein gp120 (noncovalently associated).[S1] The trigger for viral entry is recognition of CD4 on the host cell by gp120. This event induces dramatic conformational rearrangements in gp120 that expose binding regions for the coreceptor (CXCR4 or CCR5). Next, gp120 dissociates from the envelope and gp41 alone mediates membrane fusion between the virus and host cells. Gp41 contains four domains: (i) A fusion peptide (FP); (ii) An ectodomain containing N- and C-terminal heptad repeats (NHR and CHR); (iii) A transmembrane domain (TM); (iv) A cytoplasmic/intraviral domain (C) (Figure S1A). The FP inserts into the target cell membrane, resulting in a conformational intermediate known as the extended intermediate. In this state, which has a half-life of ~ 15-30 mins, the NHR and CHR are exposed to the extracellular environment and can be sequestered with entry inhibitors.[S1-S3] The NHR segment of the extended intermediate is the proposed target for D5.[S4] Next, the ectodomain folds into a six-helix bundle consisting of three segments each of

NHR and CHR. This structure is known as the 'fusion' or 'trimer-of-hairpins' intermediate and has been extensively characterized.[S5] It consists of a core parallel trimer of NHR α-helices with three peripheral CHR α-helices that run antiparallel to this core (Figure S1B). Since NHR and CHR are antiparallel relative to one another, the six-helix fold brings the FP and TM domains (and therefore the virus and cell membranes) in close proximity. The energy gained from six-helix bundle folding (~30 kcal/mol) provides a strong driving force to the thermodynamic barrier associated with fusion of the two membranes.[S6] After fusion occurs, the 'post-fusion' gp41 state is left, in which both FP and TM are incorporated into the now fused membrane. An overview of this process is shown in Figure S1C.

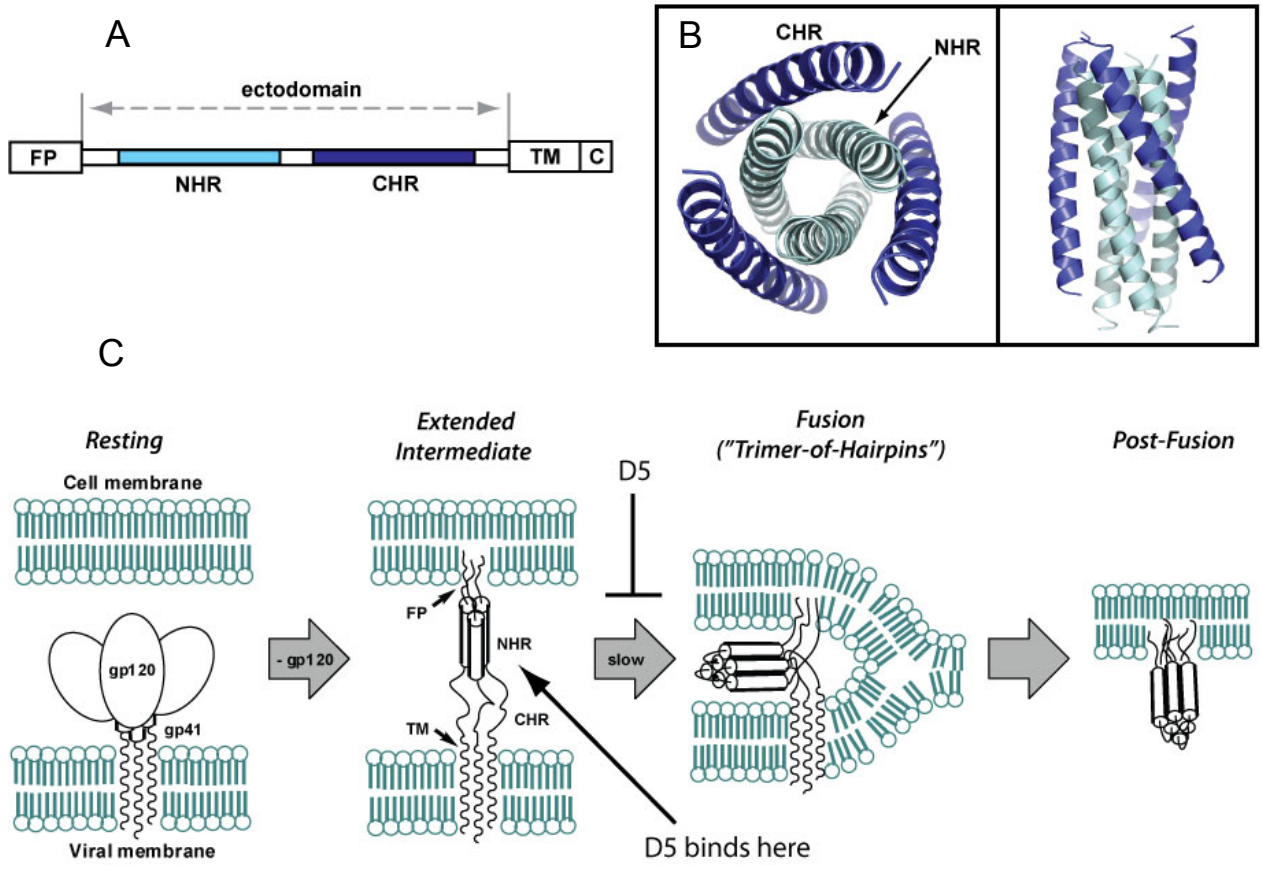

Figure S1 – HIV-1 gp41 and its role in viral membrane fusion. (A) Scheme of gp41 primary structure. The four domains are labeled (FP = fusion peptide, TM = transmembrane, C = cytoplasmic/intraviral). (B) Structure of the NHR/CHR six helix bundle, with the NHR and CHR color-coded according to panel A. Two orientations of the structure (~ 90° to one another) are shown. (C) Overview of viral membrane fusion.

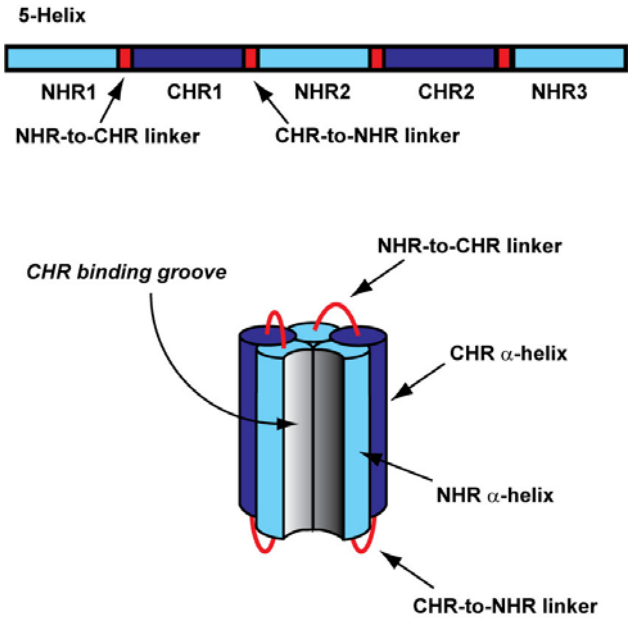

Figure S2 – Molecular design of '5-Helix'. The protein consists of alternating NHR and CHR segments (three NHRs, and two CHRs) joined by flexible linkers. When folded, the polypeptide adopts an α-helical conformation that display the CHR binding groove.

Protein Design Mimics the Extended Intermediate. 5-Helix is engineered to display the CHR-binding groove on the NHR core trimer (Figure S2) thereby mimicking aspects of the gp41 extended intermediate.[S2, S6] Suitable mimics of the NHR core require sophisticated protein engineering because NHR peptides are extremely hydrophobic and prone to aggregation. 5-Helix consists of three NHR segments, and two stabilizing CHR segments linked by flexible protein loops. When folded, this protein adopts a structure similar to the six-helix fusion state, but with one of the CHR helices displaced. This conformation results in display of the CHR-binding groove in a stable, soluble, globular protein format. Several versions of 5-Helix proteins have been reported.[S3,S7,S8] The crystal structure of D5 bound to 5-Helix was solved using a variant designed by Kim and coworkers (ref. S3). More recently, Harrison and coworkers designed a similar protein (ref. S7), also known as 'gp41-5'; we used the Harrison construct in our studies. The Kim and Harrison designs vary slightly in length of the α-helices and at the linker regions, but the residues that contact D5 are identical among these clones (Figure S3). For clarity, we refer to the Harrison construct as '5-Helix' here and throughout the main text.

NHR segment

```
Kim construct:       QLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILA
Harrison construct:     SGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARIL
```

NHR-to-CHR linker

```
Kim construct:       GGSGG
Harrison construct:  SGGRGG
```

CHR segment

```
Kim construct:       HTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLE
Harrison construct:     WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL
```

CHR-to-NHR linker

```
Kim construct:       GSSGG
Harrison construct:  GGKGGS
```

Figure S3 – Comparison of Kim and Harrison 5-Helix constructs. The sequences for each structural element are shown. The residues that contact D5 are highlighted.

Expression, Purification, and Characterization of 5-Helix. We expressed, purified, and refolded 5-Helix using established procedures.[S7] The refolded, purified 5-Helix protein was α-helical by circular dichroism (Figure S4A), and migrated as a single peak by gel filtration with an apparent molecular weight of ~20 kDa (expected monomer weight is 23.5 kDa, Figure S4B). In order to confirm that the purified, refolded 5-Helix was functional, we performed peak-shift gel filtration assay with synthetically prepared gp41 C-peptide. We incubated the 5-Helix with C-peptide at room temperature for 1 hr, and analyzed the mixture by gel filtration. We observed the 5-Helix-C-peptide complex, which migrated at an apparent molecular weigh of ~25 kDa (not shown).

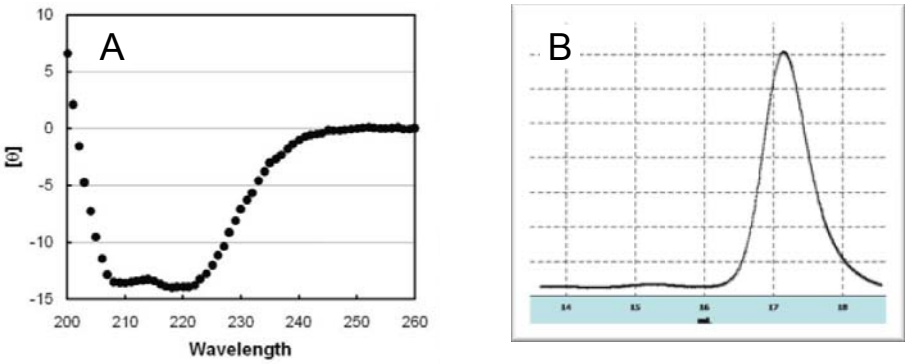

Figure S4 – Properties of 5-Helix. (A) Circular dichrosim of refolded 5-Helix shows a strong α-helical signature. (B) Size exclusion chromatography of purified, refolded 5-helix. A single peak was observed, of apparent molecular weight 20 kDa (expected molecular weight 23.5 kDa).

Functional Display of D5 scFv. A synthetic gene (codon optimized for *E. coli*) was obtained in which the V_H and V_L domains of D5 were linked by a $(GGGS)_3$ linker. (The DNA sequence of the D5scFv is shown on page S6.) This D5 scFv construct was cloned into the phagemid pAPIII6 (Figure S5A) using HindIII and SalI cloning sites to produce the phagemid pJH3.[S9] Cloning of the scFv into pAPIII6 in this way results in a chimeric protein containing an N-terminal FLAG epitope (for detection), the scFv, and the C-terminal 188 residues of the minor coat protein pIII all under *phoA* promoter control. We prepared phage particles displaying the D5 scFv monovalently (D5Φ) using standard methods [S9,S10] and performed Western Blot analysis of these phage particles by probing with a anti-FLAG/horse radish peroxidase (HRP) conjugate. As shown in Figure S2B, we observed a band corresponding to the scFv-pIII conjugate in D5Φ (~50 kDa) but

not in phage particles bearing no scFv (VCSM13). A polyreactive phage protein at 25 kDa was also observed. These results confirmed that D5Φ display the D5 scFv on their surface. To determine if the displayed D5 scFv was functional, we performed a phage enzyme-linked immunosorbent assay (ELISA) against purified 5-Helix. As shown in Figure S5C, we observed a strong binding signal for binding to 5-Helix but not BSA. Furthermore, control phage (VCSM13) lacking the D5 scFv did not exhibit binding to 5-Helix. These results indicate that the D5 scFv is expressed on phage and that the variable domains are functional in this format.

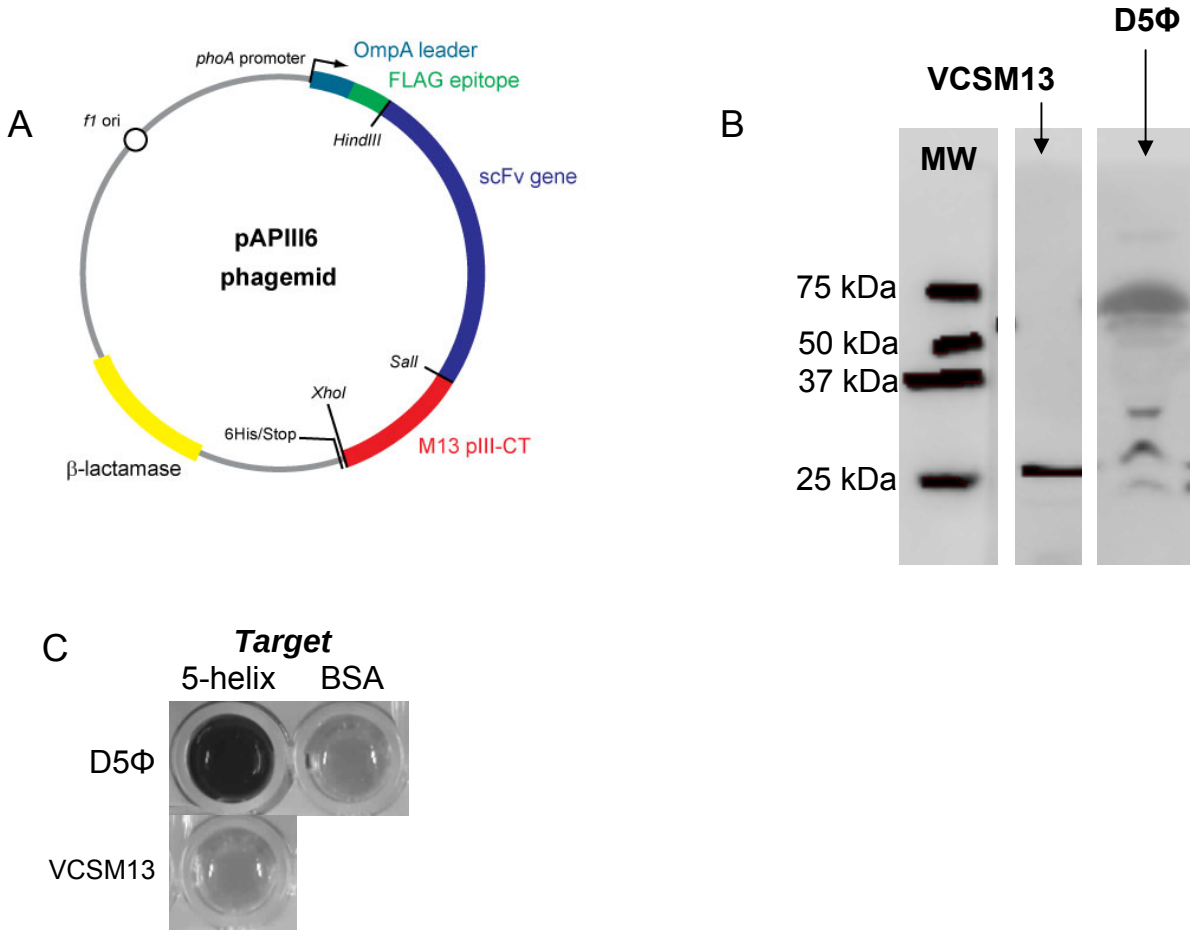

Figure S5 – pAPIII6 phagemid vector for monovalent display of D5 scFv. (A) Plasmid map of pAPIII6. A scFv gene sequence cloned with HindIII and SalI results in a chimeric gene with an N-terminal OmpA export sequence, a FLAG epitope (for detection), and a C-terminal pIII-CT fusion. (OmpA is cleaved during phage assembly.) (B) Western Blot analysis of control phage bearing no antibody (VCSM13) or phage displaying the D5 scFv (D5Φ) with anti-FLAG/HRP conjugate. The scFv-pIII-CT fusion protein (~50 kDa) is visible in D5Φ but not VCSM13. A polyreactive phage protein at 25 kDa was also observed. (C) Phage ELISA of D5Φ binding to 5-Helix or BSA target. A strong signal was observed for D5Φ to 5-Helix but not BSA. Control phage bearing no scFv (VCSM13) did not bind to 5-Helix.

Sequence of the D5 scFv (regions of VL that were randomized in the combinatorial scanning libraries highlighted in gray):

```
       HindIII
       ~~~~~~~
       LysLeuGlnVal GlnLeuVal GlnSerGly AlaGluValArg LysProGly AlaSerVal LysValSerCys LysAlaSer GlyAspThr PheSerSerTyr
1      AAGCTTCAGG TTCAACTGGT CCAGAGCGGT GCTGAGGTCC GCAAACCGGG CGCGAGCGTG AAAGTTTCTT GCAAAGCGAG CGGTGACACT TTCAGCTCTT
       TTCGAAGTCC AAGTTGACCA GGTCTCGCCA CGACTCCAGG CGTTTGGCCC GCGCTCGCAC TTTCAAAGAA CGTTTCGCTC GCCACTGTGA AAGTCGAGAA

       AlaIleSer TrpValArg GlnAlaProGly GlnGlyLeu GluTrpMet GlyGlyIleIle ProIlePhe GlyThrAla AsnTyrAlaGln AlaPheGln
101    ACGCGATTTC CTGGGTCCGC CAAGCCCCGG GTCAAGGCCT GGAATGGATG GGCGGTATTA TCCCGATCTT CGGCACCGCT AATTATGCGC AGGCGTTTCA
       TGCGCTAAAG GACCCAGGCG GTTCGGGGCC CAGTTCCGGA CCTTACCTAC CCGCCATAAT AGGGCTAGAA GCCGTGGCGA TTAATACGCG TCCGCAAAGT

        GlyArgVal ThrIleThrAla AsnGluSer ThrSerThr AlaTyrMetGlu LeuSerSer LeuArgSer GluAspThrAla IleTyrTyr CysAlaArg
201    GGGTCGTGTA ACCATCACCG CGAACGAGTC TACGTCTACC GCGTATATGG AGCTGTCTAG CCTGCGCTCC GAAGATACCG CTATCTACTA CTGCGCACGC
       CCCAGCACAT TGGTAGTGGC GCTTGCTCAG ATGCAGATGG CGCATATACC TCGACAGATC GGACGCGAGG CTTCTATGGC GATAGATGAT GACGCGTGCG

       AspAsnProThr LeuLeuGly SerAspTyr TrpGlyAlaGly ThrLeuVal ThrValSer SerGlyGlyGly GlySerGly GlyGlyGly SerGlyGlyGly
301    GATAACCCGA CCCTGCTGGG CTCTGACTAC TGGGGTGCCG GTACGCTGGT GACCGTATCC TCCGGAGGCG GTGGCAGCGG AGGCGGTGGA TCTGGCGGTG
       CTATTGGGCT GGGACGACCC GAGACTGATG ACCCCACGGC CATGCGACCA CTGGCATAGG AGGCCTCCGC CACCGTCGCC TCCGCCACCT AGACCGCCAC

         GlySerAsp IleGlnMet ThrGlnSerPro SerThrLeu SerAlaSer IleGlyAspArg ValThrIle ThrCysArg AlaSerGluGly IleTyrHis
401    GAGGAAGTGA TATTCAAATG ACCCAGTCTC CGTCTACCCT GTCTGCTTCT ATTGGCGACC GTGTAACTAT CACCTGCCGC GCGTCTGAAG GTATCTACCA
       CTCCTTCACT ATAAGTTTAC TGGGTCAGAG GCAGATGGGA CAGACGAAGA TAACCGCTGG CACATTGATA GTGGACGGCG CGCAGACTTC CATAGATGGT

       TrpLeuAla TrpTyrGlnGln LysProGly LysAlaPro LysLeuLeuIle TyrLysAla SerSerLeu AlaSerGlyAla ProSerArg PheSerGly
501    CTGGCTGGCT TGGTACCAGC AGAAACCAGG TAAAGCACCG AAACTGCTGA TCTATAAAGC GTCTTCTCTG GCATCCGGTG CTCCATCCCG TTTCTCCGGC
       GACCGACCGA ACCATGGTCG TCTTTGGTCC ATTTCGTGGC TTTGACGACT AGATATTTCG CAGAAGAGAC CGTAGGCCAC GAGGTAGGGC AAAGAGGCCG

       SerGlySerGly ThrAspPhe ThrLeuThr IleSerSerLeu GlnProAsp AspPheAla ThrTyrTyrCys GlnGlnTyr SerAsnTyr ProLeuThrPhe
601    TCCGGCTCTG GCACCGATTT TACCCTGACT ATCAGCTCCC TGCAACCAGA CGATTTTGCG ACCTACTACT GCCAGCAGTA TAGCAACTAC CCACTGACTT
       AGGCCGAGAC CGTGGCTAAA ATGGGACTGA TAGTCGAGGG ACGTTGGTCT GCTAAAACGC TGGATGATGA CGGTCGTCAT ATCGTTGATG GGTGACTGAA

                                        SalI
                                       ~~~~~~
         GlyGlyGly ThrLysLeu GluIleLysArg ValAsp
701    TTGGTGGCGG CACCAAGCTG GAGATCAAAC GCGTCGAC
       AACCACCGCC GTGGTTCGAC CTCTAGTTTG CGCAGCTG
```

Comparison of D5 V_L to Nearest Germline Progenitors. Figure S3 shows a comparison of the D5 V_L domain to the five most homologous germline segments (joined at the V/J junction). The hotspot residues Y30 is not present in the nearest germline progenitors (in all cases it is a threonine). This observation suggests that tyrosine at this position in D5 was specifically selected for function.

```
                                  30
D5-VL  DIQMTQSPSTLSASIGDRVTITCRASEGIYHWLAWYQQKPGKAPKLLIYKASSLASGAPSRFSGSGSGTDFTLTISSLQPDDFAT
Germ1  ..............V...........QS.TS.......................E.............E..............
Germ2  ..............V...........QS.TS.......................E.............E..............
Germ3  ..............V...........QS.TS...................D....E.............E..............
Germ4  ..............V...........QS.TS...................D....E.............E..............
Germ5  ..............V...........QS.TS...................A....Q........................E....

D5-VL  YYCQQYSNYPLTFGGGTKLEIK
Germ1  ......NS.S-.......V...
Germ2  ......NS.S-.......V...
Germ3  ......NS.S-..Q....V...
Germ4  ......NS.S-..Q....V...
Germ5  .....ANSFP-..Q....V...
```

Figure S3 – Alignment of D5 V domains to the five most homologous germline segments. A period indicates indentical residue identity, the positions assayed in the combinatorial scanning mutagenesis libraries are highlighted in gray. The hotspot residue Y30 of D5 is indicated.

Comparison of WT D5 and Y30A Phage Clones for Binding at Several Different 5-Helix Concentrations. Figure S4 shows the relative ELISA signal for phage display WT D5 scFv or Y30A scFv as a function of 5-Helix concentration. The disparity between WT D5 and Y30A ELISA signals decreases as the amount of 5-Helix loaded onto the well increases. At the highest amount of 5-Helix tested in this experiment (100 ng/well), the ELISA signals for WT and Y30A are within ~ 20% of one another. This result indicates that any display differences between WT D5 and Y30A scFvs are relatively minor. Similar trends were observed at several different phage concentrations.

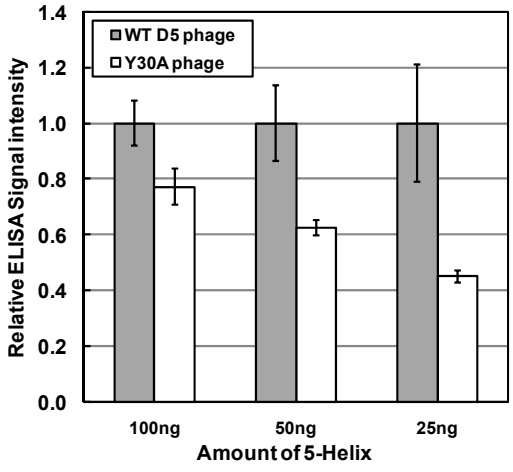

Figure S4 – ELISA of WT D5 and Y30A phage at various 5-Helix concentrations.

References

S1. (a) Harrison, S. C. *Nat. Struct. Mol. Biol.*, **2008**, *15*, 690-698.; (b) Eckert, D. M.; Kim, P. S. *Annu. Rev. Biochem.*, **2001**, *98*, 8502-8506.

S2. (a) Ferrer, M.; Kapoor, T. M.; Strassmaier, T.; Weissenhorn, W.; Skehel, J. J.; Oprian, D.; Schreiber, S. L.; Wiley, D. C.; Harrison, S. C. *Nat. Struct. Biol.*, **1999**, *6*, 953-960. (b) Eckert, D. M.; Malashkevich,

V. N.; Hong, L. H.; Carr, P. A.; Kim, P. S. *Cell*, **1999**, *99*, 103-115. (c) Eckert, D. M.; Kim, P. S. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 11187-11192.; (d) Sia, S. K.; Carr, P. A.; Cochran, A. G.; Malashkevich, V. N.; Kim, P. S. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 14664-14669.

S3. Root, M. J.; Kay, M. S.; Kim, P. S. *Science*, **2001**, *291*, 884-888.

S4. (a) Miller, M. D.; *et al. Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 14759-14764.; (b) Luftig, M. A.; *et al. Nat. Struct. Mol. Biol.*, **2006**, *13*, 740-747.

S5. (a) Chan, D. C.; Kim, P. S. *Cell*, **1997**, *89*, 263-273. (b) Weissenhorn, W.; Dessen, A.; Harrison, S. C.; Skehel, J. J.; Wiley, D. C. *Nature*¸**1997**, *387*, 426-430.

S6. Mari, D. N.; Bjeli, S. A.; Lu, M.; Bosshard, H. R.; Jelesarov, I. *J. Mol. Biol.*, **2004**, *336*¸1-6.

S7. Frey, G.; Rits-Volloch, S.; Zhang, X. Q.; Schooley, R. T.; Chen, B.; Harrison, S. C. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*¸13938-13943.

S8. Champagne, K.; Shishido, A.; Root, M. J. *J. Biol. Chem.*, **2009**, *284*, 3619-3627.

S9. Haidaris, C. G.; Malone, J.; Sherill, L. A.; Bliss, J. M.; Gaspari, A. A.; Insel, R. A.; Sullivan, M. A. *J. Immunol. Methods*, **2001**, *257*¸185-202.

S10. Sidhu, S. S.; Weiss, G. A. "Constructing phage display libraries by oligonucleotide-directed mutagenesis" in *Phage Display: A Practical Approach*. Clackson, T.; Lowman, H. B. Eds, Oxford University Press: New York, NY, 2004. Pages 27-41.

S11. Weiss, G. A.; Watanabe, C. K.; Zhong, A.; Goddard, A.; Sidhu, S. S. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*¸ 8950-8954.