SUPPLEMENTARY INFORMATION

A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain

John Wojcik,1 Oliver Hantschel,2 Florian Grebien,2 Ines Kaupe,2 Keiryn L. Bennett,2 John Barkinge,3 Richard B. Jones,3 Akiko Koide,1 Giulio Superti-Furga2 and Shohei Koide1.*

1Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA.

2Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 19, A-1090 Vienna, Austria.

3Ben May Department for Cancer Research and the Institute for Genomics and Systems Biology, The University of Chicago, 900 East 57th Street, Chicago, IL 60637, USA.
Supplementary Figure 1. Monobody phage display library development.

(a) Phage display construct schematic: a signal sequence directing either post translational (OmpT) or co-translational (DsbA) section is followed by the FN3 coding sequence, which is followed by a V5 epitope tag for display detection. The coding sequence for the phage coat protein p3 follows the V5 tag. The phage coat protein portion contained either the C-terminal domain (CTD) only, or the N1, N2 domains in addition to the CTD.

(b and c) Phage ELISA (enzyme-linked immunosorbent assay) data showing the levels of surface display of the monobody scaffold. ELISA signals are plotted as a function of the number of phages added to a microtiter plate coated with an anti-V5 antibody. In b, phages produced using the OmpT signal sequence (open boxes, dotted line) are compared to phages produced using the DsbA signal sequence (closed boxes, solid line). In c, phages from cultures grown in a baffled flask under fast shaking conditions (open boxes, dotted line) are compared with phages from cultures grown in a non-baffled flask under slow shaking conditions (closed boxes, solid line)
Supplementary Figure 2. TAP-MS analysis of HA4 binding specificity in cells.
(a) Schematic of the TAP-tag used in this study. Protein G indicates the B1 domain of Staphylococcal protein G, SBP indicates a streptavidin-binding peptide, and TEV indicates a tobacco etch virus protease recognition site.
(b) SDS-PAGE after silver staining showing recovered proteins from TAP purification from HEK293 and K562 cells expressing the HA4 monobody or a nonfunctional mutant, HA4\textsubscript{Y87A}. The left pair shows data for HEK293 cells and the right pair for the K562 cells. Bands for unambiguously identified proteins are labeled. The bands visualized in the HA4\textsubscript{Y87A} lanes are predominantly ‘background’ proteins commonly identified in similar TAP experiments. A complete list of identified proteins is provided as Supplementary Table 2.
(c) The UniProt codes for proteins detected in tandem mass spectrometry of tryptic digests of TAP eluates from the two cell lines shown in (b). The number of unique peptides identified for each protein is provided. Proteins containing an SH2 domain are shown in bold. Because of the presence of both Bcr-Abl and ABL1 in K562, ABL1 peptides could not be unambiguously assigned to either ABL1 or the Bcr-Abl fusion protein, and are grouped. For HEK293 cells, Abl2 and Abl2* denote two distinct sets of Abl2 isoforms.
Supplementary Table 1. Identity of SH2 domains spotted on the microarray

| row | column |
|-----|--------|
|     | 1      | 2      | 3      | 4      | 5      | 6      | 7      |
| 1   | TXN   | VAV2   | LCP2   | SH2B   | FES    | PIK3R3-C | PIK3R1-N-C |
| 2   | DAPP1 | GRAP2  | BRDG1  | ZAP70-N| PTPN11-C | TENS1   | CRKL    |
| 3   | ZAP70-N-C | LNK | FGR | MATK | HCK | TEC | E169291 |
| 4   | SH2D3A | E138606 | RASA1-C | E105251 | LCK | SHC1 | TXN    |
| 5   | RASA1-N | SH3BP2 | BLK | HSH2D | FER | TXK | BTK    |
| 6   | RASA1-N-C | VAV3 | PTPN11-N-C | PLCG1-C | GST-SRC | PTPN11-N | PTPN6-N-C |
| 7   | GST   | VAV1   | PIK3R3-N | SH2DIA | SHC-PTB | LYN   | CTEN   |

| row | column |
|-----|--------|
|     | 8      | 9      | 10     | 11     | 12     | 13     | 14     |
| 1   | GRB10 | PIK3R2-N | PLCG1-N-C | PIK3R2-C | TNS | SOCS3 | GST    |
| 2   | E18941| SYK-N | E109111 | SH2D2A | GRB7 | NCK1 | SOCS2  |
| 3   | PIK3R2-N-C | BMX | SLA2 | SCK-PTB | PLCG2-N-C | ZAP70-C | ABL2   |
| 4   | GST   | PTK   | YES1   | GRB14 | INPPL1 | ITK  | BLNK  |
| 5   | SOCS6 | PLCG1-N | CRK   | GRB2  | PTPN6-C | BCAR3 | SYK-N-C |
| 6   | SRC   | PIK3R1-N | ABL1  | EAT2  | SHB  | TENC1 | DAPP1  |
| 7   | PIK3R3-N-C | PIK3R1-C | SH2D3C | NCK2  | SHC3 | SYK-C | TXN    |

Each protein was spotted as duplicates. All proteins were produced as thioredoxin (TXN) fusion.
| Accession | Organism | GeneName | Synonym | GeneSymbol | % Sequence Coverage | #Spectra | #Proteins | % Peptides | #Peptides |
|-----------|----------|----------|---------|------------|-------------------|---------|----------|-----------|---------|
| IPI00013933.2 | Homo sapiens | MAP4 | Isoform 6 of Microtubule-associated protein 4 | MAP4 | 32.91 | 29 | 242 | 10 | 10 |
| IPI00013933.2 | Homo sapiens | HSPA8 | | HSPA8 | 27.47 | 5 | 292 | 10 | 10 |
| IPI00056041.5 | Homo sapiens | HNRPU | | HNRPU | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | MAP4 | Isoform 1 of Microtubule-associated protein 4 | MAP4 | 32.91 | 29 | 242 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HSPA8 | | HSPA8 | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HNRPU | | HNRPU | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | MAP4 | Isoform 6 of Microtubule-associated protein 4 | MAP4 | 32.91 | 29 | 242 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HSPA8 | | HSPA8 | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HNRPU | | HNRPU | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | MAP4 | Isoform 1 of Microtubule-associated protein 4 | MAP4 | 32.91 | 29 | 242 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HSPA8 | | HSPA8 | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HNRPU | | HNRPU | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | MAP4 | Isoform 6 of Microtubule-associated protein 4 | MAP4 | 32.91 | 29 | 242 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HSPA8 | | HSPA8 | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HNRPU | | HNRPU | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | MAP4 | Isoform 1 of Microtubule-associated protein 4 | MAP4 | 32.91 | 29 | 242 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HSPA8 | | HSPA8 | 27.47 | 5 | 292 | 10 | 10 |
| IPI00003865.1 | Homo sapiens | HNRPU | | HNRPU | 27.47 | 5 | 292 | 10 | 10 |

Supplementary Table 2. Proteins identified in tandem-affinity-purification, mass spectrometry analysis of cells transfected with TAP-tag-HA4/H4A<sub>exps</sub> fusions.

Proteins are grouped by cell line (K562/HEK293) and transfected construct (TAP-HA4/TAP-HA4/Y7TA).

*% sequence coverage* refers to the fraction of the total protein sequence that is represented by identified peptides. *% Proteins* refers to the number of different protein isoforms that are consistent with the identified peptides. *% Peptides* indicates the number of unique peptides identified for each protein. "#Peptides" indicates the number of mass spectra in which a fragment of each protein is observed.

HEK 293 HA4<sub>exps</sub>
| Gene ID | Description | Accession Number | Function |
|--------|-------------|-----------------|----------|
| IPI0019912.1 | Homo sapiens | | Perinuclear multifunctional enzyme type 2 |
| IPI0005755.1 | DSG1 | DSG1 | Desmosome-1 precursor |
| IPI0043711.1 | JPH2 | JPH2 | JPH2 precursor |
| IPI0013214.2 | Homo sapiens | | MCM10 ORF35L559, highly similar to DNA replication licensing factor MCM1 |
| IPI0028925.5 | HRX | HRX | HRX precursor |
| IPI0035775.6 | PARP10 | PARP10 | PARP10 isoform 1 |
| IPI0072974.1 | DDX1 | DDX1 | ATP-dependent RNA helicase |
| IPI0016454.2 | Tubulin beta chain | TUBB | Tubulin beta chain |
| IPI0003955.3 | Homo sapiens | | DNA replication licensing factor MCM |
| IPI0047280.7 | Homo sapiens | | Colonizing and hepatic tumor over-expressed protein isoform b |
| IPI0034049.4 | Homo sapiens | | Poly (ADP-ribose) polymerase |
| IPI0024924.3 | Homo sapiens | | Structural maintenance of chromosome protein |
| IPI0034599.4 | Homo sapiens | | Non-POD domain-containing octamer-binding protein 1 |
| IPI0010720.5 | Homo sapiens | | EZR isoform 2 of EZR/RAM-interacting/CAST family member 1 |
| IPI0028243.1 | Homo sapiens | | GRIA1 structural maintenance of chromosome protein |
| IPI0048024.2 | Homo sapiens | | JUP ORF35L4642, highly similar to Jund phospholysin |
| IPI0006843.3 | Homo sapiens | | CBL isoform 1 |
| IPI0003020.4 | Homo sapiens | | Probable ATP-dependent RNA helicase |
| IPI0000015.1 | Homo sapiens | | ABL2 isoform 2 of Abl-2 tyrosine-protein kinase |
| IPI0038849.2 | Homo sapiens | | HSPA5 isoform 2 of Abl-2 tyrosine-protein kinase |
| IPI0010503.1 | Homo sapiens | | C22orf28 isoform 2 of C22orf28 |
| IPI0008752.9 | Homo sapiens | | EIF5A2 isoeform 2 of Eif5a2 |
| IPI0049879.1 | Homo sapiens | | V63 isoform 1 of Vps13 |
| IPI0021354.1 | Homo sapiens | | Isoform Short of RNA-binding protein PUS |
| IPI0027584.1 | Homo sapiens | | DC9 Dermaid precursor |

**K562 HA4**

| Gene ID | Accession | Symbol | Description | Coverage | 1% Fasta Index | Feature |
|--------|-----------|--------|-------------|-----------|---------------|---------|
| IPI00019912.1 | Homo sapiens | | Perinuclear multifunctional enzyme type 2 | EDU1784 | 11.83 | 7 20 |
| IPI0005755.1 | Homo sapiens | | Desmosome-1 precursor | DSG1 | 5.84 | 1 6 14 |
| IPI0043711.1 | Homo sapiens | | JPH2 precursor | JPH2 | 6.98 | 1 5 13 |
| IPI0013214.2 | Homo sapiens | | MCM10 ORF35L559, highly similar to DNA replication licensing factor MCM1 | MCM10 | 5.28 | 1 5 9 |
| IPI0028925.5 | Homo sapiens | | HRX precursor | HRX | 8.15 | 1 5 40 |
| IPI0035775.6 | Homo sapiens | | PARP10 isoform 1 | PARP10 | 14.25 | 2 7 50 |
| IPI0072974.1 | Homo sapiens | | ATP-dependent RNA helicase | DDX1 | 12.77 | 2 7 51 |
| IPI0016454.2 | Homo sapiens | | Tubulin beta chain | TUBB | 9.91 | 10 4 6 |
| IPI0003955.3 | Homo sapiens | | DNA replication licensing factor MCM | MCM | 4.05 | 3 5 45 |
| IPI0047280.7 | Homo sapiens | | Colonizing and hepatic tumor over-expressed protein isoform b | JUP | 2.69 | 4 4 4 |
| IPI0034049.4 | Homo sapiens | | Poly (ADP-ribose) polymerase | POLR3C | 5.14 | 1 1 30 |
| IPI0024924.3 | Homo sapiens | | Structural maintenance of chromosome protein | SCC2 | 3.61 | 1 4 8 |
| IPI0034599.4 | Homo sapiens | | Non-POD domain-containing octamer-binding protein 1 | PODC8 | 9.56 | 5 4 8 |
| IPI0010720.5 | Homo sapiens | | EZR isoform 2 of EZR/RAM-interacting/CAST family member 1 | EZR | 9.57 | 5 7 8 |
| IPI0028243.1 | Homo sapiens | | GRIA1 structural maintenance of chromosome protein | GRIA1 | 2.21 | 1 1 3 |
| IPI0048024.2 | Homo sapiens | | JUP ORF35L4642, highly similar to Jund phospholysin | JUP | 5.23 | 1 3 130 |
| IPI0006843.3 | Homo sapiens | | CBL isoform 1 | CBL | 3.44 | 1 1 9 |
| IPI0003020.4 | Homo sapiens | | Probable ATP-dependent RNA helicase | DDX1 | 0.95 | 1 1 2 |
| IPI0000015.1 | Homo sapiens | | ABL2 isoform 2 of Abl-2 tyrosine-protein kinase | ABL2 | 5.76 | 4 3 5 |
| IPI0010503.1 | Homo sapiens | | C22orf28 isoform 2 of C22orf28 | C22orf28 | 3.95 | 1 3 3 |
| IPI0008752.9 | Homo sapiens | | EIF5A2 isoeform 2 of Eif5a2 | EIF5A2 | 6.99 | 1 2 3 |
| IPI0049879.1 | Homo sapiens | | V63 isoform 1 of Vps13 | V63 | 2.16 | 3 2 3 |
| IPI0021354.1 | Homo sapiens | | Isoform Short of RNA-binding protein PUS | PUS | 4.39 | 9 2 2 |
| IPI0027584.1 | Homo sapiens | | DC9 Dermaid precursor | DC9 | 20.2 | 2 2 23 |

**Accession**

- **Accession**
- **Symbol**
- **Description**
- **Coverage**
- **1% Fasta Index**
- **Feature**
| Accession#  | Organism     | GeneName                                                                 | Synonym                          | GeneSymbol | % Sequence Coverage | # Proteins | # Peptides | # Spectra |
|------------|--------------|--------------------------------------------------------------------------|----------------------------------|------------|---------------------|------------|------------|-----------|
| IPI0003866.1 | Homo sapiens | Isoform 1 of Heat shock cognate 71 kDa protein HSP70; HSPA10             | HSP70; HSPA10                    | HSPA10     | 37.31               | 9          | 22         | 118       |
| IPI0003766.1 | Homo sapiens | Stress-70 protein, mitochondrial precursor HSP   | HSP; HSPA2                      | HSPA2      | 10.86               | 2          | 12         | 75        |
| IPI0003942.1 | Homo sapiens | Heat shock cognate 71 kDa protein ISHAP1; JBA10; HSPA10                   | ISHAP1; JBA10; HSPA10; JBA20     | HSPA10     | 35.29               | 8          | 10         | 52        |
| IPI0003920.1 | Homo sapiens | HSAP8 protein               | HSAP8                           | HSAP8      | 22.18               | 1          | 10         | 30        |
| IPI0038601.1 | Homo sapiens | Acetyl-CoA synthetase isoform 4; Acetyl-CoA synthetase isoform 2           | ACSA; ACSB                      | ACSA       | 4.61                | 5          | 8          | 17        |
| IPI0038127.1 | Homo sapiens | Heat shock 70 kDa protein 11.1                                     | HSAP1A; HSAP1                     | HSAP1A     | 11.24               | 5          | 8          | 45        |
| IPI0030459.1 | Homo sapiens | Neutrophil-specific antithrombin-binding protein HSAP4                    | HSAP4                           | HSAP4      | 22.09               | 6          | 8          | 15        |
| IPI0030577.6 | Homo sapiens | NCBP2 protein 1 beta isoform                                          | NCBP2                           | NCBP2      | 15.70               | 2          | 6          | 24        |
| IPI0033263.1 | Homo sapiens | Heat shock 70 kDa protein 6                                               | HSP70B                          | HSP70B     | 7.18                | 3          | 5          | 40        |
| IPI0000912.3 | Homo sapiens | Peroxisomal multifunctional enzyme type 2                                | PEX17; PEX18; PEX19; PEX20; PEX21 | PEX17; PEX18; PEX19; PEX20; PEX21 | 6.25 | 2 | 4 | 8 |
| IPI0022135.4 | Homo sapiens | Elongation factor 1-alpha 1                                              | EF1a; ELAV7                      | EF1a; ELAV7 | 6.01               | 5          | 3          | 8         |
| IPI0000934.2 | Homo sapiens | Ras GTPase-activating-like protein IQGAP1                                 | KIAA0051                        | IQGAP1     | 7.25                | 1          | 3          | 11        |
| IPI0030492.5 | Homo sapiens | HSPA1B; HSPA1A Heat shock 70 kDa protein 1                               | HSPA1A; HSPA1B                   | HSPA1A; HSPA1B | 15.29 | 8 | 10 | 52 |
| IPI0000336.2 | Homo sapiens | HSPA5 protein                                                           | GRP78                           | HSPA5      | 22.14               | 1          | 10         | 30        |
| IPI0000076.2 | Homo sapiens | GRP78 protein                                                            | GRP78                           | GRP78      | 22.14               | 1          | 10         | 30        |
| IPI0030127.1 | Homo sapiens | Heat shock 70 kDa protein 11.1                                     | HSAP1A; HSAP1                     | HSAP1A     | 11.24               | 5          | 8          | 45        |
| IPI0001074.0 | Homo sapiens | Splicing factor, proline- and glutamine-rich PSF                         | PSF1; PSF2; PSF3; PSF4; PSF5; PSF6 | PSF1; PSF2; PSF3; PSF4; PSF5; PSF6 | 4.61 | 5 | 8 | 17 |
| IPI0000934.2 | Homo sapiens | Ras GTPase-activating-like protein IQGAP1                                 | KIAA0051                        | IQGAP1     | 7.25                | 1          | 3          | 11        |
| IPI0000934.2 | Homo sapiens | Ras GTPase-activating-like protein IQGAP1                                 | KIAA0051                        | IQGAP1     | 7.25                | 1          | 3          | 11        |
Supplementary Table 3. Comparison of interface characteristics of HA4/Abl1 SH2, peptide/SH2 and antibody/antigen complexes.

| Interface Characteristic     | HA4/Abl1 SH2 | Peptide/Lck SH2<sup>a</sup> | Antibody/antigen<sup>b</sup> |
|------------------------------|--------------|-----------------------------|-----------------------------|
| Interface size ASA (Å<sup>2</sup>) | 650          | 428                         | 801                         |
| Shape complementarity (S<sub>c</sub>)<sup>c</sup> | 0.72         | 0.79                        | 0.64–0.68<sup>c</sup>       |
| Hydrogen bonds per 100 Å<sup>2</sup> | 2.0          | 4.0                         | 1.2                         |
| Salt bridges per 100 Å<sup>2</sup> | 3.1          | 0.9                         | 3.2                         |
| % Nonpolar atoms             | 29           | 17                          | 40                          |

<sup>a</sup>From PDB entry, 1LCJ.<sup>1</sup>  <sup>b</sup>From protein/protein complexes in BEID database.<sup>2</sup>  <sup>c</sup>From<sup>3</sup>.
Supplementary Data
Enhanced phage-display system for FN3 monobodies

While a number of FN3 monobodies have been successfully selected using phage display, technical advances prompted us to reevaluate our phage-display system. Key factors in successful monobody selection are the level of FN3 displayed on the phage surface and library design. We therefore examined and improved these parameters. FN3 is anchored to the phage surface through fusion to coat protein p3. Ideally, each phage particle in the library should display one FN3 molecule on the surface, because low display efficiency would lead to poor sampling of the repertoire and selection of clones due to factors other than the function of the displayed molecule. High display levels require not only efficient expression of a FN3-p3 fusion protein but also efficient translocation of this protein across the E. coli inner membrane so that it can be incorporated into the phage. Recent work has revealed that the translocation step of a highly stable protein into the periplasm using the conventional, post-translational signal sequence is very inefficient, leading to poor surface display. This inefficient translocation was ameliorated through the use of a signal recognition particle (SRP)-dependent, co-translational, signal sequence. Because FN3 is also a highly stable protein, we adopted SRP phage-display for surface presentation of FN3. Here, we performed a quantitative comparison of the level of FN3 surface display achieved using a conventional, posttranslational OmpT signal sequence with that achieved using the co-translational signal sequence from DsbA. Surface display of FN3, as detected using an epitope tag located between FN3 and the phage coat protein p3 (Supplemental Figure 1A), was over 100 times greater with the DsbA sequence than with the OmpT sequence (Supplemental Figure 1B), clearly indicating the superiority of the co-translational secretion signal for FN3 display. This magnitude of increase in display level was similar to that observed for a DARPin (700 fold). We also found that the FN3 display level was strongly affected by culture conditions. A systematic test revealed that bacterial cultures grown in non-baffled flasks with fast shaking, a standard method for phage preparation (Supplementary Figure 1C). This increase in the display level was accompanied by a decrease in the final density of the E. coli culture, and a moderate increase (~2x) in phage yield. These observations suggest that the improvement in FN3 display was a consequence of the host cells devoting fewer resources to cell growth and more resources to phage production and fusion protein expression.

In addition, we have compared the effects of using different p3 fragments as the fusion partner. The p3 protein has three domains, N1, N2 and C, named for their location in the protein (Supplemental Figure 1A). In “phagemid”-based systems the protein of interest is usually fused to a p3 fragment that includes just the C domain, because the N1 and N2 domains are not necessary for surface display. However, the use of the full-length protein (except for the secretion signal) would give the flexibility of controlling the valency (monovalent versus multivalent) of a p3 fusion using different types of helper phages. We found that both types of FN3 phage-display vectors gave similar levels of surface display (data not shown). Therefore, we have chosen to use full-length p3 so that we retain the ability to control the valency. Combining all these modifications, our new system improved the FN3 display level over 1,000 fold relative to our previous methods.

Using the new phage display platform, we have constructed a combinatorial library that employed a biased-amino acid diversity strategy. The use of biased amino acid compositions in recognition loops has been shown to be highly effective in producing high-performance combinatorial libraries of antibodies and FN3 monobodies. Such ‘biased code’ libraries increase the efficiency of binding protein selection by enriching the interface with amino acids likely to mediate favorable interactions (such as Tyr, Ser and Gly), while limiting those likely to be detrimental to binding (such as Lys). This strategy is particularly useful in cases where amino acid sequence space vastly outstrips the size of library that is technically achievable.

In this work, we used a biased amino acid composition that included Tyr (30%), Ser (15%), Gly (10%), Trp (5%),...
stranded DNA fragments containing either BC or FG loop was produced following published protocols. In our design, we also permitted the lengths of the diversified loops to vary, as previous studies of antibodies and FN3 monobodies demonstrated the effectiveness of loop length variation in generating high-affinity interfaces through greater conformational diversity. The final library contained \( \sim 5 \times 10^9 \) independent clones, and sequencing of randomly chosen clones confirmed the designed amino acid composition (data not shown).

Supplementary Methods

Phage-display selection of monobodies to the Abl SH2 domain. Abl SH2 was produced as a fusion protein to an engineered glutathione S-transferase (GST). A general use GST vector was first constructed as follows. The GST gene from the pGEX2 vector was cloned into the pHFT2 vector\(^7\) that produces GST as an N-terminal His\(_{10}\) fusion protein under the control of the T7 promoter. All of the cysteine residues in GST were then changed to serine via site-directed mutagenesis. The Abl SH2 gene was cloned into this vector in such a way to fuse the SH2 domain to the C-terminus of His\(_{10}\)-GST. A single cysteine was then introduced in a linker region between the C-terminus of GST and the N-terminus of the Abl SH2 domain via site-directed mutagenesis. The fusion protein was expressed and purified using Ni affinity chromatography as described previously.\(^7\) The purified protein was then chemically biotinylated at the single cysteine site using (N-(6-(Biotinamido)hexyl)-3'-(2'-pyridyldithio)-propionamide (EZ-Link Biotin-HPDP, Thermo Scientific).

Two rounds of library selection were carried out magnetic bead-based protocols described previously.\(^12\) Target concentrations of 100 and 20 nM were used for the first and second rounds, respectively. After the second round, "loop shuffling" was performed as follows. Single-stranded, uracil-containing DNA for the enriched phage pool was produced following published protocols.\(^14\),\(^15\) Single-stranded DNA fragments containing either BC or FG loop sequences of enriched phages were produced as follows. DNA fragments corresponding to the BC and FG loops regions of the enriched phages were amplified separately using KOD DNA polymerase (Novagen). Then run-off PCR was used to enrich single-strand DNA fragments for the sense strand. These single-stranded DNA fragments were phosphorylated using polynucleotide kinase (New England Biolabs) and used as mutagenic oligonucleotides in Kunkel mutagenesis reactions with the template described above. The product of the mutagenesis reaction was used to transform \( E. \ coli \) SS320,\(^17\) and the resulting library contained \( \sim 4 \times 10^7 \) clones. This library was subjected to two rounds of selection with the biotinylated Abl SH2 protein at 10nM and 0.5nM for the first and second rounds, respectively. Individual binding clones were identified and confirmed via phage ELISA as described elsewhere.\(^18\)

Crystallization, data collection and structure determination. The HA4/Abl SH2 domain complex was purified with a Superdex 75 column (GE Lifesciences). The complex was concentrated to \( \sim 10 \text{mg/ml} \) and crystallized in 0.1M Tris HCl buffer, pH 8.5 containing 0.2M sodium acetate and 30\%w/v Polyethylene glycol 4,000 by the hanging-drop vapor-diffusion method. Glycerol (20\%) was used as a cryoprotectant. X-ray diffraction data were collected at the Advanced Photon Source beamline 21 ID-D (Argonne National Laboratory) at a wavelength of 0.97872 Å and a temperature of 100 K. Data collection and structure determination statistics are given in Table 1. Diffraction data were processed and scaled with the HKL2000 package.\(^19\) The structures were solved by molecular replacement using the MOLREP program in the CCP4 program suite.\(^20\),\(^21\) A multicopy search was performed with the Abl SH2 domain and the FN3 scaffold, without the loop regions, as the search models (PDB IDs 2ABL and 1FNA, respectively). Rigid-body refinement was carried out using REFMAC5 in the CCP4 program suite. TLS (translation/libration/screw) groups were defined using the TLSMD server,\(^22\) and TLS refinement, B-factor refinement, bulk solvent parameters, final positional refinement, and the search for and
refinement of water molecules was carried out using REFMAC5. Model building and evaluation were carried out using the Coot program, and molecular graphics were generated using PyMOL (DeLano Scientific). The final structure had 100% of residues within allowed Ramachandran regions, and 98.7% in favored regions as measured by MOLPROBITY. 23 Surface area calculations were performed using the PROTORP protein–protein interaction server. 24

HEK293 transfection. The composition of the IP buffer is 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% (v/v) NP-40, 5 mM ethylenediaminetetraacetic acid (EDTA), 5 mM ethyleneglycoltetraacetic acid (EGTA), 25 mM NaF, 1 mM orthovanadate, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 µg ml⁻¹ Tosyl-phenylalanine chloromethyl ketone (TPCK), 5 µg ml⁻¹ tosyl-lysine chloromethyl ketone (TLCK), 1 µg ml⁻¹ leupeptin, 1 µg ml⁻¹ aprotinin, and 10 µg ml⁻¹ soybean trypsin inhibitor.

References for Supplementary Information

1. Eck, M.J., Shoelson, S.E. & Harrison, S.C. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* **362**, 87-91 (1993).

2. Tong, J.C., Tan, P.T.J., Ren, E.C. & Sinha, A.A. BEID: database for sequence-structure-function information on antigen-antibody interactions. *Bioinformatics* **3**(2008).

3. Lawrence, M.C. & Colman, P.M. Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**, 946-50 (1993).

4. Smith, G.P. & Petrenko, V.A. Phage Display. *Chem Rev* **97**, 391-410 (1997).

5. Steiner, D., Forrer, P. & Pluckthun, A. Efficient selection of DARPinS with sub-nanomolar affinities using SRP phage display. *J Mol Biol* **382**, 1211-27 (2008).

6. Steiner, D., Forrer, P., Stumpp, M.T. & Pluckthun, A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat Biotechnol* **24**, 823-31 (2006).

7. Koide, A., Gilbreth, R.N., Esaki, K., Tereshko, V. & Koide, S. High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci U S A* **104**, 6632-7 (2007).

8. Clackson, T.a.L., H.B. (ed.) *Phage Display: A Practical Approach*, 332 (Oxford University Press, New York, 2004).

9. Karlsson, F., Borrebaeck, C.A., Nilsson, N. & Malmborg-Hager, A.C. The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains. *J Bacteriol* **185**, 2628-34 (2003).

10. Rondot, S., Koch, J., Bretling, F. & Dubel, S. A helper phage to improve single-chain antibody presentation in phage display. *Nat Biotechnol* **19**, 75-8 (2001).

11. Fellouse, F.A., Wiesmann, C. & Sidhu, S.S. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci U S A* **101**, 12467-72 (2004).

12. Gilbreth, R.N., Esaki, K., Koide, A., Sidhu, S.S. & Koide, S. A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. *J Mol Biol* **381**, 407-18 (2008).

13. Fellouse, F.A. et al. High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol* **373**, 924-40 (2007).

14. Rock, E.P., Sibbald, P.R., Davis, M.M. & Chien, Y.H. CDR3 length in antigen-specific immune receptors. *J Exp Med* **179**, 323-8 (1994).

15. Barrios, Y., Jirholt, P. & Ohlin, M. Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor. *J Mol Recognit* **17**, 332-8 (2004).

16. Huang, J., Koide, A., Nettle, K.W., Greene, G.L. & Koide, S. Conformation-specific affinity purification of proteins using engineered binding proteins: application to the estrogen receptor. *Protein Exp Purif* **47**, 348-54 (2006).

17. Sidhu, S.S., Lowman, H.B., Cunningham, B.C. & Wells, J.A. Phage display for selection of novel binding peptides. *Methods Enzymol* **328**, 333-63 (2000).

18. Koide, A. & Koide, S. Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. *Methods Mol Biol* **352**, 95-109 (2007).

19. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-26 (1997).

20. Collaborative Computational Project. The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* **D50**, 760-763 (1994).

21. Potterson, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **59**, 1131-7 (2003).

22. Painter, J. & Merritt, E.A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-50 (2006).

23. Davis, I.W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83 (2007).

24. Reynolds, C., Damerell, D. & Jones, S. ProtorP: a protein-protein interaction analysis server. *Bioinformatics* **25**, 413-4 (2009).