HHS Public Access

Author manuscript

*Nat Struct Mol Biol*. Author manuscript; available in PMC 2010 October 01.

Published in final edited form as:

*Nat Struct Mol Biol*. 2010 April ; 17(4): 519–527. doi:10.1038/nsmb.1793.

# 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-Furga**[2], and **Shohei Koide**[1,*]

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

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

[3]Ben 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

## Abstract

Interactions between SH2 domains and phosphotyrosine sites regulate tyrosine kinase signaling networks. Selective perturbation of these interactions is challenging due to the high homology among the 120 human SH2 domains. Using an improved phage-display selection system, we generated a small antibody-mimic or 'monobody', termed HA4, that bound to the Abl kinase SH2 domain with low nanomolar affinity. SH2 protein microarray analysis and mass spectrometry of intracellular HA4 interactors demonstrated HA4's exquisite specificity, and a crystal structure revealed how this specificity is achieved. HA4 disrupted intramolecular interactions of Abl involving the SH2 domain and potently activated the kinase *in vitro*. Within cells, HA4 inhibited processive phosphorylation activity of Abl and also STAT5 activation. This work provides a design guideline for highly specific and potent inhibitors of a protein interaction domain and demonstrates their utility in mechanistic and cellular investigations.

## Introduction

Protein interaction domains (PIDs) are molecular recognition modules that mediate interactions with a broad array of biological ligands.[1] PIDs play central roles in the spatiotemporal activation of individual proteins and in the organization of cellular protein networks responsible for complex biological processes. Because signaling proteins often



*Corresponding author. skoide@uchicago.edu.

**Accession codes.** The coordinates and structure factors for the HA4/Abl1 SH2 domain complex have been deposited in the PDB with the accession code 3K2M.

Author Contributions: J.W. A.K. and S.K. designed phage-display library, selection and biophysical characterization; A.K. and J.W. optimized phage display methods. J.W. performed selections, biophysical characterization, crystallization and structure determination; J.B. and R.B.J. designed and made SH2 microarrays; J.W. and J.B. conducted microarray experiments; J.W. F.G., O.H. and G.S.F. designed cellular experiments and kinase assays. J.W. O.H. F.G. and I.K. conducted cellular studies and kinase assays. K.L.B. conducted mass spectrometry and data analysis. J.W., O.H., F.G., R.B.J., G.S.F. and S.K. wrote the manuscript.

contain multiple PIDs, each of which can serve as an interaction node, defining the function of each PID is critically important for our understanding of cellular networks. Such knowledge will also be of use in the identification and validation of PID targets for the development of novel therapeutics.

Large-scale interactome studies have identified many potential interactions involving PIDs,[2-4] but their functional importance is difficult to evaluate. This is due, in part, to the technical challenge of specifically disrupting the function of an individual PID without perturbing the function of the rest of the protein. Gene knockout and RNA interference eliminate the entire protein, and directed mutagenesis requires modification of the host genome and often affects gene expression level. Moreover, genetic perturbations can result in a phenotype that differs considerably from the phenotype caused by a reagent that binds to the same protein.[5,6] Characterization of PID function, therefore, would benefit greatly from the development of minimally invasive tools to selectively inhibit an individual PID within the cell, but this represents a considerable challenge. PIDs often comprise large families containing many structurally conserved members.[1] A specific PID inhibitor would have to distinguish among the many members of a PID family within a given genome, all of which utilize a conserved interaction interface to recognize similar ligands.

In order to assess the feasibility of selective PID inhibition, we have chosen the SH2 domain family, which recognizes phosphotyrosine (pY)-containing peptide ligands, as our model system. The SH2 domain family has all of the features that hinder the development of specific inhibitors. The phosphopeptide binding pocket is highly conserved among the 120 human SH2 domains.[7] Moreover, because roughly one-half of total phosphopeptide binding energy is derived from the pY moiety,[8] individual SH2 domains often recognize generic pY-peptides with only slightly weaker affinity than optimal pY-peptides, making it difficult to achieve high specificity using pY-based compounds.[9] Indeed, there is substantial skepticism as to whether it is possible to discriminate one SH2 domain from the others through interactions limited to the phosphopeptide-binding interface.[10,11] Although a few high-affinity SH2 domain inhibitors, most of which are small phosphopeptide mimics, have been developed, no rigorous analyses of their specificity toward SH2 targets have been reported.[12-15]

Protein-based inhibitors are a promising but unproven alternative to small molecule- or phosphopeptide-mediated inhibition of SH2–phosphopeptide interactions. Because protein-protein interactions generally involve larger interfaces than protein-small molecule or protein-peptide interactions, protein-based inhibitors could promote specificity through interactions outside of the conserved phosphopeptide-binding interface. Currently, antibodies are the most common protein-based inhibitors[16], but as large, multi-chain, disulfide-stabilized proteins, they are not likely to fold into the functional form in the reducing environment of the cytoplasm. An attractive alternative are engineered single-domain binding proteins. One of the best established systems is based on the tenth human fibronectin type III domain (FN3), a small (10 kDa), highly stable β–sandwich protein with surface loops tolerant to extensive mutations.[17] These surface loops form a binding interface analogous to that presented by the complementarity determining regions of antibodies (Fig. 1a). FN3-based binding proteins, called ‘monobodies,’ do not contain disulfides and are thus

functional in both oxidizing and reducing environments, making them amenable to *in vitro* and cellular studies. Moreover, because monobodies tend to recognize binding 'hot spots' of the target protein, they often act as competitive inhibitors.[18,19]

In this work, we chose the SH2 domain of human Abl kinase as our model target. Abl kinase is involved in a wide array of physiological processes and its oncogenic counterpart, the Bcr-Abl fusion protein, causes chronic myelogenous leukemia.[20] Moreover, structure-function studies have established the importance of the SH2 domain in Abl kinase regulation.[21-23] Using an improved phage-display platform, we generated a high-affinity and remarkably specific monobody inhibitor, HA4, to the Abl SH2 domain. The crystal structure of the HA4/Abl SH2 complex reveals how HA4 achieves such high degrees of affinity and specificity, thereby providing a guide to the development of PID inhibitors. We also assessed the consequences of the binding of HA4 to the SH2 domain within full-length Abl *in vitro* and in cells. Together, our results demonstrate the feasibility of highly specific PID inhibition, and illustrate the utility of monobody inhibitors as tools to precisely define the *in vitro* and cellular functions of an individual PID.

## Results

### Selection of FN3 monobodies to the Abl SH2 domain

We have made improvements to vector design and phage preparation methods (see Supplementary Data), that markedly enhanced the level of FN3 monobody displayed on the phage surface, resulting in a greater success rate in monobody selection. We constructed a new library in which FN3 loops were diversified with highly biased amino acid mixtures (Fig. 1b) and selected FN3 monobodies to the SH2 domain of Abl. Although we initially obtained a large number of monobodies, their affinity ($K_d$ >500 nM) was only slightly higher than typical peptide ligands for SH2 domains. Subsequent affinity maturation using a single cycle of 'loop shuffling'[24] resulted in several monobodies with higher affinity (Fig. 1b). The most frequently recovered clone, dubbed HA4, was expressed in *Escherichia coli* as a soluble protein, and its binding properties were analyzed using surface plasmon resonance (SPR). HA4 bound to the Abl SH2 domain with 7 nM $K_d$ (Fig. 1c), regardless of the presence of inorganic phosphate in solution, an affinity level similar to that of antibodies and 500–1000 times higher than that of physiological ligands of the Abl SH2.[25]

### HA4 inhibits Abl SH2–phosphopeptide interactions

Because FN3 monobodies tend to bind to hot spots of protein-protein interactions even in the absence of a tailored selection strategy, [18,19] we tested whether this was the case for HA4. The addition of HA4 to GST-Abl SH2 bound to a fluorescently labeled phosphopeptide led to a dose-dependent decrease in fluorescence polarization, indicating displacement of SH2-bound peptide by HA4. A non-binding HA4 mutant ($HA4_{Y87A}$; see below under "Interface Energetics" for its design) had no effect (Fig. 1d). Because of the low affinity of the peptide ligand, a high concentration of Abl SH2 (10 μM) was required to achieve substantial polarization change of the peptide fluorescence, a prerequisite for inhibition assays. Consequently, high concentrations of HA4 were also required to detect inhibition in this assay. The decrease in polarization caused by HA4 was nearly

stoichiometric, indicating that HA4 is a high-affinity competitive inhibitor of Abl SH2-phosphopeptide interactions.

## Protein microarray demonstrates that HA4 is highly specific

We sought to rigorously characterize the specificity of HA4 using SH2 domain protein microarrays.[4] Critically, these protein microarrays retain most of the spotted SH2 domains in their folded, functional state. Of the of 84 SH2 domains present on the chip (Supplementary Table 1), HA4 interacted most strongly with the Abl and Abl2 (Arg) SH2 domains (Fig. 2a, left panel), indicating a high level of specificity. Because the Abl and Abl2 SH2 domains share >90% sequence identity, this cross reactivity was expected. At higher concentrations of HA4, binding to a small number of additional SH2 domains was also detected (Fig. 2a, right panel). Due to differences in the amount and quality of SH2 domains at individual spots, signal intensity does not directly correlate with the strength of interaction. Therefore, we analyzed the microarray data for each spot as a function of HA4 concentration. The apparent $K_d$ values for the Abl and Abl2 SH2 domains were nearly identical (22 nM and 18 nM, respectively; Fig. 2b), and were consistent with those obtained by SPR. The majority of the spotted SH2 domains (63 domains) show no signal above background levels even with 500 nM HA4, the highest concentration used (data not shown). Furthermore, 15 randomly selected SH2 domains had no detectable interactions with HA4 as tested with SPR, indicating $K_d > 5$ μM and confirming the microarray results (data not shown).

The SH2 domains that exhibit detectable interactions with HA4 can be divided into two groups. The first group (15 domains) exhibited no signal saturation at the highest HA4 concentrations used (500 nM), indicating weak interactions (e.g. SH2D5, Fig. 2b). The second group (a total of three domains in addition to Abl and Abl2 SH2) displayed signal saturation and thus likely represents cross-reactivity of at least moderate affinity (Fig. 2b). This group includes the tandem SH2 domains of Syk (an apparent $K_d$ of 28 nM), but neither of the individual SH2 domains showed signal saturation, indicating weak interactions and suggesting the presence of a secondary HA4 binding site in the tandem SH2 domain. Grb7 SH2, which had an apparent Kd of 169nM, forms non-covalent dimers[26] that would present two SH2 molecules in close vicinity, but we could not compare monomer and dimer interactions because there was no simple means to disrupt dimerization. Finally, Blk was the only monomeric SH2 domain other than Abl and Abl2 that exhibited signal saturation. The $K_d$ values obtained from the microarray data (109 nM) and from SPR (240 nM) are substantially higher than for the HA4/Abl SH2 interaction. Together, these protein microarray and SPR analyses showed that HA4 is highly specific for Abl and Abl2.

## Interactome analysis shows highly specificity in cells

In order to evaluate the specificity of HA4 in the cellular environment, we stably expressed HA4 and $HA4_{Y87A}$ as tandem affinity-purification (TAP) tag fusion proteins in HEK293 and K562 cells.[27,28] Both cell lines express Abl, and K562 additionally expresses the constitutively active Bcr-Abl fusion protein. Endogenous proteins interacting with expressed HA4 and $HA4_{Y87A}$ were purified, visualized by silver staining, and identified using tandem protein mass spectrometry (MS) (Supplementary Fig. 2 and Supplementary Table 2). The

most abundant proteins retrieved from both cell lines were only identified when HA4, but not $HA4_{Y87A}$, was expressed, and were identified as ABL1, ABL2 and SYK (Supplementary Fig. 2c and Supplementary Table 2). In K562 cells, the TAP procedure also retrieved peptides from the BCR protein that correspond to portions present in the Bcr-Abl fusion.

A number of other proteins were also detected only when HA4 but not $HA4_{Y87A}$ were used as bait, but at lower levels. Many of these were previously identified as interactors of either Abl/Abl2 (FYN, INPPL1, CYFIP, G2DF1, NCKAP1 and FRYL)[28-31] or Bcr-Abl (BTK, INPPL1, GRB2, SHC1, UBASH3B and AP2A1).[28,32] Although some of these proteins also contain SH2 domains (FYN, INPPL1, GRB2 and SHC1), none were identified as HA4 interactors in our SH2 microarray experiments, and no binding of HA4 to the SH2 domains of Grb2, Shc1 or Fyn was detected in SPR experiments (data not shown). Together, these data indicate that these additional SH2-containing proteins associate, directly or indirectly, with Abl/Abl2 or Bcr-Abl, rather than with HA4. Importantly, the capture of these additional SH2-containing proteins provides strong support for HA4's high specificity in the cellular context. For example, the association between Bcr-Abl and Grb2 is mediated through the Grb2 SH2 domain.[33] The identification of Grb2, therefore, indicates that HA4 did not markedly interfere with the function of the endogenous Grb2 SH2 domain.

The TAP-MS data indicates that, in the two different cell types, HA4 has strong interactions with only the SH2 domains of Abl, Abl2, and Syk. Gene expression analyses indicate that 65 genes containing a total of 73 SH2 domains (out of the120 SH2 domains encoded in the human genome) are expressed in K562 cells [34], and 67 are expressed in HEK293 cells,[35] yielding a total of 92 SH2 genes expressed. Thus, HA4 successfully recognizes its targets in a complex cellular milieu with exquisite specificity. Moreover, the concordance of the highest affinity interactors identified by protein microarray and by TAP-MS provides cross-validation for these very different methods of specificity assessment.

## Crystal structure of the HA4/Abl SH2 domain complex

In order to understand how HA4 achieves such remarkable levels of specificity and affinity, we determined the crystal structure of HA4 in complex with the Abl SH2 domain at 1.75 Å resolution (Table 1). The asymmetric unit contains two HA4/Abl SH2 complexes (Fig. 3a); the two HA4 molecules form an intermolecular β-sheet through antiparallel association of β-strand A and the N-terminal tail, a common mode of crystal contacts observed for β-rich proteins. HA4 maintains the FN3 scaffold structure as evidenced by small difference from previously determined monobody structures ($C^{\alpha}$ root mean squared deviation (RMSD) <0.7 Å, excluding the three mutated loops).[19] The overall structure of the Abl SH2 domain is in good agreement with published NMR structures and with the corresponding portion within the crystal structures of larger segments of Abl kinase ($C^{\alpha}$ RMSD <1 Å).[36,37]

HA4 binds to the phosphopeptide-binding site of the Abl SH2 domain, consistent with the inhibition data. The interface buries a total of 684$Å^2$ of monobody surface area and 651 $Å^2$ of the SH2 domain, roughly in line with the surface area burial observed in crystal structures of monobody/maltose binding protein complexes.[19] Interestingly, the electron density map contained a well-resolved density in the pY binding pocket, consistent with an inorganic

phosphate ion, (probably from a phosphate buffer used for SH2 preparation). This density is adjacent to the hydroxyl group of Tyr87 of HA4 such that Tyr87 and the putative inorganic phosphate together mimic pY (Fig. 3b). Consistent with such mimicry, the βB-βC loop of the SH2 domain is in the 'closed' conformation characteristic of an SH2 domain with a bound phosphopeptide.[38] As noted above, inclusion of phosphate did not alter the affinity of HA4 for the Abl SH2 domain (Fig. 1c), indicating that the bound phosphate ion contributes little to the energetics of the HA4-SH2 interaction.

The HA4 structure resembles a cupped hand consisting of the 'palm' and 'fingers' upon which the Abl SH2 domain rests (Fig. 3b). The palm consists of one β-sheet of the FN3 scaffold and the C-terminal residue of the BC loop (Tyr35). The contacts made by HA4 'palm' residues, contributing ∼40% of the total interface area, are at the periphery of the phosphopeptide-binding interface. A β-hairpin formed by a long FG loop of HA4 (residues 79–90) corresponds to the fingers that recognize the center of the phosphopeptide-binding interface, which make up ∼60% of the interface. The finger residues in the FG loop closely mimic the canonical backbone conformation of an SH2-bound phosphopeptide (Fig. 3c). In the stereotypical mode of SH2-phosphopeptide interaction, the phosphopeptide lies perpendicular to the SH2 central β-sheet, and interacts with surface cavities on either side of the central β-strand (βD) (Fig. 4b).[38],[39] One of these cavities accommodates pY, while the other forms a pocket for the amino acid three-residues C-terminal to pY (termed "Y+3"). In the HA4 FG loop, Tyr87 is inserted into the pY pocket where it forms cation-π interactions with two arginines along the sides of the pocket (Arg194 and Arg153; note that two-digit numbers refer to HA4 residues and three-digit numbers to SH2 residues), and polar contacts with Ser181 at the base (Fig. 3d). Residues Ser84–Gly86 form a network of polar interactions with SH2 residues at the periphery of the pY pocket. Although Asp83 does not contact the SH2 domain, it forms intramolecular hydrogen bonds with Gly86 and Tyr87, which may be important for supporting the FG loop conformation. Met88 makes hydrophobic contacts (Fig. 3b) and forms a main-chain H-bond with His192 of the SH2 domain (Fig. 3d). C-terminal to Met88, the main chain bends away from the surface such that the 'Y+3 pocket' of the SH2 domain is not engaged by the amino acid three residues C-terminal to Tyr87 (Fig. 3c). Instead, distinct from the canonical mode of peptide-SH2 interactions, Trp80, at the N-terminus of the FG loop, forms a large hydrophobic surface that blankets the Y+3 binding pocket.

Unlike the combinatorially optimized FG loop residues, residues mediating the 'palm' interactions (Tyr35, Tyr36, Arg38, Glu52 and Tyr78) are located in the invariant FN3 scaffold, suggesting that these contacts are serendipitous. These contacts, both hydrophobic and polar, are with portions of the Abl SH2 domain at the periphery of the phosphopeptide-binding interface (Fig. 3 and 4c).

The HA4-Abl SH2 interface shows intermediate characteristics, in terms of size, chemistry and shape complementarity,[40] between a typical SH2-phosphopeptide interaction and typical antibody-antigen interactions (Supplementary Table 3). These statistics are consistent with the structural observation that HA4 utilizes a combination of peptide-like and antibody-like modes of recognition to achieve high affinity binding.

The HA4-binding interface of Abl SH2 includes nearly the entire pY-peptide-binding interface, as revealed by comparison with the peptide-binding interface of the Lck SH2 domain,[39] the closest Abl homologue for which a peptide-bound structure is available (Fig. 4a). Of the 18 positions in the peptide-binding interface in the Lck SH2 structure, 13 are identical between Abl and Lck SH2 domains, and three are conservative substitutions; many of these residues are also highly conserved across the entire Src family (Fig. 4a and c). Therefore, inhibitors targeting this area alone are unlikely to achieve high specificity.

About one third of the HA4-binding interface lies outside the peptide-binding interface (Fig. 4a). While these residues are nearly identical between Abl and Abl2, they are less conserved between Abl/Abl2 and the Src family (Fig. 4a), providing the potential for unique contacts. In addition, Abl SH2 lacks the extended βC-βD loop characteristic of the Src family (Fig. 4). This loop abuts the βD residues of the SH2 domain that interact with HA4 at the periphery of the interface. The CD-loop of Src family SH2 domains is packed with charged and high-entropy residues, which forms a large electronegative 'knob' that is predicted to clash with the HA4 DE loop, negatively impacting HA4 binding (Fig. 4d). Consistent with this notion, the SH2 domain of Blk, the only member of the Src family with which HA4 interacts, has a knob that is reduced in size and charge (Fig. 4a). These structural features likely explain both the observed cross-reactivity of HA4 to Abl/Abl2, and the absence of cross-reactivity to members of the Src family and other SH2 domains.

## Interface Energetics

We performed alanine-scanning mutagenesis to assess the energetic contributions of individual monobody residues. We constructed individual point mutants of seven HA4 residues, each burying more than 40 $Å^2$ and altogether responsible for >75% of surface area burial, and measured their binding to the Abl SH2 domain by SPR. Mutants at four finger positions (Y36A, W80A, Y87A, M88A) had no detectable binding, corresponding to a $\Delta\Delta G \geq 3.8$ kcal $mol^{-1}$ (Fig. 3f). We chose binding-defective mutant Y87A as a negative control for biochemical and cellular experiments. Two 'palm' residues (R38A and E52A) contributed considerably ($\Delta\Delta G = \sim 2.2$ and 2.4 kcal $mol^{-1}$, respectively), but another, Y35A, marginally ($\Delta\Delta G \sim 0.8$ kcal/mol). These results validate the contacts observed in the crystal structure, and highlight the importance of both palm and finger residues to binding energetics.

## HA4 activates the Abl kinase *in vitro*

The SH2 domain plays an important role in Abl autoinhibition by interacting with the C-terminal lobe of the kinase domain.[36] The binding of Abl SH2 to phosphopeptide ligands activates Abl *in vitro*, and is thought to mimic the activation of Abl by cellular substrates.[23,41] Superposition of the HA4/Abl SH2 structure on the autoinhibited structure of Abl indicates steric clashes between HA4 and the kinase domain (Fig. 5a). In addition, HA4 directly interacts with SH2 residues that form salt bridges and hydrogen bonds with the C-lobe of the kinase in autoinhibited Abl (Fig. 5b). These observations suggest that the binding of HA4 to the SH2 domain should disrupt the autoinhibited conformation, thereby activating the kinase. Indeed, in *in vitro* Abl kinase assays HA4 and a phosphopeptide derived from c-Jun ($K_d = \sim 7$ μM)[41] caused a dose-dependent increase in Abl activity but $HA4_{Y87A}$ did not

(Fig. 5c, left). HA4 was a substantially more potent activator than the phosphopeptide, in line with its ∼1000× higher affinity for the SH2 domain.

In contrast to its effect on the autoinhibited wild-type kinase, HA4 did not further increase the activity of mutated forms of Abl in which autoinhibition is already disrupted (Fig. 5c, right). This supports the model that HA4 activates Abl by disrupting the interactions that maintain the autoinhibited conformation. It is also notable that HA4 did not inhibit these activated Abl mutants indicating that the HA4-interaction interface of Abl SH2 is distinct from the interface required for kinase activation.[23]

The concentration of HA4 at which activation of Abl was half-maximal (∼0.4–0.5 μM) was ∼60 times higher than the $K_d$ of HA4 for the isolated SH2 domain. Likewise, the concentration of phosphopeptide required for Abl half-maximal activation is much higher than the measured $K_d$ for the SH2-peptide interaction. This disparity may indicate the challenge that HA4 faces in competing with the kinase domain for binding to the SH2 domain. Because the kinase domain is a part of the same polypeptide as the SH2 domain, its effective concentration for SH2 interaction is extremely high. These data also imply that the phosphopeptide-binding interface of the SH2 domain is largely inaccessible in autoinhibited c-Abl, but the conformational dynamics of Abl occasionally permit HA4 to bind the SH2 domain, shifting the equilibrium away from the closed, autoinhibited conformation toward a more active open state.

## HA4 blocks processive phosphorylation by Abl in cells

For tyrosine kinases containing an SH2 domain, the SH2 domain may aid in the phosphorylation of substrates containing multiple phosphorylation sites by binding a singly-phosphorylated substrate and positioning the kinase domain for subsequent rounds of phosphorylation.[42] This mechanism is referred to as processive phosphorylation and is believed to increase substrate specificity. To test whether the SH2 domain of Abl is responsible for processive phosphorylation in the cellular context, we co-transfected HEK293 cells with a constitutively active form of Abl, HA4, and paxillin, a *bona fide* Abl substrate containing multiple phosphorylation sites (Fig. 6a).[43] An active Abl form was chosen to eliminate complications from HA4's ability to activate wild-type Abl, and to ensure efficient phosphorylation of paxillin by Abl.

In the presence of $HA4_{Y87A}$, or in the absence of any monobody, Abl efficiently phosphorylated paxillin at multiple sites, giving rise to a characteristic smear of slower-migrating species observed on an anti-pY blot (Fig. 6b). In contrast, HA4 dramatically reduced multi-site phosphorylation of paxillin. Moreover, at low paxillin concentrations, HA4 also substantially reduced Abl auto-phosphorylation when compared with $HA4_{Y87A}$ (Lanes 5 and 6). These results demonstrate that the SH2 domain of Abl is critical for the processive phosphorylation of paxillin and suggest that HA4 prevents the assembly of activated signaling complexes that facilitate the phosphorylation and activation of Abl (Fig. 6a). These results are, to our knowledge, the first demonstration of the intracellular inhibition of processive phosphorylation, and provide evidence for the potential physiological importance of a phenomenon mainly observed *in vitro*.[44]

### HA4 disrupts Bcr-Abl-mediated signaling

The ability of HA4 to disrupt processive phosphorylation illustrates its potential as an intracellular inhibitor of endogenous signaling pathways. Thus, we tested the effect of intracellular expression of HA4 on the phosphorylation level of STAT5 in K562 cells. Tyrosine phosphorylation and subsequent activation of STAT5 is a critical step for the induction of chronic myeloid leukaemia (CML) and a commonly used read-out for Bcr-Abl function.[45] The level of STAT5 phosphorylation was inversely correlated with the level of HA4 expression, while $HA4_{Y87A}$ had no effect (Fig. 6c and d). These results clearly demonstrate that HA4 can disrupt endogenous Bcr-Abl dependent signaling events and indicate that the Abl SH2 domain function is necessary for STAT5 activation by Bcr-Abl. Our results are consistent with previous work demonstrating that the deletion of both SH2 and SH3 domains abolishes the ability of Bcr-Abl to activate STAT5,[46] and they further provide mechanistic evidence for a major role of the SH2 domain in STAT5 activation.

## Discussion

In this work we successfully selected a highly specific, high-affinity binding protein to a small protein domain optimized for interactions with phosphopeptides. In terms of affinity, HA4 compares favorably to previously developed SH2 domain inhibitors based on peptide mimics, which showed $K_d$ values in the range of several to several hundred nanomolar.[12-15]. The sole SH2 domain to which inhibitors of considerably higher affinity (Kd <500 pM) have been reported is Grb2.[13,47]

Synthetic SH2 inhibitors are mostly conformationally constrained phosphopeptide mimetics, designed to optimize contacts in the binding hotspots of SH2 domains, the pY and Y+3 pockets, while reducing the unfavorable entropic loss incurred upon binding through cyclization. The FG loop of HA4 structurally resembles these inhibitors: it interacts extensively with the pY pocket and the Y+3 pocket, and it is effectively cyclized. Its β-turn structure, if formed in the unbound state, could help further reduce the entropic loss upon binding.

The convergence of HA4 and synthetic inhibitors to a similar mode of interaction suggests that there are inherent structural requirements for high-affinity binding to the SH2 domain phosphopeptide-binding interface. Since these requirements are shared among SH2 domain family members, it may be extremely difficult to achieve both high-affinity and high-specificity recognition through interactions limited to the phosphopeptide-binding interface. Recent thermodynamic studies of c-Src SH2 domain interactions with pY-peptide and synthetic ligands highlight this challenge,[10] by showing that the Y+3 ‘specificity pocket’ is poorly equipped to mediate specificity, since it is highly flexible and can bind diverse aliphatic ligands with similar affinity. The pY pocket, on the other hand, is rigid and cannot easily accommodate ligands that are structurally divergent from phosphotyrosine. Since the pY pocket is highly conserved, ligands binding to this pocket should bind with similar affinity to many other SH2 domains, as supported by recent molecular dynamics studies.[11]

While HA4 structurally resembles pY-peptides and pY-peptide-like inhibitors, it does not employ a pY or a pY mimic, and hence does not rely on extensive charge-charge

interactions with the conserved pY pocket. In fact, HA4 interactions with SH2 resemble those of the few SH2 domains that bind peptide substrates in a pY-independent manner. One such SH2 domain, SAP, achieves high affinity by forming a very large peptide/SH2 interface,[48] with contacts with less conserved portions of the SH2 domain that promote specificity. Recent work has established that enhanced specificity in SH2/target interactions can be achieved through a secondary interface away from the phosphopeptide-binding site.[49] These findings suggest that, to achieve high-specificity for an SH2 domain, it is necessary to 'branch out' and make contacts with less-conserved surfaces outside the peptide-binding interface. The structural information presented here will be useful for improving HA4's affinity and specificity and for designing monobody libraries specifically targeted to SH2 domains.

The fact that HA4 generation did not require extensive cycles of affinity maturation highlights the effectiveness of our library design. We applied a 'biased-diversity' strategy that has been successfully employed to select synthetic antibodies to a diverse set of targets.[50] Our success with the monobody scaffold illustrates the generality of this strategy and validates that, with respect to molecular recognition, not all amino acids are equal. Binding interfaces consisting solely of serine and tyrosine have been shown to mediate high-affinity and high-specificity molecular recognition,[51],[52] and even when the amino acid palette is expanded to incorporate other amino acid types, tyrosine still plays a dominant role in the interaction interface.[19],[53] In the HA4/Abl SH2 domain structure, tyrosine contributes >40% of the HA4 interface.

HA4 is the first monobody to be functionally expressed within mammalian cells. Thus, this work places monobodies as a useful tool in the manipulation of intracellular networks for research and therapeutic ends, alongside other intracellular inhibitors (e.g., the so-called peptide aptamers and intrabodies).[6],[54] More specifically, HA4 should be of immediate use in the dissection of Abl signaling networks, to help identify interaction partners of Abl SH2 in the cellular context and determine physiological consequences of interrupting SH2 interactions. HA4 should also prove valuable in the study of oncogenic signaling mediated by Bcr-Abl. Mutation of the SH2 domain drastically reduces Bcr-Abl oncogenicity, suggesting that the SH2 domain recruits some of the functionally critical downstream targets.[55],[56]

In conclusion, this work establishes a platform for the rapid selection of highly specific, intracellular inhibitors of any desired PID. PID inhibitors will enable researchers to precisely define the roles of individual domains in cellular protein interaction networks without otherwise altering the genome, thereby yielding insight into basic cellular physiology and aiding in the identification of novel targets for therapeutic intervention.

## Methods

### Vector Construction

We constructed the phage display vector, DsbFNp3FL as follows. We amplified the 5′ portion of the M13 gene 3 from M13mp18 using a forward primer that introduced an MluI site N-terminal to the mature p3 protein (5′-

GCATACGCGTGCTGAAACTGTTGAAAG-3′) and a reverse primer annealing to the downstream of a natural ClaI site in M13 gene 3. After digesting with the two enzymes, we cloned the fragment into a phage-display vector containing the gene for the C-terminal domain of p3, described previously,[57] to regenerate the full-length p3 gene. *Escherichia coli* cells used for phage display harbored a LacI-containing plasmid, pMCSG21 (a generous gift of Dr. Mark Donnelly, Argonne National Laboratory), based on pCDFDuet-1 (Novagen). The Abl1 and Abl2 expression vectors were gifts of Dr. Piers Nash, University of Chicago.[58] All other SH2 domain genes were from Open Biosystems (catalog number, OHS4902), and we cloned them into a vector for an N-terminal $His_{10}$-tag fusion protein,[51] as described in Supplementary Methods. We constructed alanine mutants by Kunkel mutagenesis as described previously.[19]

### Phage display, library construction and selection

We diversified FN3 loop residues in the library using Kunkel mutagenesis with a custom trimer phosphoramidite oligonucleotide mix (Glen Research) as described previously.[50,57] Amino acid compositions are provided in Figure 1 legend. Phage display selection is described in Supplementary Methods. Phage ELISA analysis of selected clones was described previously.[57]

### Protein sample preparation

We produced proteins with an N-terminal His-tag and purified using a Ni–Sepharose column (GE Lifesciences) to apparent homogeneity as described previously.[51] Monobodies contained a N-terminal His-tag and a unique Cys as a C-terminal extension. We removed the His-tag moiety by protease cleavage prior to SPR measurements and crystallization.

### Surface plasmon resonance

We performed measurements in 10 mM sodium phosphate buffer, pH 7.4 containing 150 mM NaCl, 0.005% (w/v) Tween 20 and 50 μM ethylenediaminetetraacetic acid at 25 °C on a BIAcore2000 instrument. We immobilized purified SH2 domains via a His-tag to a nickel–NTA sensor chip (BIAcore). We then flowed a monobody over the sensor chip at a rate of 50 μl $min^{-1}$ and monitored the association and dissociation kinetics. We corrected all data by subtraction of data for a blank run (with no monobody injection) and fitted them to a 1:1 Langmuir binding model by global fitting of multiple kinetic traces using BIAevaluation software (BIAcore). For HA4 alanine mutants, we measured binding data at monobody concentrations of 250, 500 and 1,000 nM.

### Fluorescence polarization assay for the inhibition of SH2-peptide interaction

We performed measurements in 10 mM sodium phosphate buffer, pH 7.4 containing 150 mM NaCl, 1 mM dithiothreitol and 1% (w/v) bovine serum albumin, using 5 nM rhodamine-labeled phosphopeptide, KEEGpYELPYNP, at room temperature using a Beacon 2000 Instrument (Invitrogen).

### SH2 microarray

We prepared the SH2 microarrays according to procedures described elsewhere.[4] We labeled the HA4 protein at a unique cysteine residue with DyLight 800 maleimide (Thermo Scientific) and incubated with individual SH2 domain arrays, as described elsewhere.[4] We detected HA4 using a LI-COR Odyssey scanner. We determined mean fluorescence intensity values for each spot using GenePix Pro 6.0 software (Molecular Devices). We fit signals displaying saturation with a 1:1 binding model using Origin7 (OriginLab).

### Mammalian Cell-culture and TAP-MS

We cloned the HA4 gene as an N-terminal fusion to the GS-TAP tag Gateway destination vector described previously.[27] We introduced the HA4-TAP fusion (and a negative control with $HA4_{Y87A}$) into HEK293 and K562 cells by retroviral gene transfer. We sorted transfected cells based on GFP expression and grew them in DMEM with 10% (v/v) fetal calf serum, 100 units $ml^{-1}$ penicillin and 0.1 mg $ml^{-1}$ streptomycin. We prepared cell lysates and carried out TAP/MS as described previously.[27,28] We identified proteins by automated database search, with a minimal threshold value of two peptides for a positive protein identification, and grouped protein identifications according to shared peptides. We dropped background proteins commonly found in similar TAP experiments from further analysis. We scored proteins as HA4-interacting only if they failed to interact with the $HA4_{Y87A}$ mutant. We established interactions between individual HA4-interacting proteins identified in TAP-MS on the basis of UNIProt database searches for binary interactions. We estimated the total number of expressed SH2-containing proteins in HEK293 cells on the basis of expression calls from control samples contained in NCBI GEO dataset GDS2426.[35] We counted an SH2 domain as expressed if ‘present’ in more than 50% of control replicates.

### Kinase Assay

We transiently transfected HEK293 cells with vectors for c-Abl, constitutively active Abl PP (P242E/P249E) or Bcr-Abl, and immunoprecipitated the Abl proteins using an anti-Abl antibody. We performed *in vitro* kinase assays by monitoring the incorporation of $^{32}P$ into a preferred Abl substrate, biotin-GGEAIYAAPFKK-amide, as described elsewhere.[41]

### HEK293 transfectionf

We cultured HEK293 cells in Dulbecco's Modified Eagle's Medium supplemented with 10% (v/v) fetal calf serum. We transfected cells with the designated amounts of expression vectors for AblG2APP, myc-HA4, myc-$HA4_{Y87A}$, and HA-Paxillin using PolyFect (Qiagen). Following a 40 hr growth period, we lysed cells in IP buffer (see Supplementary Methods for its composition), and insoluble material was removed by centrifugation. We performed Western blotting with anti-HA (Covance), anti-Abl (Oncogene Science), anti-phosphotyrosine (Cell Signaling) and anti-myc (Rockland) antibodies.

### K562 transfection and flow cytometry

We transfected $1 \times 10^6$ K562 cells with 5 μg of expression vectors encoding GFP-HA4 or GFP-$HA4_{Y87A}$ fusion proteins or GFP alone using the Nucleofector Kit V (Amaxa), applying program T-016. We harvested cells 48 hrs later, fixed in 3.2% (v/v)

paraformaldehyde and stored in methanol overnight. We stained cells using an Alexa-Fluor-647-conjugated anti-phospho-STAT5 antibody (BD Biosciences) and analyzed on a LSRII flow cytometer (BD Biosciences) using FlowJo software (Treestar).

### Crystallization, data collection and structure determination

Methods are described in the Supplementary Methods.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank R. Gilbreth and Dr. E. Duguid for assistance with X-ray structure determination; Drs. E. Duguid and C. He for assistance with DNA synthesis; M. Ciaccio, R. Gilbreth and Drs. P. Nash, B. Liu and A.A. Kossiakoff for discussion; the staff of the Life Sciences Collaborative Access Team (LS-CAT) beamline at the Advanced Photon Source and the University of Chicago DNA Sequencing Core facility for technical support; Dr. A.A. Kossiakoff for access to the BIAcore instrument; A. Müller, N. Venturini and M. Planyavsky for expert assistance with the MS analysis; the Structural Genomics Consortium for making the SH2 vectors available through Open Biosystems. This work was supported by National Institutes of Health grants R01-GM72688, U54-GM74946 and R21-CA132700 to S.K., by the University of Chicago Cancer Research Center and by the Austrian Academy of Sciences. J.W. was supported in part by NIH grant 5T32GM07281-33 and EMBO fellowship ASTF 293.00-2009. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817).

## References

1. Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. Science. 2003; 300:445–52. [PubMed: 12702867]
2. Rual JF, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005; 437:1173–8. [PubMed: 16189514]
3. Kocher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. Nat Methods. 2007; 4:807–15. [PubMed: 17901870]
4. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006; 439:168–74. [PubMed: 16273093]
5. Weiss WA, Taylor SS, Shokat KM. Recognizing and exploiting differences between RNAi and small-molecule inhibitors. Nat Chem Biol. 2007; 3:739–44. [PubMed: 18007642]
6. Visintin M, Melchionna T, Cannistraci I, Cattaneo A. In vivo selection of intrabodies specifically targeting protein-protein interactions: a general platform for an “undruggable” class of disease targets. J Biotechnol. 2008; 135:1–15. [PubMed: 18395925]
7. Campbell SJ, Jackson RM. Diversity in the SH2 domain family phosphotyrosyl peptide binding site. Protein Eng. 2003; 16:217–27. [PubMed: 12702802]
8. Bradshaw JM, Mitaxov V, Waksman G. Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. J Mol Biol. 1999; 293:971–85. [PubMed: 10543978]
9. Ladbury JE, Arold S. Searching for specificity in SH domains. Chem Biol. 2000; 7:R3–8. [PubMed: 10662684]
10. Taylor JD, et al. Structure, dynamics, and binding thermodynamics of the v-Src SH2 domain: implications for drug design. Proteins. 2008; 73:929–40. [PubMed: 18536014]
11. Gan W, Roux B. Binding specificity of SH2 domains: insight from free energy simulations. Proteins. 2009; 74:996–1007. [PubMed: 18767163]
12. Shakespeare WC. SH2 domain inhibition: a problem solved? Curr Opin Chem Biol. 2001; 5:409–15. [PubMed: 11470604]

13. Machida K, Mayer BJ. The SH2 domain: versatile signaling module and pharmaceutical target. Biochim Biophys Acta. 2005; 1747:1–25. [PubMed: 15680235]
14. Mandine E, et al. High-affinity Src-SH2 ligands which do not activate Tyr(527)-phosphorylated Src in an experimental in vivo system. Biochem Biophys Res Commun. 2002; 298:185–92. [PubMed: 12387813]
15. Garcia-Echeverria C. Antagonists of the Src homology 2 (SH2) domains of Grb2, Src, Lck and ZAP-70. Curr Med Chem. 2001; 8:1589–604. [PubMed: 11562287]
16. Reichert JM, Rosensweig CJ, Faden LB, Dewitz MC. Monoclonal antibody successes in the clinic. Nat Biotechnol. 2005; 23:1073–8. [PubMed: 16151394]
17. Koide A, Bailey CW, Huang X, Koide S. The fibronectin type III domain as a scaffold for novel binding proteins. J Mol Biol. 1998; 284:1141–51. [PubMed: 9837732]
18. Koide A, Abbatiello S, Rothgery L, Koide S. Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor. Proc Natl Acad Sci U S A. 2002; 99:1253–8. [PubMed: 11818562]
19. Gilbreth RN, Esaki K, Koide A, Sidhu SS, Koide S. A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. J Mol Biol. 2008; 381:407–18. [PubMed: 18602117]
20. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science. 1990; 247:824–30. [PubMed: 2406902]
21. Nagar B, et al. Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. Mol Cell. 2006; 21:787–98. [PubMed: 16543148]
22. Hantschel O, Superti-Furga G. Regulation of the c-Abl and Bcr-Abl tyrosine kinases. Nat Rev Mol Cell Biol. 2004; 5:33–44. [PubMed: 14708008]
23. Filippakopoulos P, et al. Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. Cell. 2008; 134:793–803. [PubMed: 18775312]
24. Hackel BJ, Kapila A, Wittrup KD. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. J Mol Biol. 2008; 381:1238–52. [PubMed: 18602401]
25. Tokonzaba E, Capelluto DG, Kutateladze TG, Overduin M. Phosphoinositide, phosphopeptide and pyridone interactions of the Abl SH2 domain. Chem Biol Drug Des. 2006; 67:230–7. [PubMed: 16611216]
26. Porter CJ, et al. Grb7 SH2 domain structure and interactions with a cyclic peptide inhibitor of cancer cell migration and proliferation. BMC Struct Biol. 2007; 7:58. [PubMed: 17894853]
27. Burckstummer T, et al. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. Nat Methods. 2006; 3:1013–9. [PubMed: 17060908]
28. Brehme M, et al. Charting the molecular network of the drug target Bcr-Abl. Proc Natl Acad Sci U S A. 2009; 106:7414–9. [PubMed: 19380743]
29. Basu D, El-Assal Sel D, Le J, Mallery EL, Szymanski DB. Interchangeable functions of Arabidopsis PIROGI and the human WAVE complex subunit SRA1 during leaf epidermal development. Development. 2004; 131:4345–55. [PubMed: 15294869]
30. Ewing RM, et al. Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol. 2007; 3:89. [PubMed: 17353931]
31. Wu C, et al. Systematic identification of SH3 domain-mediated human protein-protein interactions by peptide array target screening. Proteomics. 2007; 7:1775–85. [PubMed: 17474147]
32. MacPartlin M, Smith AM, Druker BJ, Honigberg LA, Deininger MW. Bruton's tyrosine kinase is not essential for Bcr-Abl-mediated transformation of lymphoid or myeloid cells. Leukemia. 2008; 22:1354–60. [PubMed: 18548107]
33. Pendergast AM, et al. BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. Cell. 1993; 75:175–85. [PubMed: 8402896]
34. Hantschel O, et al. The chemokine interleukin-8 and the surface activation protein CD69 are markers for Bcr-Abl activity in chronic myeloid leukemia. Mol Oncol. 2008; 2:272–81. [PubMed: 19383348]

35. El Hader C, et al. HCaRG increases renal cell migration by a TGF-alpha autocrine loop mechanism. Am J Physiol Renal Physiol. 2005; 289:F1273–80. [PubMed: 16033922]
36. Nagar B, et al. Structural basis for the autoinhibition of c-Abl tyrosine kinase. Cell. 2003; 112:859–71. [PubMed: 12654251]
37. Overduin M, Rios CB, Mayer BJ, Baltimore D, Cowburn D. Three-dimensional solution structure of the src homology 2 domain of c-abl. Cell. 1992; 70:697–704. [PubMed: 1505033]
38. Waksman G, Shoelson SE, Pant N, Cowburn D, Kuriyan J. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. Cell. 1993; 72:779–90. [PubMed: 7680960]
39. Eck MJ, Shoelson SE, Harrison SC. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. Nature. 1993; 362:87–91. [PubMed: 7680435]
40. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol. 1993; 234:946–50. [PubMed: 8263940]
41. Hantschel O, et al. A myristoyl/phosphotyrosine switch regulates c-Abl. Cell. 2003; 112:845–57. [PubMed: 12654250]
42. Patwardhan P, Miller WT. Processive phosphorylation: mechanism and biological importance. Cell Signal. 2007; 19:2218–26. [PubMed: 17644338]
43. Schaller MD, Schaefer EM. Multiple stimuli induce tyrosine phosphorylation of the Crk-binding sites of paxillin. Biochem J. 2001; 360:57–66. [PubMed: 11695992]
44. Mayer BJ, Hirai H, Sakai R. Evidence that SH2 domains promote processive phosphorylation by protein-tyrosine kinases. Curr Biol. 1995; 5:296–305. [PubMed: 7780740]
45. Ye D, Wolff N, Li L, Zhang S, Ilaria RL Jr. STAT5 signaling is required for the efficient induction and maintenance of CML in mice. Blood. 2006; 107:4917–25. [PubMed: 16522816]
46. Nieborowska-Skorska M, et al. Signal transducer and activator of transcription (STAT)5 activation by BCR/ABL is dependent on intact Src homology (SH)3 and SH2 domains of BCR/ABL and is required for leukemogenesis. J Exp Med. 1999; 189:1229–42. [PubMed: 10209040]
47. Schoepfer J, et al. Highly potent inhibitors of the Grb2-SH2 domain. Bioorg Med Chem Lett. 1999; 9:221–6. [PubMed: 10021933]
48. Poy F, et al. Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. Mol Cell. 1999; 4:555–61. [PubMed: 10549287]
49. Bae JH, et al. The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. Cell. 2009; 138:514–24. [PubMed: 19665973]
50. Fellouse FA, et al. High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. J Mol Biol. 2007; 373:924–40. [PubMed: 17825836]
51. Koide A, Gilbreth RN, Esaki K, Tereshko V, Koide S. High-affinity single-domain binding proteins with a binary-code interface. Proc Natl Acad Sci U S A. 2007; 104:6632–7. [PubMed: 17420456]
52. Fellouse FA, et al. Molecular recognition by a binary code. J Mol Biol. 2005; 348:1153–62. [PubMed: 15854651]
53. Fellouse FA, Barthelemy PA, Kelley RF, Sidhu SS. Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. J Mol Biol. 2006; 357:100–14. [PubMed: 16413576]
54. Hoppe-Seyler F, Crnkovic-Mertens I, Tomai E, Butz K. Peptide aptamers: specific inhibitors of protein function. Curr Mol Med. 2004; 4:529–38. [PubMed: 15267224]
55. Mayer BJ, Jackson PK, Van Etten RA, Baltimore D. Point mutations in the abl SH2 domain coordinately impair phosphotyrosine binding in vitro and transforming activity in vivo. Mol Cell Biol. 1992; 12:609–18. [PubMed: 1370711]
56. Skorski T, et al. Transformation of hematopoietic cells by BCR/ABL requires activation of a PI-3k/Akt-dependent pathway. Embo J. 1997; 16:6151–61. [PubMed: 9321394]
57. Koide A, Koide S. Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. Methods Mol Biol. 2007; 352:95–109. [PubMed: 17041261]

58. Machida K, et al. High-throughput phosphotyrosine profiling using SH2 domains. Mol Cell. 2007; 26:899–915. [PubMed: 17588523]

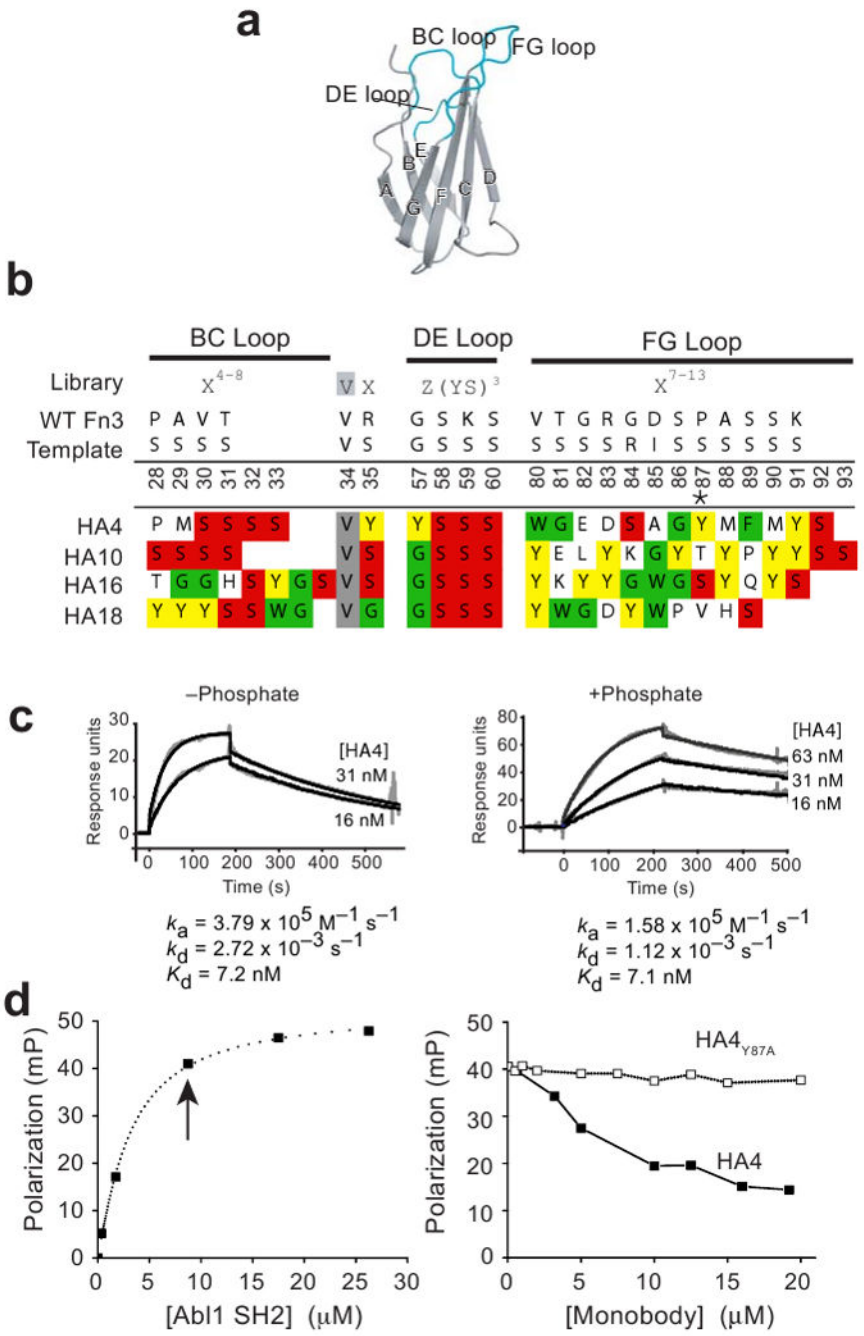


**Figure 1. Library design and selected Abl SH2-binding monobodies**

**(a)** Schematic of the FN3 scaffold. β-Strands are labeled with A–G and loop regions diversified in the combinatorial library are in cyan. Figure generated using PyMOL (www.pymol.org). **(b)** Library design and loop sequences of Abl SH2-binding monobodies. X refers to a mixture of 30% Tyr (yellow), 15% Ser (red), 10% Gly (Green), 5% each of Trp, Phe and Arg (Green) and 2.5% each of all other amino acids except Cys. Z refers to a mixture of 50% Gly, 25% Tyr and 25% Ser. The numbers indicate positions for HA4. The Tyr87 position, mutated in the $HA4_{Y87A}$ non-binding control, is marked with the asterisk.

Because of differences in loop lengths, the numbering does not correspond to previously published monobodies. **(c)** SPR traces for HA4 binding to immobilized Abl SH2 domain, corrected by subtraction of the sensorgram for a blank run (gray) Parameters for the global Langmuir fit are provided, and the black lines show the best fit. Left, measurements in non-phosphate buffer. Right, measurements in phosphate buffer. **(d)** Left, fluorescence polarization changes of a rhodamine-labeled pY-peptide as a function of GST-Abl SH2 added to the solution. The concentration of GST-Abl SH2 required to give ∼80% maximum polarization (10 μM, indicated with the arrow) was used for HA4 competition assay shown on the right panel. Right: Fluorescence polarization of the rhodamine-labeled pY-peptide in the presence of GST-Abl SH2 is plotted versus the concentration of monobody added to the solution.

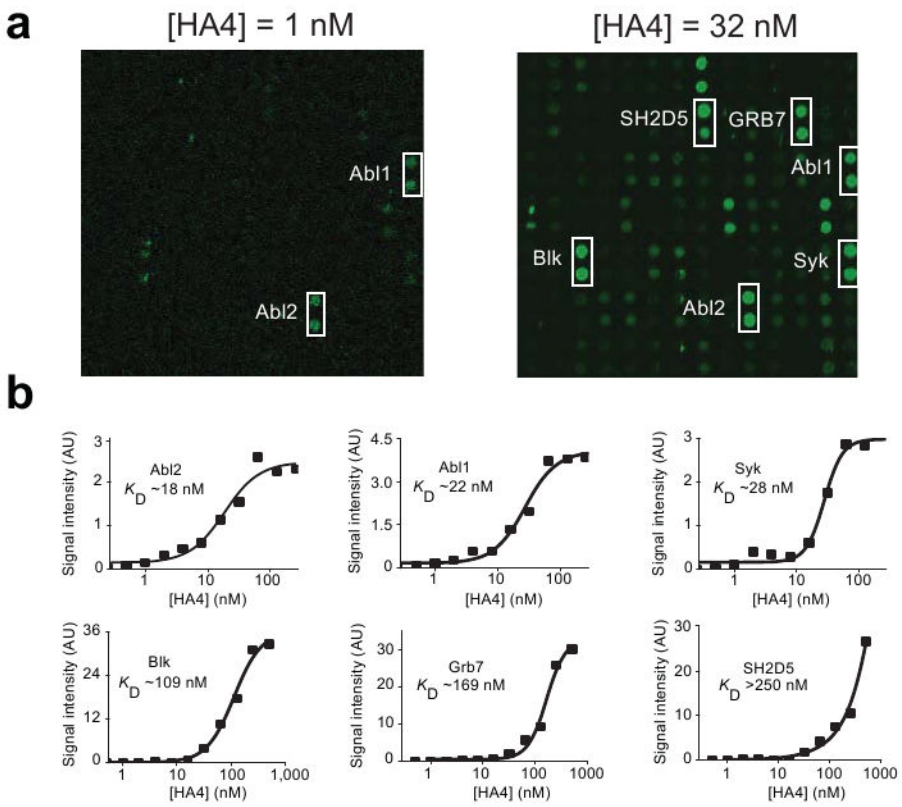


**Figure 2. SH2 protein microarray assays of HA4 specificity**

**(a)** Fluorescence images of SH2 chips following incubation with fluorophore-labeled HA4. Each SH2 domain is spotted in duplicate. A listing of spotted SH2 domains is provided in Supplementary Table 1. Spots for the SH2 domains shown in (b) are enclosed in the boxes. **(b)** The fluorescence signal strengths of individual SH2 domain spots are plotted as a function of HA4 concentration. The curves show the best fit of a 1:1 binding model, and the estimated dissociation constants are provided.

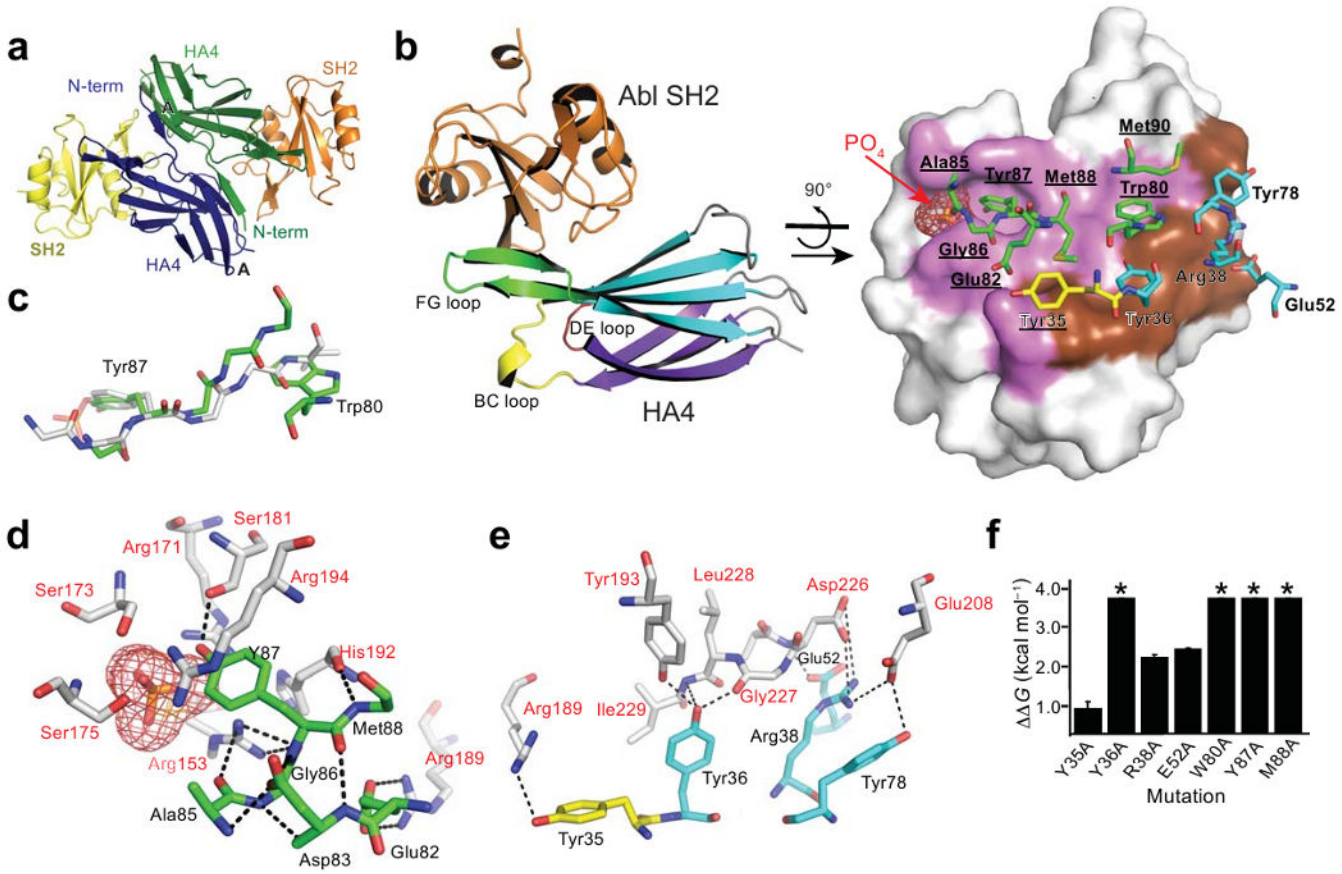


**Figure 3. The crystal structure of the HA4/Abl SH2 complex**

**(a**) The asymmetric unit containing two copies of the complex. β-strand A and the N-terminal extension, forming an intermolecular β-sheet, are labeled. **(b**) The HA4/Abl SH2 interface. Left, cartoon representation of the complex. Right, orthogonal view, where the SH2 domain is shown as a surface model, HA4 residues as stick models with residues in the BC, DE, and FG loops in yellow, cyan and green, respectively. Amino acids that differ from the library template are underlined. SH2 surfaces contacted by FG loop 'finger' residues are in pink, and 'palm' contacts are in brown. A putative inorganic phosphate in the pY-binding pocket is shown (red mesh). **(c**) Backbone traces comparing HA4 FG loop residues near the pY pocket with the typical position of a pY-peptide (PDB ID: 1LCJ).[39] The side chains of Tyr87 and Trp80 are shown for HA4, and the pY and Y+3 residues of the pY-peptide. **(d**) Polar interactions (dashed lines) made by 'finger' residues (green sticks with black labels) near the pY pocket. SH2 residues are shown as white sticks and labeled in red. **(e**) Polar contacts made by 'palm' residues. Residues in the monobody scaffold are shown in cyan. Tyr35 is in yellow. Residues are labeled as in (d). **(f**) Changes in the binding free energy ($\Delta\Delta G_{binding}$) for alanine mutants of HA4 residues. Asterisks indicate the lower limit of $\Delta\Delta G_{binding}$ for mutants for which no binding was detected.

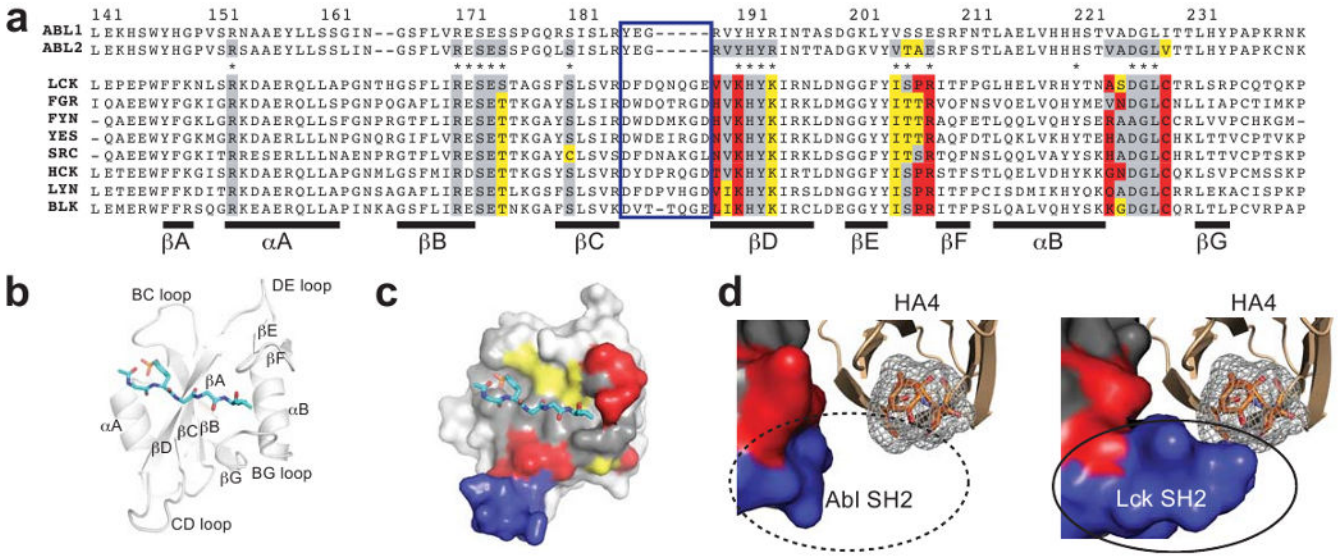


**Figure 4. Structural basis for HA4's specificity toward Abl and Abl2**

**(a**) Amino-acid sequence alignment of Abl, Abl2 and Src family SH2 domains. Residues within 5Å of the HA4 interface are colored as follows: gray, residues where the consensus amino acid among the Src family members is identical to that of Abl; yellow, conservative substitutions; and red, non-conservative substitutions. Residues in the SH2 CD loop are in the blue box. Residues in the peptide-binding interface, inferred from the Lck structure[39] are indicated with asterisks. **(b**) Cartoon model of the Lck SH2 domain structure with a bound peptide (1LCJ).[39] The phosphopetide (sticks) lies across the central strand (βD). The pY-binding pocket and the Y+3 pocket are on either side of βD. **(c**) Conservation of HA4-interacting residues shown on the surface of the Lck SH2 domain. A phosphopeptide (sticks) highlights the overlap between phosphopeptide-binding and HA4-binding interfaces. **(d**) HA4/Abl complex (left) and hypothetical HA4/Lck complex (right, modeled by aligning Lck SH2 with Abl SH2 in the HA4/Abl SH2 complex), emphasizing SH2 CD-loop residues (blue). The two SH2 domains are viewed from an equivalent direction. HA4 is shown as a cartoon model, with the DE loop shown as orange sticks and gray mesh. The CD loop of the Lck SH2 domain that creates a protruding knob and is predicted to clash with the DE loop of HA4 is enclosed in the solid circle. The equivalent region in the Abl SH2 domain is marked with the dotted circle.

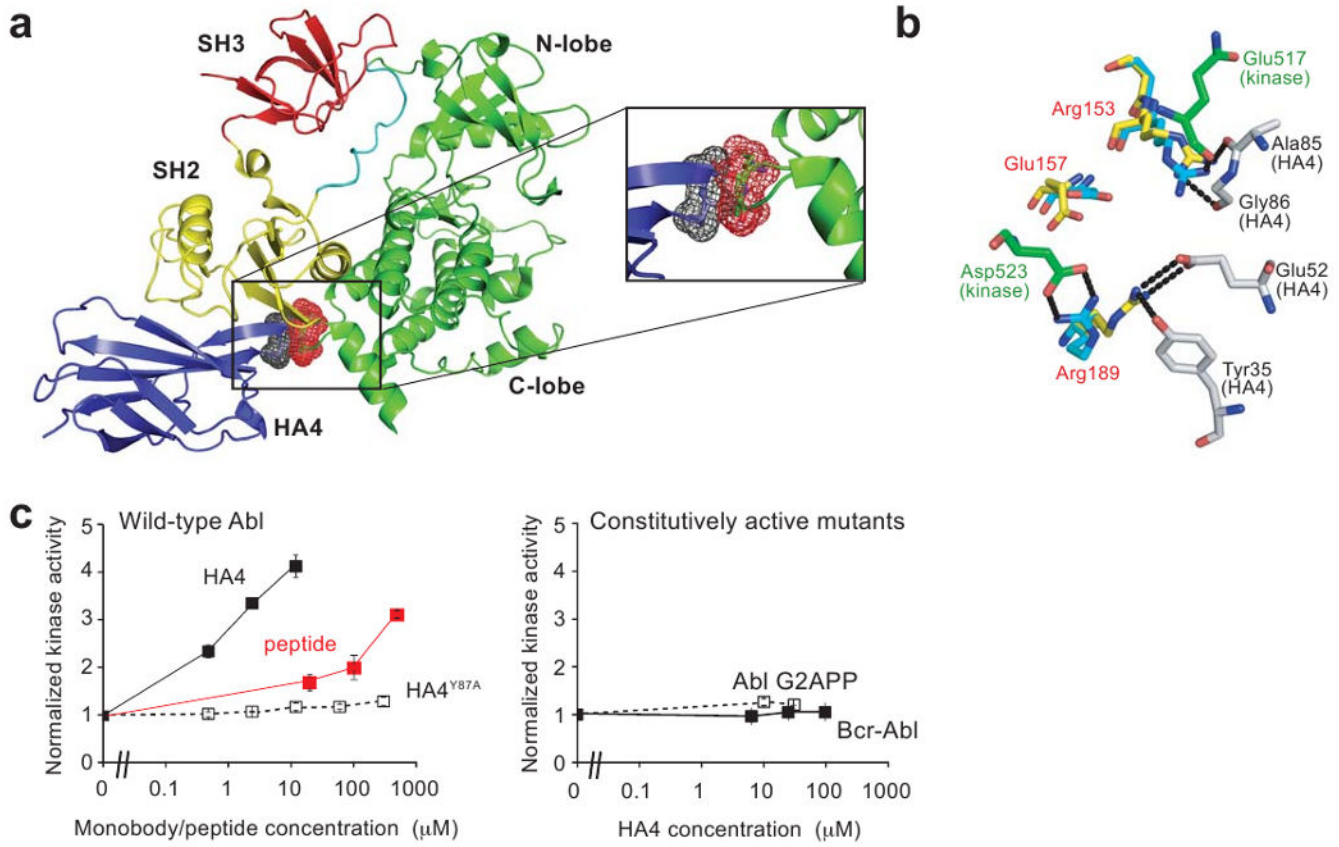


**Figure 5. HA4 activates autoinhibited Abl**

(**a**) Left, an overlay of the HA4/Abl SH2 complex (blue; only HA4 is shown for clarity) and autoinhibited Abl (1OPL)[36] highlighting steric clash (boxed) between the HA4 FG loop (black mesh) and the Abl αI-αI′ helix (red mesh). The SH3, SH2 and kinase domains of Abl are colored in red, yellow and green, respectively. In the blow-up, SH2 is omitted for clarity. (**b**) Mutually exclusive interactions of two SH2 residues (Arg153 and Arg189) between the two structures. Residues in the SH2 and kinase domains of the autoinhibited Abl structure are shown with the carbon atoms in cyan and green, respectively. Residues in the SH2 domain and HA4 in the HA4/SH2 complex are shown with the carbon atoms in yellow and white, respectively. Polar interactions involving Arg153 and Arg189 of the SH2 domain are shown with dashed lines. SH2 residues are labeled in red. (**c**) Effects of HA4 on Abl activation. Left, the activity of the Abl kinase is plotted as a function of HA4 (solid black), phosphopeptide (red), and $HA4_{Y87A}$ (open squares) concentrations. The activity was normalized relative to that in the absence of monobody or peptide. Right, the normalized activity of an autoinhibition-defective mutant of Abl kinase (G2APP: open boxes, dashed line) and the activity of the highly activated Bcr-Abl fusion protein (solid boxes, solid line) plotted as a function of HA4 concentration.

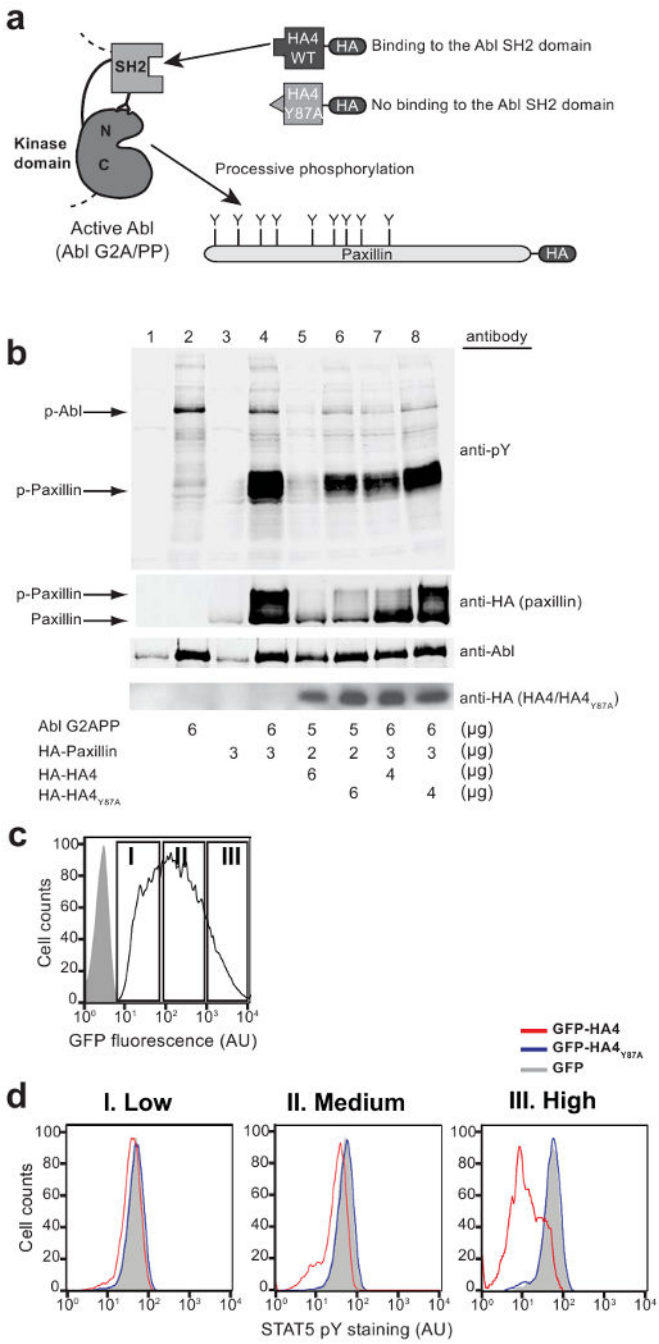


**Figure 6. HA4 blocks processive phosphorylation of an Abl substrate in cells and inhibits STAT5 phosphorylation in leukemia cells**

(**a**) Schematic of constructs used to monitor Abl-mediated processive phosphorylation of paxillin in HEK293 cells. The active Abl mutant, G2APP, was cotransfected with HA4 (or $HA4_{Y87A}$) and paxillin, which contains multiple phosphorylation sites. Paxillin and $HA4/HA4_{Y87A}$ were expressed as HA-tag (gray oval) fusions to facilitate detection in Western analysis. (**b**) Western blots of lysates from HEK293 cells transfected with the indicated amounts of DNA. The putative positions of Abl kinase and paxillin are indicated by arrows. HA4-mediated disruption of paxillin processive phosphorylation is observed,

with a more pronounced effect with lower levels of paxillin (**c**) Histograms showing the numbers of K562 cells transfected with a HA4-green fluorescent protein (GFP) fusion vector (line) or untreated K562 cells (gray area) as a function of GFP fluorescence intensity as measured by flow cytometry. The cells were binned in three classes, low, medium and high, based on GFP fluorescence as indicated. (**d**) Flow cytometric analysis of tyrosine phosphorylation of STAT5 in transfected K562 cells. Histograms depict signal intensities of cells stained with anti-phospho-Stat5 antibody. The fusion proteins used for transfection are indicated. The three panels show results of the three GFP-gated classes as shown in (c).

## Table 1
## Data collection and refinement statistics (molecular replacement)

| | Monobody HA4/Abl1 SH2 complex (3K2M) |
|---|---|
| **Data collection** | |
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 33.63, 88.18, 131.08 |
| α, β, γ (°) | 90, 90, 90 |
| Resolution (Å) | 1.75 (1.81−1.75) |
| $R_{sym}$ | 4.6 (43.8) |
| $I / \sigma I$ | 26.3 (1.9) |
| Completeness (%) | 95.4 (65) |
| Redundancy | 6.1 (2.8) |
| **Refinement** | |
| Resolution (Å) | 1.75 |
| No. reflections | 36480 |
| $R_{work}$ / $R_{free}$ | 0.18/0.22 |
| No. atoms | 3394 |
| Protein | 3089 |
| Ligand/ion | 10 |
| Water | 295 |
| *B*-factors | |
| Protein | 30.2 |
| Ligand/ion | 36.2 |
| Water | 36.1 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.017 |
| Bond angles (°) | 1.500 |

* Values in parentheses are for highest-resolution shell.