Biophysical characterization of the ETV6 PNT domain polymerization interfaces

Chloe A.N. Gerak[1], Sophia Y. Cho[1], Maxim Kolesnikov[2], Mark Okon[1], Michael E.P. Murphy[2], Richard B. Sessions[3], Michel Roberge[1], and Lawrence P. McIntosh[1,4,*]

From the [1]Department of Biochemistry and Molecular Biology, and [2]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada; [3]School of Biochemistry, University of Bristol, Bristol, United Kingdom; and [4]Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada

Edited by Wolfgang Peti

ETV6 is an E26 transformation specific family transcriptional repressor that self-associates by its PNT domain to facilitate cooperative DNA binding. Chromosomal translocations frequently generate constitutively active oncoproteins with the ETV6 PNT domain fused to the kinase domain of one of many protein tyrosine kinases. Although an attractive target for therapeutic intervention, the propensity of the ETV6 PNT domain to polymerize via the tight head-to-tail association of two relatively flat interfaces makes it challenging to identify suitable small molecule inhibitors of this protein–protein interaction. Herein, we provide a comprehensive biophysical characterization of the ETV6 PNT domain interaction interfaces to aid future drug discovery efforts and help define the mechanisms by which its self-association mediates transcriptional repression. Using NMR spectroscopy, X-ray crystallography, and molecular dynamics simulations, along with amide hydrogen exchange measurements, we demonstrate that monomeric PNT domain variants adopt very stable helical bundle folds that do not change in conformation upon self-association into heterodimer models of the ETV6 polymer. Surface plasmon resonance–monitored alanine scanning mutagenesis studies identified hot spot regions within the self-association interfaces. These regions include both central hydrophobic residues and flanking salt-bridging residues. Collectively, these studies indicate that small molecules targeted to these hydrophobic or charged regions within the relatively rigid interfaces could potentially serve as orthosteric inhibitors of ETV6 PNT domain polymerization.

ETV6 is a modular transcriptional repressor of the ETS (E26 transformation specific) family for which head-to-tail polymerization of its PNT (or SAM) domain facilitates cooperative binding to tandem DNA sites by its ETS domain (1, 2). The defining DNA-binding ETS domain is conserved among all ETS transcription factors, whereas the PNT domain is present in approximately one-third of ETS paralogs (3). Unlike the monomeric PNT domains of most ETS factors, the ETV6 PNT domain self-associates in a head-to-tail fashion to form an open-ended, left-handed helical polymer (4, 5). The only other known self-associating PNT domains in the ETS family are those of *Drosophila* Yan (6) and possibly human ETV7 (7–9), as other PNT domains lack suitable interfaces because of amino acid differences or steric blockage (10).

ETV6 is biologically important in embryonic development and hematopoietic regulation (11, 12). Although reported to recruit corepressors such as mSin3A, SMRT, and N-CoR (13), the mechanisms by which ETV6 regulates transcription require further investigation. Polymeric DNA-bound ETV6 is proposed to cause localized chromatin compaction to block access of the transcriptional machinery to target genes. This speculative model is based on the observation that the repeat distance of the helical polymer formed by the PNT domain is comparable to the width of the nucleosome core particle (13).

ETV6 also has preeminent roles in cancer. Frequently, chromosomal translocations fuse gene fragments encoding the PNT domain of ETV6 with the kinase domain of one of many diverse receptor protein tyrosine kinases (PTKs) or to the DNA-binding domain from one of several transcription factors (14). Known receptor PTK fusion partners of ETV6 include PDGFβ, JAK2, FGFR3, and NTRK3. The resulting constitutively active self-associated oncoproteins have been linked to over 40 human leukemias, as well as fibrosarcomas, breast carcinomas, and nephromas (14–16).

Owing to its presence in numerous fusion oncoproteins, the ETV6 PNT domain is an attractive target for therapeutic intervention (15). However, its propensity to form long insoluble polymers via the tight head-to-tail association (K$_D$ $\sim$ nM) of two relatively flat interfaces (13) hinders identification of suitable small molecule inhibitors (17). These interfaces, termed the ML- and EH-surfaces (mid-loop and end-helix, respectively), lie roughly on opposite sides on the globular PNT domain (Fig. S1). Each is composed of a hydrophobic patch encompassed by polar and charged side chains. The introduction of an ionizable residue into either hydrophobic patch yields a monomeric PNT domain as judged by several techniques including equilibrium ultracentrifugation and native gel electrophoresis (10, 13). Examples include the V112E or V112R mutations that disrupt the EH-surface or the

A93D mutation that disrupts the ML-surface. Two such mutant PNT domains with complementary wild-type interfaces can still form a heterodimer. The availability of these monomeric and heterodimeric forms of the PNT domain facilitates studies of ETV6 self-association.

In the case of the ETV6-NTRK3 (EN) fusion oncoprotein, the introduction of monomerizing mutations blocked the ability of EN to polymerize, to activate its PTK, and to transform NIH3T3 cells (18). Furthermore, when co-expressed, the isolated PNT domain had a dominant-negative effect on EN-transformed cells (18). Subsequent studies showed that weakening polymerization by disrupting a peripheral inter-molecular salt bridge (K99-D101) also abrogated the ability of EN to transform NIH 3T3 cells (19). Collectively, these studies demonstrated that inhibiting PNT domain polymerization is indeed a viable therapeutic strategy against ETV6-driven cancers.

Understanding the mechanisms of PNT domain polymerization will yield insights in the transcriptional repression properties of ETV6 and the oncogenic properties of ETV6 fusions. In particular, defining structural and thermodynamic differences between monomeric and heterodimeric forms of the PNT domain and determining "hot spot" regions for self-association are both important for delineating the mechanisms underlying polymerization and for developing strategies to inhibit this process. Herein, using NMR spectroscopy, X-ray crystallography, and molecular dynamics (MD) simulations, we demonstrate that the structures of monomeric ETV6 PNT domain variants do not change significantly upon mutation or self-association into heterodimers. Amide hydrogen exchange (HX) measurements confirmed that the monomeric PNT domain is very stable in solution. Protection against exchange increased upon heterodimerization for amides clustering near the self-association interfaces. Complementary alanine scanning mutagenesis revealed several hot spot regions within these ML- and EH-surfaces. These residues partake in both hydrophobic and electrostatic interactions and can serve as starting points for targeted rational drug design.

Results

PNT domain dimerization characterized by NMR spectroscopy

The NMR spectroscopy can give insights into the thermodynamic, kinetic, and structural mechanisms of protein–ligand interactions. A particularly convenient approach is to use ^{15}N-heteronuclear single quantum coherence (HSQC) spectra to monitor the titration of a ^{15}N-labeled protein with an unlabeled, and hence NMR silent, ligand such as another protein (20). Amide chemical shifts are highly sensitive to even subtle environmental changes, and thus an interaction with the unlabeled species can usually be detected through chemical shift perturbations (CSPs) of the labeled protein. Amides exhibiting CSPs typically cluster around the protein–ligand interface, yet may also be distal if binding causes longer range (allosteric) structural changes (20).

To begin this interrogation, purified samples of uniformly ^{13}C/^{15}N-labeled ETV6 fragments (residues 40–125; Table S1)

that contain either an A93D mutation or V112E mutation (henceforth described as the A93D-PNT or V112E-PNT domains, respectively) were prepared for NMR spectroscopic characterization. Under neutral pH solution conditions, the A93D-PNT and V112E-PNT domains yielded well-dispersed NMR spectra indicative of stably folded structures (Fig. 1A). However, the latter showed some propensity to self-associate, and improved spectra were obtained at a sample pH value of 8.0. Presumably, this reflects the deprotonation of E112, which may have an anomalously high pK_a value when buried at the polymer interface (21, 22). Unfortunately, the more alkaline conditions resulted in some loss of signal intensity because of base-catalyzed HX.

Upon addition of the unlabeled V112E-PNT domain to the ^{15}N-labeled A93D-PNT domain, many amides exhibited CSPs in the slow exchange regime (Figs. 1B and S2). That is, at an intermediate titration point, separate ^{1}H^N-^{15}N peaks corresponding to the unbound (monomeric) and bound (heterodimeric) forms of the labeled protein were observed. This is consistent with the previously reported K$_D$ value of ∼2 nM for the high-affinity binding equilibrium (13). Comparable results were observed for the reciprocal titration of the unlabeled A93D-PNT domain to the ^{15}N-labeled V112E-PNT domain (Fig. 1B). Although the oligomerization states of the PNT domains in these experiments were not directly determined, the results are entirely consistent with previous studies showing that the A93D-PNT and V112E-PNT domains are monomeric when separated and heterodimeric when combined (13).

Through a combination of scalar and NOE correlation experiments, NMR signals were assigned from most main chain ^{1}H, ^{13}C, and ^{15}N nuclei of the V112E-PNT and A93D-PNT domains in their monomeric forms and as heterodimers with their unlabeled partners (Figs. S3–S6). Utilizing the motif identification from chemical shift algorithm (23), the secondary structural elements for the four species were predicted from these chemical shifts. In each case, four distinct helical regions were detected (Fig. 1C). These coincide well with the four α-helices (H1: R63-E76; H2: G91-L94; H3: K99-R105; H4: G110-K122) identified in the X-ray crystal structures of the self-associated ETV6 PNT domains (PDB: 1JI7 and 1LKY) by PDBsum (24). Matching two short N-terminal 3_{10}-helices observed in these crystal structures, residues A52-L54 and, to a lesser extent, P58-Y60 also have chemical shifts indicative of helical character. In contrast, such diagnostic chemical shifts were not seen for residues S84-T86 even though they are classified as a forming a 3_{10}-helix in a subset of the monomer subunits of PDB file 1LKY. This minor discrepancy may arise because these residues are within an extended, solvent exposed polypeptide segment between helices H1 and H2. Amides in this region have chemical shift-derived random coil index-squared order parameter values (RCI-S^2, a proxy for backbone dynamics (25)) indicative of increased flexibility relative to the well-ordered helices (Fig. 1C). Most importantly, these analyses demonstrated that the A93D-PNT and V112E-PNT domains have very similar secondary structures in their monomeric and heterodimeric forms. Thus, the proteins in solution do not undergo any significant conformational
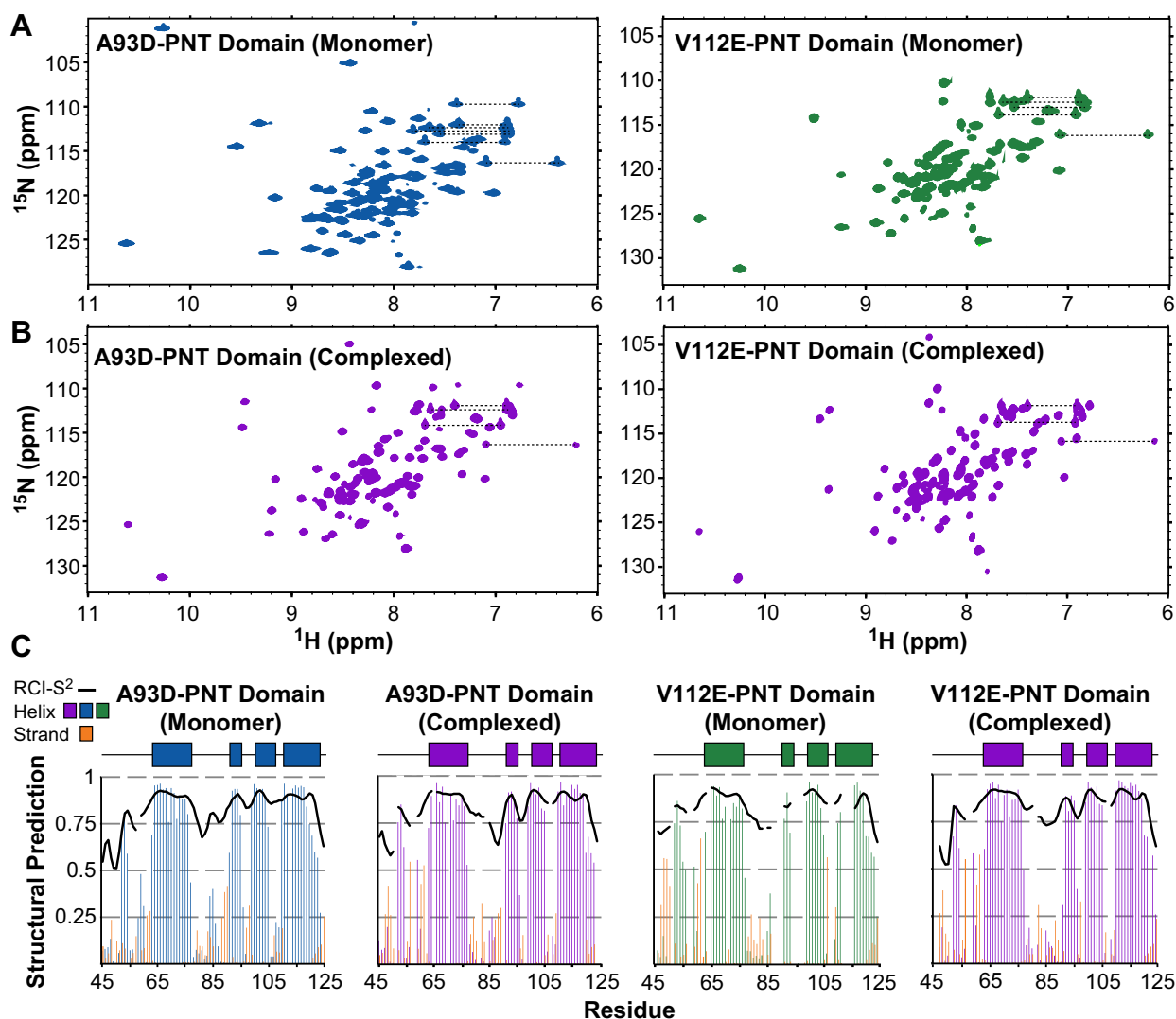
Figure 1. The PNT domains adopt stable folded structures in their monomeric and heterodimeric states. The ^{15}N-HSQC spectra of the ^{15}N-labeled A93D-PNT domain (*left*) and V112E-PNT domain (*right*) in the absence (*A*) and presence (*B*) of unlabeled partner PNT domain. Assignments and sample conditions are provided in Figures S3–S6. Apparent differences in ^{15}N-HSQC linewidths result from the use of different conditions, spectral acquisition parameters, and contour display parameters. *C*, The predicted secondary structural elements and RCI-S^2 values (*black lines*; decreasing values from 1 to 0 indicate increasing flexibility), calculated with the MICS algorithm, are plotted for the PNT domains in their monomeric and complexed states. Missing data correspond to residues lacking chemical shift assignments. The locations of the four α-helices (H1-H4) observed in the X-ray crystal structures of the ETV6 PNT domain are indicated as rectangles above each plot. RCI-S^2, random coil index-squared order parameter; MICS, motif identification from chemical shift.

changes upon mutation or association into the structures previously characterized by X-ray crystallography.

Armed with chemical shift assignments, the amide ^{1}H^N-^{15}N CSPs resulting from PNT domain dimerization were readily calculated. Of note, residues in the ^{15}N-labeled A93D-PNT and V112E-PNT domains that experienced the greatest spectral perturbations cluster within the EH- and ML-surfaces, respectively (Fig. 2). This confirms that, as seen by X-ray crystallography, the two monomerized PNT domains indeed associate in solution through their wild-type interfaces.

Crystallographic comparison of monomeric and dimeric PNT domains

In their original studies, the Bowie group obtained crystals of the V112E-PNT domain (PDB: 1JI7, *C*2 space group, three monomers in the asymmetric unit) (13). Despite burial of

E112, the monomers assembled in the crystal lattice *via* their ML- and EH-surfaces to form an extended helical polymer with an approximate 6$_5$ screw symmetry (Fig. 3*A*). Subsequently, they determined the structure of a heterodimer composed of a A93D-PNT domain bound to a V112R-PNT domain *via* their complementary wild-type interfaces (PDB: 1LKY, *P*1 space group, 3 heterodimers in the asymmetric unit) (5). Although no longer polymeric within the crystal lattice, a model built from PDB: 1LKY using appropriate monomers subunits with native interfaces closely matched the polymeric structure of PDB: 1JI7 with the V112 E substitution. Thus, the latter serves as a reliable experimental structure of the ETV6 PNT domain polymer.

The NMR spectroscopic studies presented above indicate that the secondary structures and binding interfaces of the A93D-PNT and V112E-PNT domains in solution closely
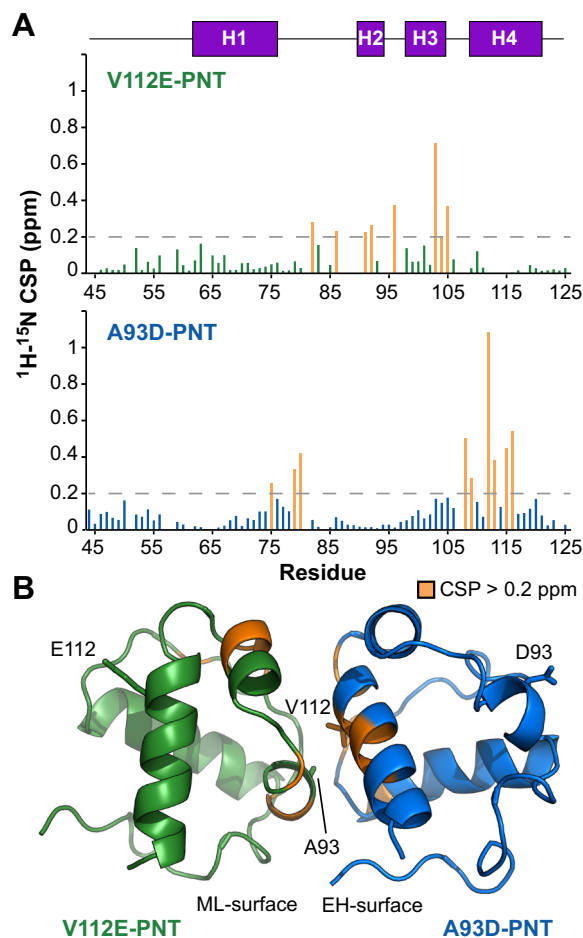
Figure 2. PNT domain dimerization interface identified by amide chemical shift perturbations. *A,* Backbone amide $^1H^N$-^{15}N CSPs resulting from the heterodimerization of the V112E-PNT (*top*) and A93D-PNT (*bottom*) domains. Missing data correspond to prolines or residues without assigned NMR signals in both protein states. Most residues showed small CSPs which may be due in part to the different conditions under which spectra were assigned (V112E-PNT domain monomer, pH 8.0; V112E-PNT domain complexed with the A93D-PNT domain, pH 7.5; A93D-PNT domain monomer, pH 7.0; A93D-PNT domain complexed with the V112E-PNT domain, pH 7.0). *B,* However, amides with CSP values >0.2 ppm, which are highlighted in *orange* on a model of the V112E- (*green*)/A93D- (*blue*) PNT domain heterodimer, map to the interfacial regions. The cartoon is derived from PDB: 1LKY with a V112R-PNT domain. CSP, chemical shift perturbation.

Figure 3. Structural comparison of the monomeric and polymeric ETV6 PNT domain variants. *A,* The X-ray crystallographic structure of the V112E-PNT domain from PDB: 1JI7. The three subunits of the V112E-PNT domain in the asymmetric unit (*green*) form a helical polymer with neighboring subunits (*gray*). The mutated E112 side chain (EH-surface) is shown as *red* spheres and the wild-type A93 (ML-surface) as *blue* spheres. *B,* The structure of the A93D-V112E-PNT domain determined herein. The two A93D-V112E-PNT domains in the asymmetric unit are highlighted in *orange* (E112 as red spheres, D93 as *purple* spheres), and neighboring asymmetric units are in *gray*. The crystal contacts are unlike those in the polymeric structure and do not involve the ML- or EH-surfaces. *C,* Structural overlay of an A93D-PNT domain subunit that was in complex with a V112R-PNT domain (*blue,* PDB: 1LKY chain B) and the A93D-V112E-PNT domain (*orange*), determined herein. Residues at positions 93 and 112 are highlighted. In contrast to well-aligned interior side chains, variations in the rotamer conformations of surface side chains and the N-terminal residues can be seen.

resemble those observed by X-ray crystallography. However, the PNT domain in a complexed form may have subtle structural differences relative to a monomeric form. To examine this possibility, a variant with both A93D and V112 E mutations was generated and found to crystallize in 2.8 M sodium acetate at pH 7.0. Its structure was solved to 1.86 Å resolution using molecular replacement (Table S2). The A93D-V112E-PNT domain crystallized in the space group $P6_522$ with two monomers in the asymmetric unit (Fig. 3B). Most importantly, the presence of both monomerizing mutations prevented any intermonomer interactions within the crystal lattice *via* the ML- and EH-surfaces. Rather, nearest neighbor contacts were *via* alternative interfaces that are not relevant to polymerization of the wild-type PNT domain.

Regardless of differing crystallization conditions and mutations, the structure of the A93D-V112E-PNT domain closely
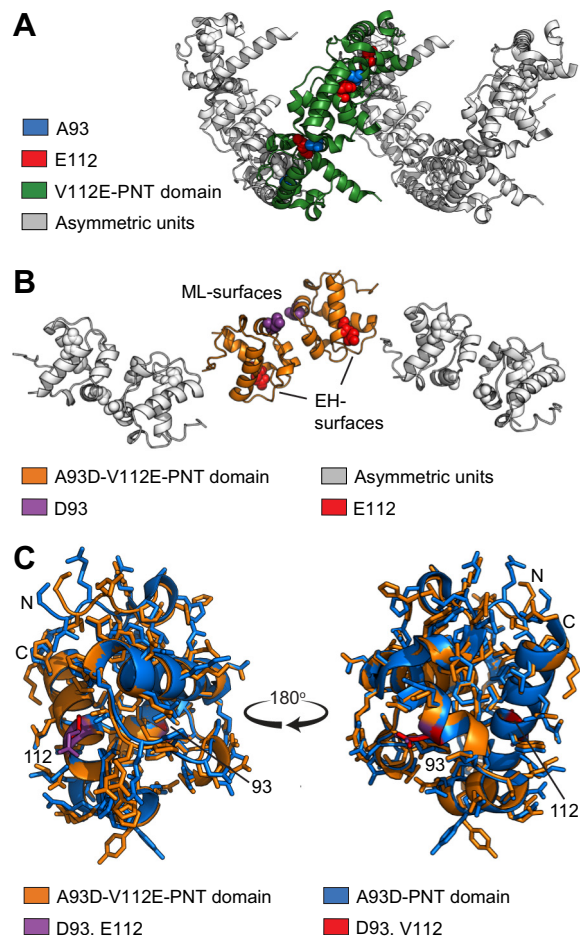
resembles those previously determined for other ETV6 PNT domain variants. Using the DALI server to compare one subunit from this structure with a A93D-PNT domain subunit from PBD entry 1LKY, a total of 77 residues were aligned with a root mean squared deviation (RMSD) value of 0.7 Å and Z-score of 16.8 (26). However, a few subtle differences can be seen upon detailed comparison (Fig. 3C). For example, residues N-terminal to helix H1 have variable conformations. This is consistent with their RCI-S^2 scores indicating a degree of flexibility (Fig. 1C). Not unexpectedly, several surface residues adopted different side chain rotamer conformations, whereas residues within the interior hydrophobic core of the PNT domain superimposed well. Most importantly, the local structural features of the ML- (including residue 93) and

EH- (including residue 112) surfaces do not differ despite the presence or absence of monomerizing mutations or their association upon heterodimer or polymer formation. Thus, interactions of the ETV6 PNT domain do not contribute to any discernible conformational changes between its monomeric, heterodimeric, or polymeric forms.

Amide hydrogen exchange data show increased protection of interfacial residues upon dimerization

Amide HX is a useful technique to characterize protein structure, stability, and dynamics, as well as identifying ligand-binding interfaces (27–29). Through a continuum of local to global conformational fluctuations, main chain amide hydrogens are constantly exchanging with the hydrogens of solvent water (30). If a labile amide proton exchanges for a deuteron, it

will become silent for ^{1}H-detected NMR, and thus, its signal will disappear from an ^{15}N-HSQC spectrum. The rate at which it disappears is determined by its structural features (*e.g.,* hydrogen bonding and solvent accessibility), as well as the experimental conditions (*e.g.,* pH and temperature). To account for the latter, the observed exchange rate constant can be compared with the predicted rate constant for a random coil polypeptide with the same sequence and under the same conditions. The ratio of the predicted versus observed rate constants is the protection factor (PF).

To gain further insights into the ETV6 PNT domain dynamics, NMR spectroscopy was used to measure the amide PFs of its monomeric and heterodimeric forms. By comparing the spectra recorded after 3 days of exchange (Fig. 4A), it is immediately obvious that more amides were
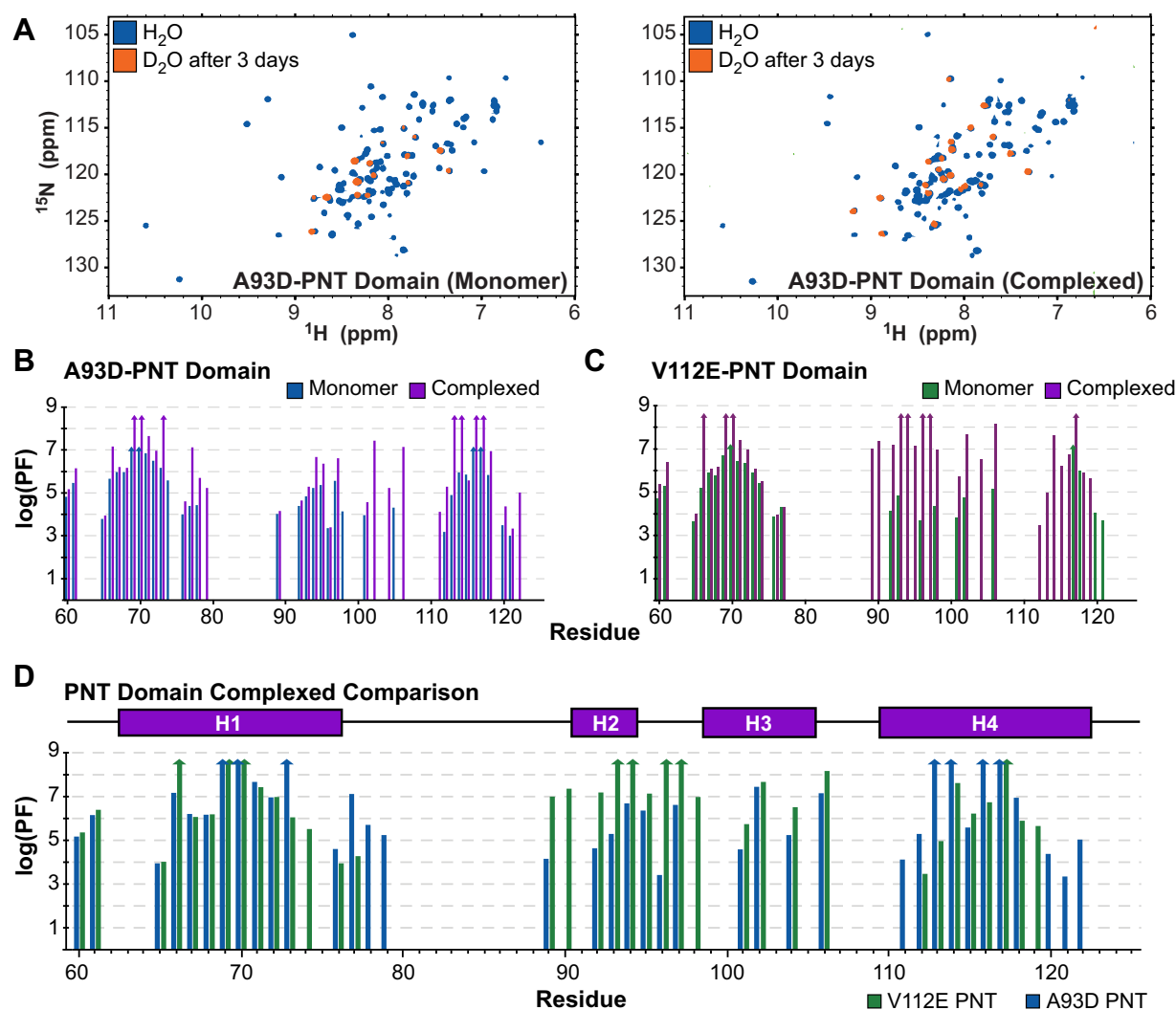

Figure 4. PNT domain amide HX protection factors increase upon heterodimerization. *A,* Overlaid ^{15}N-HSQC spectra for the monomeric (*left*) and complexed (*right*) forms of the A93D-PNT domain in H$_2$O (*blue*) and after 3 days at 21°C in D$_2$O buffer (*orange*). More amides in the heterodimeric species were protected from HX. A summary of the protection factors (PFs) for the (*B*) A93D-PNT and (*C*) V112E-PNT domains in their monomeric and complexed states (values in Table S3). Missing data correspond to prolines, residues with unassigned or overlapping amide chemical shifts, or residues that exchanged before collection of the first ^{15}N-HSQC spectrum after transfer into D$_2$O buffer and thus have log(PF) values less than an estimated upper limit of ∼3. The latter include amides preceding residue 60, which were not observed for any protein after transfer into D$_2$O buffer. In the cases of amides with that had not exchanged significantly after 3 months, estimated lower limits of log(PF) values >7 (monomeric) and >8.5 (heterodimeric) are indicated by upwards arrows. *D,* A comparison of the amide PFs of the V112 E PNT-domain and A93D-PNT domains when associated within a heterodimer. The most protected amides are located near the complementary wild-type ML- (containing A93) and EH- (containing V112) interfaces of the two species, respectively (Fig. 5). The helical regions are indicated with rectangles. HX, hydrogen exchange; HSQC, heteronuclear single quantum coherence.

protected from HX in the heterodimeric versus monomeric species. Exchange rate constants for most amides were determined by fitting their time-dependent $^1H^N$-^{15}N signal intensities to single exponential decays. These values were converted into the PFs summarized in Figure 4, *B* and *C* and tabulated in Table S3.

The monomeric A93D-PNT and V112E-PNT domains exhibited very similar patterns of amide HX. All residues N-terminal to Y60 and most of those in interhelical regions exchanged rapidly under these experimental conditions. This is consistent with their surface exposure and general lack of intramolecular hydrogen bonding interactions (Table S3). Conformational flexibility of these residues was also indicated by their RCI-S^2 values (Fig. 1C). Conversely, amides within or near the four α-helices showed substantial protection from HX. In particular, W69 and L70 in helix H1 and L116 and L117 in helix H4 of both proteins exchanged very slowly, with several of these residues having log(PF) values >7. Under the commonly observed EX2 conditions, with pH-dependent bimolecular kinetics, PFs reflect the residue-specific free energy changes, $\Delta G_{HX} = 2.303RTlog(PF)$, governing local or global conformational equilibria leading to exchange (31). Assuming that these most protected amides exchange through global or near-global structural fluctuations, these HX data provide an estimation of the unfolding free energy for each monomeric PNT domain of >40 kJ/mol. Such a value is consistent with the view that, even without polymerizing, the ETV6 PNT domain adopts a very stable folded conformation. This further indicates that the A93D and V112 E mutations prevent polymerization without disrupting the structure or stability of the monomeric PNT domain.

Heterodimerization resulted in increased HX protection for many residues in both the A93D-PNT and V112E-PNT domains (Fig. 4, *B* and *C*). Indeed, several amides in both proteins did not exchange significantly even after 3 months in D$_2$O buffer and thus have log(PF) > 8.5 (and ΔG_{HX} > 48 kJ/mol). Although global stabilization upon heterodimer formation is expected, many of the residues with at least a 100-fold increase in HX protection localized around the complementary interfacial regions of the two PNT domains (Figs. 4D and 5, *A* and *B*). These are exemplified by amides within or near the wild-type ML-surface of the V112E-PNT domain (N90, K92, A93, L96, T98, D101, F102) and the wild-type EH-surface of the A93D-PNT domain (F77, V112, L113, Y114). Given that the structures of the A93D-PNT and V112E-PNT domains do not change significantly upon heterodimerization, the increased protection of interfacial residues against HX may result from their local stabilization against conformational fluctuations allowing exchange. Alternatively, if exchange occurs predominantly through transient monomers, then the increased protection of amides in the heterodimer would reflect the equilibrium population distribution of these two species (27). Regardless of mechanism, the HX data are consistent with the role of these interfaces in ETV6 PNT domain polymerization.

Alanine scanning mutagenesis at the PNT domain self-association interface

Alanine scanning mutagenesis was used to identify which residues at the PNT domain heterodimer interface contribute most to binding affinity. The mutation of a residue to alanine reduces its side chain to a single methyl group, thereby eliminating its contributions to intermolecular binding, while also avoiding the introduction of any additional non-native interactions. Alanine was chosen over glycine as the latter may lead to increased backbone flexibility. For these studies, surface plasmon resonance (SPR) was used to characterize the binding interaction as this technique is rapid and uses only small quantities of bacterially expressed proteins.

Initial controls were performed to reproduce the results of previously reported SPR studies of the PNT domain dimerization (13). Either the biotinylated A93D-PNT or V112E-PNT domain was immobilized on a streptavidin chip, and the complementary (positive control) or same (negative control) PNT domain was applied as the analyte at various concentrations (Fig. S7). High affinity binding was seen for the heterodimer pairings, whereas the identical PNT domains did not measurably interact. The A93D-PNT domain analyte bound the V112E-PNT domain ligand with a fit K$_D$ value of 7.5 nM, and the V112E-PNT domain analyte bound the A93D-PNT domain ligand with a fit K$_D$ value of 5.1 nM. These results agreed well with previously reported K$_D$ values of 1.7 to 4.4 nM for the PNT domain interactions as measured by SPR (13) and isothermal titration calorimetry (19). Thus, SPR was used to characterize the effects of alanine substitutions of 18 residues within or around the EH-surface of the A93D-PNT domain and 14 residues within or around the ML-surface of the V112E-PNT domain (Tables 1 and 2).

Six out of 18 on the A93D-PNT domain and seven out of 14 tested residues on the V112E-PNT domain had large detrimental effects on binding ($\Delta\Delta G$ >6 kJ/mol) and can be classified as hot spots (Fig. 6A). Although more difficult to interpret than K$_D$ values, these mutations acted to varying degrees by slowing the association rate constants, k$_{on}$, and/or increasing the dissociation rate constants, k$_{off}$. The higher proportion of hotspot interfacial residues on the V112E-PNT domain, including the three most detrimental of the entire set of alanine mutants, suggests that the features of this interface contribute most significantly to PNT domain polymerization.

Many of the hydrophobic residues at the center of the heterodimer interfaces are hotspots for binding. In particular, the L96 A mutation on the ML-surface of the V112E-PNT domain resulted in an ~1000-fold decrease in affinity. As shown in Figure 6B, L96 protrudes out of the V112E-PNT domain as a "bump" to fit into the "hole" between residues F77, V112, E115, and H119 on the EH-surface of the A93D-PNT domain. Thus, in addition to contributing to heterodimerization through the hydrophobic effect, L96 partakes in favorable van der Waals interactions at the protein–protein interface. Other hydrophobic residues that are important
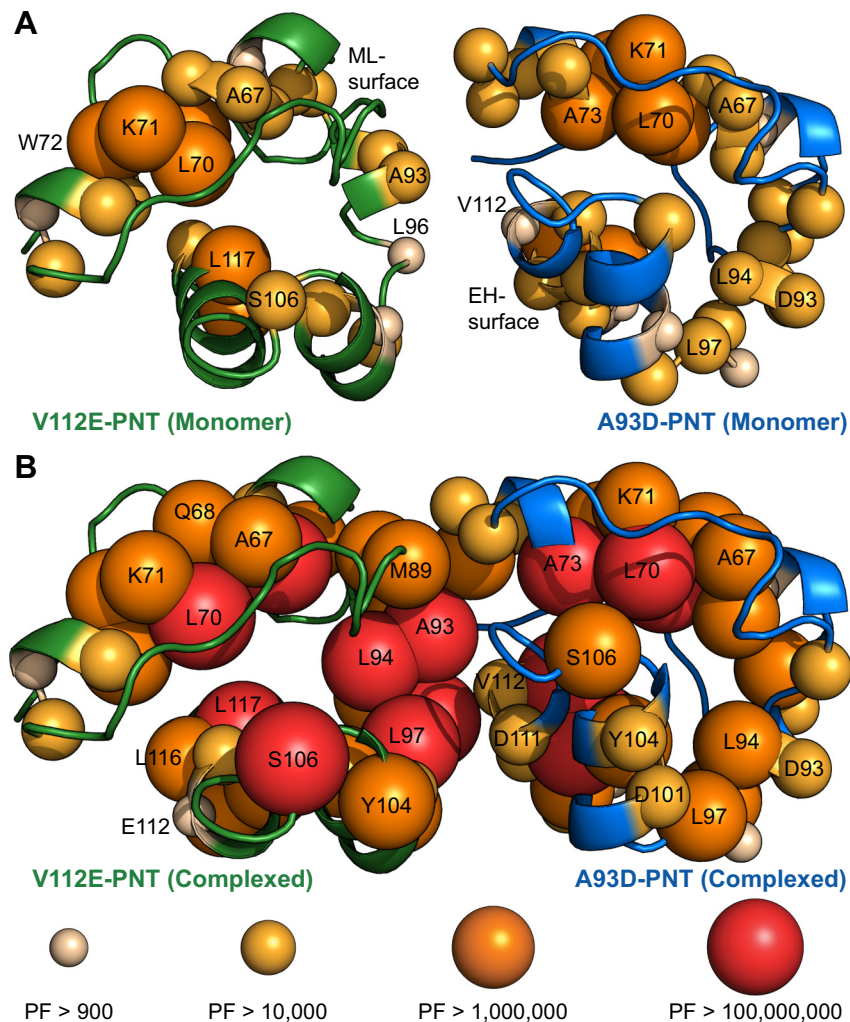
Figure 5. Mapping of HX protection factors on the monomeric and complexed PNT domain structures. Amide protection factors (PFs) are encoded as spheres on the ribbon diagrams (from PDB: 1LKY) of the V112E-PNT (*green*) and A93D-PNT (*blue*) domains in their (*A*) monomeric and (*B*) heterodimeric forms according to the indicated color and size scheme. Residues without spheres are either prolines, lack an assigned or fully resolved $^1H^N$-^{15}N signal, or exchanged too fast for reliable HX quantitation. Both monomeric PNT domains showed similar HX profiles, with the most protected amides located in structured helical regions. Upon association, the PFs of many amides at the complementary dimer interfaces increased. HX, hydrogen exchange.

include L97 and M89 on the ML-surface and F77, L79, and V112 on the EH-surface.

Surrounding the hydrophobic center of the heterodimer interface is a ring of residues that generally have moderate contributions toward binding affinity (Figure 6*B*). However, an important exception is the peripheral K99 on the A93D-PNT domain EH-surface. This residue forms an intermolecular salt bridge with D101 on the ML-surface of the V112E-PNT domain. Alanine mutations of K99 or D101 resulted in decreased binding affinities of 930 nM or 400 nM, respectively. Thus, consistent with their salt bridge pairing, either mutation weakened binding relative to the wild-type by a factor of ~100. Mutation of K99 has been shown to interfere with EN oncogenic cellular transformation, confirming the functional importance of this salt bridge (19). A second intermolecular salt bridge consistently seen in the PNT domain heterodimer structures involves R105-D111. The R105 A and D111 A mutations also severely weakened binding to K_D values of 2.3 μM and 400 nM, respectively, thus confirming the importance of this interaction. In contrast,

several additional salt bridges involving K99-E100, R103-E100, and R103-D101 are seen in some, but not all crystallographically defined interfaces. However, alanine substitutions of either E100 or R103 did not significantly impact binding. Thus, the K99-D101 and R105-D111 salt bridges play important roles in PNT domain association, whereas those involving E100 or R103 do not. It is also notable that the polar residue N90 on the ML-surface of the V112E-PNT domain surface is also a hotspot, but it does not appear to have a potential reciprocal hydrogen bond donor or acceptor. Overall, the alanine scanning mutagenesis study illustrated that PNT domain dimerization is a result of interactions involving both hydrophobic and charged interfacial residues.

Molecular dynamics simulations of the PNT monomer and dimer

MD simulations were run on the A93D-V112E-PNT domain monomer, reported herein, and the PNT domain heterodimer, comprising an A93D-PNT domain and

Table 1
Alanine scanning mutagenesis of the EH-surface on the A93D-PNT domain[a]

Ala mutation	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_D (nM)	$\Delta\Delta G$ (kJ/mol)[b]
None	$(2.0 \pm 0.03) \times 10^5$	$(1.5 \pm 0.01) \times 10^{-3}$	7.5 ± 0.10	
S47A	$(2.1 \pm 0.02) \times 10^5$	$(1.4 \pm 0.01) \times 10^{-3}$	6.9 ± 0.09	-0.2 ± 0.05
I48A	$(3.4 \pm 0.03) \times 10^5$	$(4.0 \pm 0.03) \times 10^{-3}$	12 ± 0.15	1.2 ± 0.05
E76A	$(6.2 \pm 0.10) \times 10^5$	$(6.2 \pm 0.08) \times 10^{-3}$	9.9 ± 0.20	0.7 ± 0.06
F77A	$(3.4 \pm 0.11) \times 10^4$	$(1.5 \pm 0.04) \times 10^{-2}$	450 ± 21	10.2 ± 0.12
S78A	$(2.8 \pm 0.58) \times 10^6$	$(9.3 \pm 1.8) \times 10^{-3}$	3.3 ± 0.96	-2.0 ± 0.71
L79A	$(8.1 \pm 0.10) \times 10^3$	$(1.6 \pm 0.03) \times 10^{-3}$	200 ± 4	8.2 ± 0.06
R80A	$(9.1 \pm 0.11) \times 10^4$	$(2.9 \pm 0.02) \times 10^{-3}$	32 ± 0.4	3.6 ± 0.05
K99A	$(2.4 \pm 0.02) \times 10^3$	$(2.3 \pm 0.01) \times 10^{-3}$	930 ± 9	12.0 ± 0.04
E100A	$(3.1 \pm 0.02) \times 10^5$	$(1.5 \pm 0.01) \times 10^{-3}$	5 ± 0.04	-1.0 ± 0.04
R103A	$(2.5 \pm 0.27) \times 10^5$	$(5.4 \pm 0.48) \times 10^{-3}$	21 ± 3	2.6 ± 0.35
P107A	$(2.6 \pm 0.03) \times 10^5$	$(2.1 \pm 0.01) \times 10^{-3}$	8.1 ± 0.10	0.2 ± 0.05
H108A	$(2.6 \pm 0.02) \times 10^5$	$(1.4 \pm 0.005) \times 10^{-3}$	5.4 ± 0.04	-0.8 ± 0.04
D111A	$(3.3 \pm 0.04) \times 10^3$	$(1.4 \pm 0.02) \times 10^{-3}$	420 ± 8	9.9 ± 0.06
V112A	$(2.0 \pm 0.06) \times 10^4$	$(1.1 \pm 0.02) \times 10^{-2}$	550 ± 19	10.7 ± 0.09
Y114A	$(3.5 \pm 0.04) \times 10^4$	$(4.5 \pm 0.03) \times 10^{-3}$	130 ± 2	7.1 ± 0.05
E115A	$(2.3 \pm 0.08) \times 10^5$	$(2.5 \pm 0.06) \times 10^{-3}$	11 ± 0.5	1.0 ± 0.12
L116A	$(2.3 \pm 0.004) \times 10^5$	$(7.6 \pm 0.02) \times 10^{-4}$	3.3 ± 0.01	-2.0 ± 0.04
H119A	$(2.7 \pm 0.02) \times 10^5$	$(3.4 \pm 0.02) \times 10^{-3}$	13 ± 0.12	1.4 ± 0.04

SPR, surface plasmon resonance.
[a] All A93D-PNT domain analytes were run on the streptavidin SPR chip bound with the biotinylated V112E-PNT domain ligand. Tabulated are fit values of k_{on} and k_{off} with standard errors, and derived values of K_D and $\Delta\Delta G$ with propagated errors.
[b] Calculated as $\Delta\Delta G = RT \ln(K_{D,mutant}/K_{D,wild\text{-}type})$ where wild-type is the top-listed protein with an unmodified interface.

V112R-PNT domain from PBD: 1LKY. Both the monomer and the heterodimer remained structurally stable throughout the 1 μs simulations with no drift in the RMSDs of backbone atoms after equilibration (Fig. 7A). Importantly, the PNT domain subunits remained associated with each other, indicating that the heterodimer can be modeled in silico. The transition occurring around 150 ns in the monomer trajectory corresponds to a rearrangement of its N-terminal residues. Indeed, the RMSD values of these ETV6 truncations (residues 47–123) are dominated by motions of their termini. Restricting the analyses to residues 57 to 120 reduced the backbone RMSD values to ∼1 Å. This further demonstrates that, regardless of mutation or association state, the PNT domain core helical bundle is relatively rigid.

Plots of the all-atom root mean squared fluctuations (RMSFs) of each residue over the trajectories are similar for the monomeric and heterodimeric PNT domains, again showing low mobility in the core and higher mobility at the termini (Fig. 7B). This is consistent with the RCI-S^2 values presented in Figure 1C. Principal component analysis of the trajectories was also performed to explore low frequency conformational motions. In each case, the proportion of the motion (variance) was largely confined to the first few principal components (Fig. S8). Inspection of the first two principal components (Videos S1–S4) shows that these motions in both the monomeric and heterodimeric PNT domains mostly comprise random conformational changes of their N- and C-terminal segments.

A closer look at the data presented in Figure 7B shows that the RMSF values of residues near the termini of the A93D-PNT domain in the heterodimer are lower than those of both its V112R-PNT domain partner and the monomeric A93D-V112E-PNT domain. This can be better seen in a plot of the difference in RMSF values between corresponding residues in the two components of the heterodimer (Fig. 7C). When mapped onto the structure of the heterodimer, residues with

Table 2
Alanine scanning mutagenesis of the ML-surface on the V112E-PNT domain[a]

Ala mutation	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_D (nM)	$\Delta\Delta G$ (kJ/mol)[b]
None	$(4.4 \pm 0.09) \times 10^5$	$(2.3 \pm 0.03) \times 10^{-3}$	5.1 ± 0.13	
I59A	$(4.3 \pm 0.03) \times 10^5$	$(2.5 \pm 0.01) \times 10^{-3}$	5.8 ± 0.05	0.3 ± 0.06
R63A	$(3.6 \pm 0.04) \times 10^5$	$(5.4 \pm 0.05) \times 10^{-3}$	15 ± 0.2	2.7 ± 0.07
N85A	$(3.5 \pm 0.04) \times 10^5$	$(3.4 \pm 0.03) \times 10^{-3}$	9.9 ± 0.13	1.6 ± 0.07
E88A	$(5.0 \pm 0.03) \times 10^5$	$(4.0 \pm 0.02) \times 10^{-3}$	8.1 ± 0.07	1.1 ± 0.06
M89A	$(9.7 \pm 0.16) \times 10^4$	$(3.3 \pm 0.03) \times 10^{-2}$	340 ± 6	10.4 ± 0.08
N90A	$(7.8 \pm 0.20) \times 10^4$	$(3.7 \pm 0.10) \times 10^{-2}$	480 ± 17	11.3 ± 0.11
K92A	$(3.5 \pm 0.06) \times 10^5$	$(2.2 \pm 0.04) \times 10^{-2}$	62 ± 1.6	6.2 ± 0.09
L96A	$(9.5 \pm 1.3) \times 10^3$	$(9.5 \pm 1.1) \times 10^{-3}$	1000 ± 180	13.1 ± 0.45
L97A	$(9.8 \pm 0.16) \times 10^4$	$(6.8 \pm 0.06) \times 10^{-2}$	700 ± 13	12.2 ± 0.08
T98A	$(6.2 \pm 0.05) \times 10^5$	$(5.4 \pm 0.04) \times 10^{-3}$	8.8 ± 0.09	1.3 ± 0.07
E100A	$(5.0 \pm 0.04) \times 10^5$	$(3.5 \pm 0.02) \times 10^{-3}$	7 ± 0.07	0.8 ± 0.07
D101A	$(6.3 \pm 0.12) \times 10^3$	$(2.5 \pm 0.04) \times 10^{-3}$	400 ± 10	10.8 ± 0.09
Y104A	$(7.3 \pm 0.10) \times 10^5$	$(1.5 \pm 0.02) \times 10^{-2}$	21 ± 0.37	3.5 ± 0.08
R105A	$(2.3 \pm 0.11) \times 10^3$	$(5.4 \pm 0.11) \times 10^{-3}$	2340 ± 120	15.2 ± 0.14

SPR, surface plasmon resonance.
[a] All V112E-PNT domain analytes were run on the streptavidin SPR chip bound with the biotinylated A93D-PNT domain ligand. Tabulated are fit values of k_{on} and k_{off} with standard errors, and derived values of K_D and $\Delta\Delta G$ with propagated errors.
[b] Calculated as $\Delta\Delta G = RT \ln(K_{D,mutant}/K_{D,wild\text{-}type})$ where wild-type is the top-listed protein with an unmodified interface.
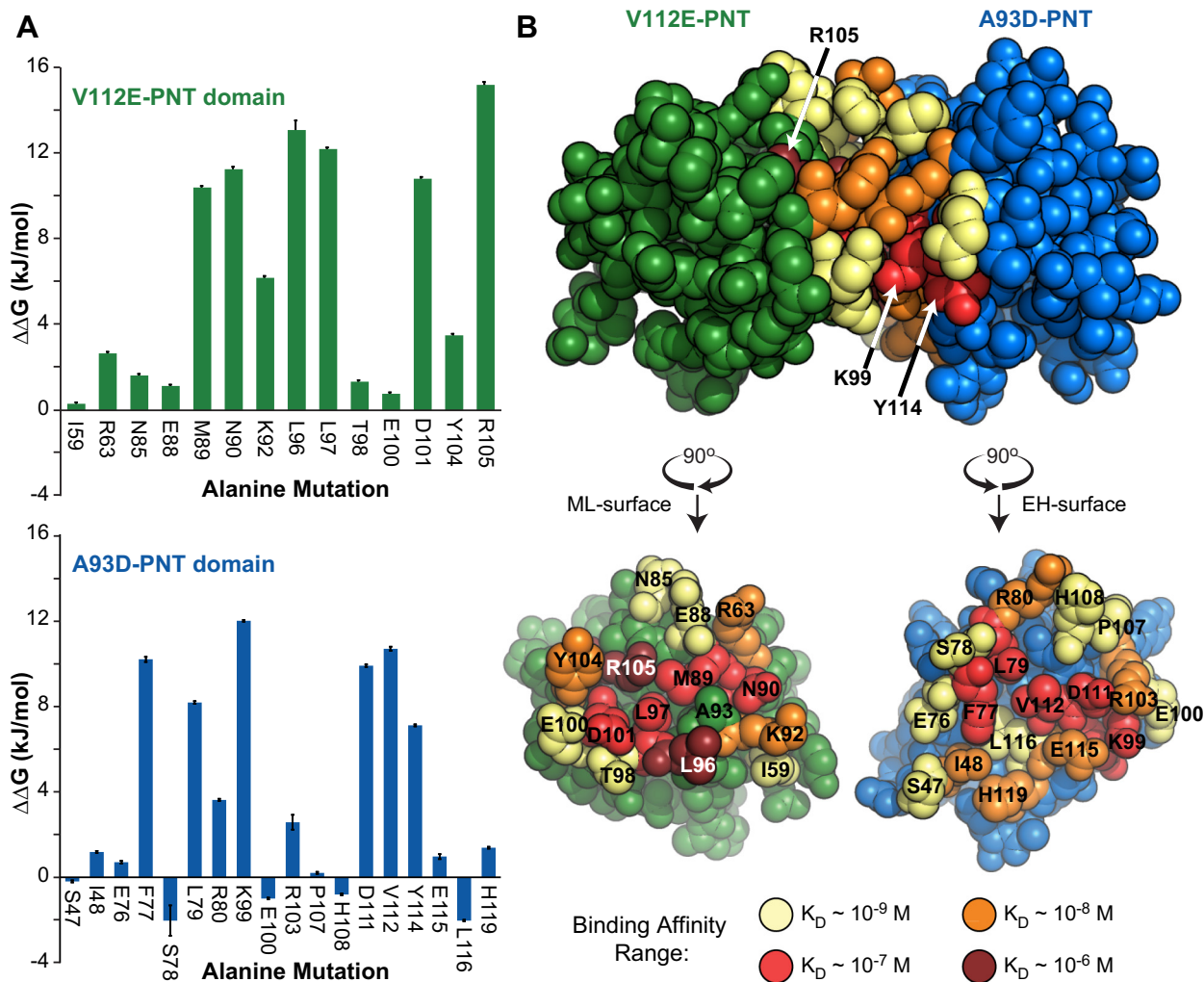
Figure 6. Characterization of the PNT domain interface by alanine scanning mutagenesis. *A*, A summary of the effects of alanine substitutions on the ML-surface of V112E-PNT domain (*top*) and the EH-surface of the A93D-PNT domain (*bottom*). Each ΔΔG value corresponds to the change in binding free energy to the complementary PNT domain relative to that of the reference protein with an unmodified wild-type interface (Tables 1 and 2). Standard errors are shown by the black lines. *B*, The effects of alanine substitutions on the K_D values for heterodimer dissociation are mapped onto the structures of the V112E- (*green* background) and A93D- (*blue* background) PNT domains. Residues for which alanine mutations maintained binding affinity approximately the wild-type K_D (~10^{-9} M) are indicated in *pale yellow*, whereas those weakening binding by approximately 10-fold (*orange*), 100-fold (*red*) and 1000-fold (*burgundy*) are in increasingly darker colors.

the largest ΔRMSD values clustered near the interface (Fig. 7D). That is, over the course of the MD simulations, residues near or at the wild-type ML-surface of the V112R-PNT domain exhibited smaller fluctuations than the corresponding residues of the mutated unbound ML-surface of the A93D-PNT domain. Similarly, residues near or at the wild-type EH-surface of the A93D-PNT domain, which includes its termini, exhibited smaller fluctuations than the corresponding residues of the mutated EH-surface of the V112R-PNT domain. Thus, although overall relatively rigid in MD simulations, to some degree heterodimerization reduced the mobility of residues at the bound interfacial regions of the heterodimeric PNT domains.

We also explored whether the dynamics simulations could be correlated with the experimental HX PFs. Despite time scales differing by many orders of magnitude, such a correlation between hydrogen bond persistence in MD

simulations with PFs, measured with NMR spectroscopy, was observed in a short helical peptide (32). Presumably, the persistence of hydrogen bonding reflects the underlying rigidity of the structure responsible for damping conformational motions required for exchange. Hence, the two backbone carbonyl oxygens closest to every amide nitrogen in the proteins were selected as candidates for backbone hydrogen bonding. The averages of these distances were calculated over the trajectories and amides nitrogens within 3.5 Å to carbonyl oxygens were assigned as being hydrogen bonded. As summarized in Figure S9, most of these amides, which are located in α-helices, exhibited measurable protection from HX (*i.e.*, with log(PF) > 3). However, several additional hydrogen-bonded amides in loop regions or the N-terminal segments of the PNT domain constructs exchanged more rapidly, with log(PF) values less than an upper measurable limit of ~3. Thus, in contrast to the very
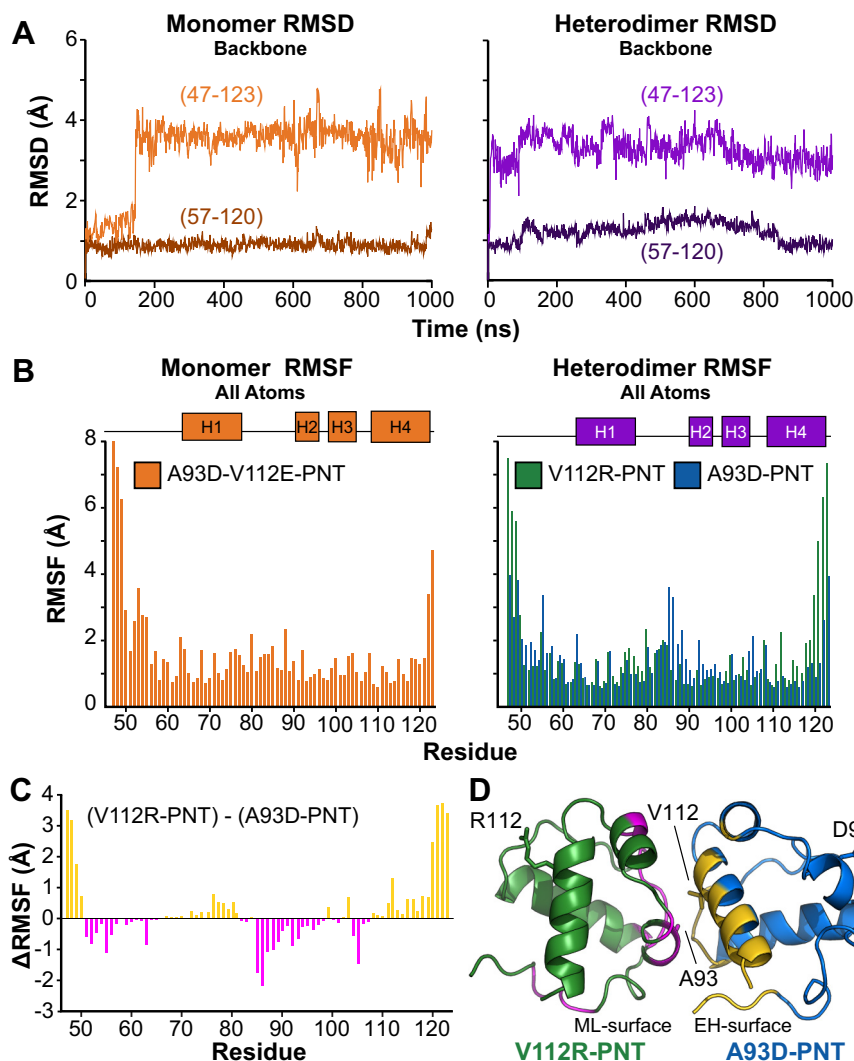
Figure 7. Analysis of PNT domain structural fluctuations from MD simulations. *A,* The RMSD values of the backbone atoms of residues 47 to 123 throughout the 1 μs MD simulation are plotted relative to the starting structures at 0 ns. After the initial equilibration, both the monomer (A93D-V112E-PNT domain) and heterodimer complex (A93D-PNT and V112R-PNT domains) showed little conformational drift. The event around 150 ns in the monomer simulation corresponds to a rearrangement of the N-terminal residues. The lower darker lines in both graphs correspond to the same analysis for the core residues 57 to 120 with the flexible N-terminal (47–56) and C-terminal (121–123) segments omitted. The low and consistent RMSD values ∼1 Å across the trajectories indicate the rigidity of the core PNT domain. *B,* The RMSF values of all atoms in each residue along the simulation trajectory are plotted for the monomer and heterodimer. The left plot shows the A93D-V112E-PNT monomer (*orange*) and the right plot shows both the V112R-PNT domain (*green*) and A93D-PNT domain (*blue*) in the heterodimer. In the heterodimer, the termini of the A93D-PNT domain, which are close to the interface, have lower mobility than those of the partner V112R-PNT domain or the monomeric A93D-V112E-PNT domain. *C,* The graph shows the ΔRMSF values obtained by subtracting the corresponding RMSF values of the A93D-PNT domain from those of the V112R-PNT domain in the heterodimer complex. Positive values (*gold*) and negative values (*magenta*) correspond to lower fluctuations for residues in the A93D-PNT domain and the V112R-PNT domain, respectively. *D,* The ΔRMSF values (>0.3 Å *gold*, < −0.3 Å *magenta*) are mapped onto the structure of the V112R- (*green*)/A93D- (*blue*) PNT domain heterodimer. Residues with reduced mobility cluster around the heterodimer interface. RMSF, root mean squared fluctuation; RMSD, root mean squared deviation; MD, molecular dynamics.

stable α-helices, these regions of the PNT domain undergo more facile local conformational fluctuations, including hydrogen bond breakage, and hence have reduced protection from HX.

Discussion

The ETV6 PNT domain retains a similar structure in monomeric and heterodimeric states

Our biophysical studies of the ETV6 PNT domain show that its tertiary structure is very stable in the monomeric and heterodimeric forms and is not perturbed by the presence of monomerizing mutations. There are no

substantial conformational differences between these species as seen both by X-ray crystallography and by NMR spectroscopy (secondary structural propensities and chemical shift perturbations). Furthermore, despite differences in contact surfaces within crystal lattices, the X-ray crystallographic structure of a monomeric A93D-V112E-PNT domain closely resembles that of the heterodimers. Similarly, the 1 μs MD simulations did not show any evidence for large conformational fluctuations and only small changes in the mobility of interfacial residues upon heterodimer formation. Thus, the well-characterized ETV6 PNT domain structure is retained both before and after polymerization. In

this respect, the PNT domain can be viewed as a rigid "lego block", poised to self-assemble without any requisite change in shape.

It is noteworthy that upon association of the A93D-PNT and V112E-PNT domains, amides exhibiting chemical shift perturbations and substantially increased HX protection localized to the wild-type ML- and EH-surfaces and not to the mutated surfaces. This confirms that only the former participate in heterodimer formation. Furthermore, along with a lack of any structural differences between the crystallographically characterized variants, this demonstrates that the "head" and "tail" interaction surfaces of a PNT domain are not allosterically coupled. Thus, the initial formation of the wild-type ETV6 PNT homodimer is unlikely to change the binding affinity of subsequent monomers to a growing polymer chain.

Alanine scanning mutagenesis identifies interfacial hotspot residues

Detailed SPR-monitored alanine scanning mutagenesis studies revealed that the ETV6 PNT domains heterodimerize with high affinity ($K_D \sim$ nM) because of residues partaking in electrostatic, van der Waals, and hydrophobic interactions. Although it has been reported that hotspots are generally not enriched in electrostatic interactions ([33]), we found the opposite with residues in two intermolecular salt bridges (K99-D101 and R105-D111) being critical for high-affinity binding. Previous studies have shown that a K99 R substitution also resulted in weaker binding, indicating that both the charge and amino acid structure are important, at least for one of these salt bridges ([19]). In contrast, the E100-R103 salt bridge does not contribute significantly as alanine substitutions of these residues did not result in a reduction in binding affinity or the oncogenic properties of the EN fusion protein ([19]). The presence of this salt bridge, and others involving these residues, varies between the reported ETV6 PNT domain structures. This also suggests that it is dynamic and not as persistent as the two formed by K99-D101 and R105-D111. Indeed, it is noteworthy that essentially all of the hotspot residues (*i.e.*, $\Delta\Delta G > 6$ kJ/mol), including these charge pairs, adopt similar well-ordered side chain conformations in all known X-ray crystallographic structures, regardless of intermolecular packing within the crystal lattice. In contrast, residues less critical for self-association adopt more variable side chain conformations.

In addition to A93 and V112 (the founding sites for monomerizing mutations), several hydrophobic residues with the EH- and ML-surfaces are also hotspots for heterodimerization. In particular, L96 on the ML-surface plays a critical role in binding as removal of its side chain resulted in ~1000-fold weaker affinity. The leucine makes contacts with several residues on the reciprocal EH-interface for which alanine substitutions also reduced the binding affinity, albeit each to a lesser extent. Targeting a molecule to bind near these EH-surface residues may exclude L96 and thereby inhibit PNT polymerization.

Many of the hotspot residues have increased protection from amide HX

Alanine scanning mutagenesis and amide HX experiments provide complementary insights into the self-association of the ETV6 PNT domain. That is, alanine scanning reveals the effect of removing the side chain of a given residue on the affinity for heterodimer formation, whereas HX experiments show how dimerization changes the conformational fluctuations of backbone amide hydrogens leading to exchange with water. In general, many of the hydrophobic hotspot residues within the core regions of the EH- and ML-surfaces, including L79, L96, L97, T98, Y104, V112, Y114, and L116, also exhibited enhanced HX protection upon heterodimerization. This is consistent with their burial, and likely dampened dynamics, within the heterodimer relative to the monomeric species. In contrast, although partaking in important intermolecular interactions, K99 and R105 underwent fast amide HX in both the monomeric and heterodimeric PNT domains, whereas their salt bridge partners D101 and D111 exhibited increased HX protection. This is also consistent with the location of the amides of these residues at the periphery of the dimer interface.

Hotspot residues are conserved among self-associating PNT domains

Although present in about one-third of ETS family members, the only other known self-associating PNT domains are those of *Drosophila* Yan ([6]) and possibly human ETV7 ([7–9]). The structure of the Yan PNT domain has been reported (PDB: 1SV4) and that of ETV7 can be reliably modeled based on its ~60% sequence identity with the ETV6 PNT domain. Both Yan and ETV7 have ML- and EH-surfaces with physicochemical features very similar to those of ETV6. Moreover, all ETV6 hotspot residues identified in this study are either identical or highly similar in the PNT domains of Yan and ETV7, whereas less critical interfacial residues show more sequence variability. These conserved residues likely play central roles in the self-association of all three ETS family PNT domains.

Small molecule inhibition of PNT domain polymerization

Recently, we developed and implemented a mammalian cell-based assay, utilizing a protein-fragment complementation approach with split *Gaussia* luciferase, and a yeast two-hybrid assay to screen chemical libraries for potential inhibitors of ETV6 PNT domain self-association ([17]). In parallel, virtual screening using the Bristol University Docking Engine ([34]) was performed to identify compounds predicted to bind the ETV6 polymerization interfaces. Although numerous candidate molecules were tested, none had inhibitory effects in cellular assays and none bound to the isolated PNT domain. This work highlighted the difficulty in disrupting with small molecules the polymerization of the ETV6 PNT domain which occurs through two relatively large flat interfaces.

Insights into the factors that drive ETV6 PNT domain polymerization may aid inhibitor design. For example, small

molecules that bind interaction surfaces near hotspot hydrophobic (*e.g.*, L96) or salt-bridging (*e.g.*, K99-D101 or R105-D111) residues may competitively prevent PNT domains from self-associating. However, designing or identifying such small molecules is challenging as the polymerization occurs with high affinity, and there does not appear to be any structural differences between the PNT domains in their monomeric and heterodimeric states that could be exploited for potential binding sites. One strategy that could be implemented for screening assays is to utilize an alanine mutant that weakens binding. This way, it may be easier to identify an initial compound that inhibits a micromolar affinity interaction, as opposed to the nanomolar interaction of the wild-type PNT domains. Once discovered, such a lead molecule could be optimized for higher affinity binding. As previously shown, mutations of K99 weaken PNT domain polymerization and alter the oncogenic properties of EN (19). This residue is not at the center of the heterodimer interface, and the K99 A or K99 R mutants may be well suited for such a screening strategy.

Alternatively, from the alanine scanning mutagenesis, we know which residues are not hotspots and thus tolerant to modifications. An example of a technique that would benefit from this acquired knowledge is disulphide tethering where weakly binding chemical fragments are tethered *via* an introduced cysteine residue near the protein-protein interaction interface (35). In principle, one could modify an ETV6 PNT domain residue, that does not affect binding and is near a hot spot, to a cysteine for this approach.

Helix "stapling" is another method that has been used to design molecules that disrupt protein–protein interactions. The general principle is to covalently stabilize the secondary structure of residues that normally form a helix along an interaction surface (36). The PNT domain, a subset of SAM domains, is a helical bundle, with helix H4 and helices H2 and H3 forming complementary interaction surfaces. Residues D111, V112, and Y114 in helix H4 are all hotspots and experience substantial increases in HX protection upon dimerization. Thus, a stapled helical polypeptide or a polypeptide mimetic encompassing these residues might be sufficient to prevent polymerization. In contrast, whereas several residues in H2 and H3 are hotspots, such as D101 and R105, they are not adjoining as a single helix. A similar methodology has been used to target the Ship2 and EphA2 SAM-SAM domain interactions, whereby a penta-amino acid motif found in EphA2 binds to the SAM domain of Ship2, albeit with a K_D value in the high micromolar range (37).

As a closing comment, SPR experiments demonstrated that the lifetime of the PNT domain heterodimer ($1/k_{off}$) is ∼10 min. The lifetime of polymeric forms will be longer since dissociation must occur at multiple interfaces to completely monomerize. If the dissociation of the PNT domain polymer is slow in the context of a PNT–PTK fusion oncoprotein, then a prospective small molecule inhibitor that would have the greatest effect may likely need to act on newly translated PNT–PTK fusion oncoproteins, before they polymerize.

Experimental procedures

PNT domain expression and purification

The DNA construct encoding residues 1 to 125 of human ETV6 (Genbank Gene ID: 2120), preceded by a thrombin-cleavable N-terminal His$_6$-affinity tag (Table S1), was initially cloned into the pET28a vector (Invitrogen) as described (19). The monomerizing A93D, V112 E, and A93D-V112 E substitutions were introduced through QuikChange site-directed mutagenesis (Stratagene). Subsequently, it was recognized that a thrombin cleavage site is present within the intrinsically disordered N-terminal region of these constructs (residues V37-P38-R39↓A40). This is located just before an alternative start site (M43) for ETV6 expression (14). Thus, ETV6 fragments were expressed from available clones as residues 1 to 125 with a His$_6$-tag and cleaved with thrombin to yield final purified samples spanning residues 40 to 125.

PNT domain-containing proteins were expressed in *Escherichia coli* BL21 (λDE3) cells grown at 37 °C to an OD$_{600}$ ∼ 0.6 and induced with 1 mM IPTG overnight. Unlabeled proteins were produced in lysogeny broth (LB) media, and M9 minimal media was supplemented with either 1 g/L of ^{15}NH$_4$Cl for ^{15}N-labeled proteins or 1 g/L of ^{15}NH$_4$Cl and 3 g/L ^{13}C$_6$-glucose for ^{13}C/^{15}N-labeled proteins. In all cases, 35 mg/L kanamycin was included for plasmid selection. After continued growth at 37 °C overnight, the cells were harvested by centrifugation (Sorvall GSA rotor; 5000 rpm) and frozen at −80 °C. The cell pellet was thawed for purification and resuspended in denaturing buffer (4 M GdnHCl, 20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.5) and lysed by sonication on ice. The lysate was then cleared by centrifugation (Sorvall SS34 rotor; 15,000 rpm), and the resulting supernatant was passed through either a 0.45 or 0.8 μm pore size filter and loaded onto a 5 ml Ni^{+2}-NTA HisTrap HP column (GE Healthcare) pre-equilibrated with binding buffer (20 mM imidazole, 50 mM sodium phosphate, 500 mM NaCl, pH 7.5). After washing with several column volumes of binding buffer, the protein was eluted using a 120 ml linear gradient of elution buffer (500 mM imidazole, 50 mM sodium phosphate, 400 mM NaCl, pH 7.5).

The collected fractions were analyzed by SDS-PAGE, and those containing the desired protein were pooled. In general, the ETV6 fragments eluted as two major peaks off the Ni^{2+} column. Previous studies demonstrated that proteins from these fractions had the same masses yet showed small differences in their ^{15}N-HSQC spectra (38). Despite significant efforts, the origin of these differences was never elucidated. For consistency, only the fastest eluting peak was collected and dialyzed overnight at 4 °C in thrombin cleavage buffer (20 mM Tris, 0.15 mM NaCl, 2.5 mM CaCl$_2$, 0.5 mM EDTA, 1 mM DTT, pH 8.4) with ∼1 unit of thrombin (Millipore) per 20 ml of collected fractions. The cleaved protein was separated from any His$_6$-tagged species by passage through the Ni^{+2}-NTA HisTrap HP column and purified further using size-exclusion chromatography (Superdex S75, GE Healthcare). This also served to exchange the protein in a buffer optimized for NMR experiments (noted below). The concentration of each protein

sample was determined by ultraviolet absorbance at 280 nm based on its predicted molar absorptivity (39).

General NMR spectroscopy methods

NMR spectra were recorded with cryoprobe-equipped Bruker Avance III 500, 600, and 850 MHz spectrometers. All data acquired were processed and analyzed with NMRPipe (40) and NMRFAM-Sparky (41, 42). Typically, protein samples were 150 μM to 600 μM in ~450 μl of standard buffer (20 mM MOPS, 50 mM NaCl, and 0.5 mM EDTA) with D_2O (5% v/v) added for signal locking. Unless otherwise noted, experiments involving the A93D-PNT domain were conducted at pH 7.0, and those involving the V112E-PNT domain were conducted at pH 8.0 because of its propensity to self-associate under lower sample pH conditions. In the case of the heterodimer species, the unlabeled partner protein was added to its isotopically labeled partner in a 1.1 M excess. The heterodimer containing an isotopically labeled V112E-PNT domain was studied at pH 7.5, whereas that containing an isotopically labeled A93D-PNT domain was studied at pH 7.0.

Spectral assignments and chemical shift analyses

The signals from mainchain ^1H, ^{13}C, and ^{15}N nuclei of the $^{13}C/^{15}N$-labeled PNT domain constructs were assigned using standard heteronuclear ^1H-^{13}C-^{15}N scalar correlation experiments (43), including HNCACB, HNCO, CBCA(CO)NH, HNCACO, along with HSQC-NOESY spectra, recorded on a 600 MHz spectrometer at 25 °C. Spectra of the monomeric A93D-PNT domain were assigned manually, whereas those of the monomeric V112E-PNT domain and the two heterodimer complexes were automatically interpreted using PINE (44) and verified manually. The assigned chemical shifts of these four species have been deposited in the BioMagResBank.

Secondary structure analyses were carried out using the motif identification from chemical shift online server (23). Combined CSPs (Δδ) for the $^1H^N$ (Δδ_H) and ^{15}N (Δδ_N) signals of the corresponding amides in the dimer versus monomer species were calculated as $\Delta\delta = [(\Delta\delta_H)^2 + (0.14\Delta\delta_N)^2)]^{1/2}$ (20).

Amide hydrogen exchange by NMR spectroscopy

Protium–deuterium HX experiments for the PNT domains in the absence (monomeric) and presence of their unlabeled partner (heterodimeric) were conducted on a Bruker Avance 600 MHz spectrometer at 21 °C (matching ambient room temperature). The initial pH values were 7.0 for the samples containing the ^{15}N-labeled A93D-PNT domains and 7.5 for those containing the ^{15}N-labeled V112E-PNT domain. Initial reference ^{15}N-HSQC spectra of the proteins in H_2O buffer were recorded. Subsequently a 450 μl aliquot was lyophilized, resuspended with the same volume of D_2O, and immediately put into the NMR spectrometer to begin data recording within 4 to 7 min. Initially, a series of ~ 5-min ^{15}N-HSQC spectra were acquired with two scans/free induction decay to characterize amides exchanging on the minutes timescale. Then ~20-min ^{15}N-HSQC spectra with eight scans/free induction decay were collected back-to-back for ~3 h, followed by a 20-

min spectrum every hour for ~24 h, and then intermittent 20-min spectra over a period up to 3 months. After the first week, the sample was removed from the spectrometer and stored at ambient room temperature between recording spectra. Upon completion of data recording, the pH* (pH meter reading uncorrected for the deuterium isotope effect) of each sample was measured as 7.3 (monomeric A93D-PNT domain), 7.6 (monomeric V112E-PNT domain), 7.4 (^{15}N-labeled A93D-PNT domain in complex with V112E-PNT domain), and 7.5 (^{15}N-labeled V112E-PNT domain in complex with A93D-PNT domain).

For each amide with measurable signals at times t after resuspension in D_2O, the pseudo-first order exchange rate constant k_{obs} was obtained by fitting with NMRFAM-Sparky the $^1H^N$-^{15}N peak intensity I_t, scaled for number of acquisitions/free induction decay, to the equation for a single exponential decay:

$$I_t = I_0 e^{-(k_{obs})(t)}$$

I_0 is the fit initial intensity extrapolated to t = 0. The protection factor ($PF = \frac{k_{pred}}{k_{obs}}$) for each amide was determined as the ratio of its predicted intrinsic exchange rate constant (k_{pred}) in an unstructured polypeptide of the same amino acid sequence versus its experimentally measured k_{obs}. The k_{pred} values were determined with the program Sphere (45) which uses reference data based on poly-DL-alanine and corrected for amino acid type, temperature, pH and isotope effects (46, 47). In the cases of amide that had not exchanged significantly after 3 months, lower limits on their PFs were estimated based on the largest measured PFs for the given sample.

Site-directed mutagenesis and construct cloning

To enable site-directed biotinylation during protein expression in *E. coli*, a gene encoding residues 43 to 135 of V112E ETV6 with an N-terminal His_6-tag and Avitag (Table S1) was constructed in the pET28a vector using polymerase incomplete primer extension cloning techniques (48, 49). The A93D and E112V mutations were sequentially introduced to generate the complementary A93D-PNT domain construct with the His_6-tag and Avitag.

Interfacial residues present in the X-ray crystal structure of the PNT domain dimer (PDB: 1LKY) were identified using the online Solvent accessibility-based Protein-Protein Interface iDEntification and Recognition server [SPPIDER (50)]. Alanine substitutions were encoded at these sites in the respective A93D- or V112E-PNT domain clones using QuikChange site-directed mutagenesis techniques. All but two constructs were successfully generated in-house, and genes encoding N90A-V112E-PNT domain and L116A-A93D-PNT domain were purchased commercially (Biomatik).

Protein expression and purification for surface plasmon resonance

Each plasmid encoding either the biotinylated A93D-PNT and V112E-PNT domain was co-transformed into *E. coli*

BL21 (λDE3) with the pET21a-BirA plasmid, which produces biotin ligase. Selection of both plasmids was maintained by including kanamycin (35 mg/L) and ampicillin (100 mg/L) in all media. Overnight seed cultures were used to inoculate LB media, supplemented with 0.05 mM biotin, and grown at 37 °C to an $OD_{600} \sim 0.6$. Protein expression was induced with 0.5 mM IPTG, and the cells were grown at 30 °C overnight. The cells were collected by centrifugation and cell pellets were frozen at –80 °C.

Protein purification was carried out fully as described above with minor modifications. The final size exclusion purification step was omitted after thrombin cleavage and removal of the His$_6$-tag by passage through a Ni^{+2}-NTA HisTrap HP column. The final protein samples were exchanged into 20 mM MOPS, 50 mM NaCl, 0.5 mM EDTA buffer, at pH 8.0, and concentrated to ~1 ml with an Amicon 3K MWCO centrifugal filter. The protein samples were then snap frozen in liquid nitrogen and stored at –80 °C before SPR analysis. The proteins were ~50 to 95% biotinylated as judged by MALDI-ToF mass spectrometry.

Surface plasmon resonance

SPR experiments were performed at 25 °C on a Biacore X100 instrument using the streptavidin Sensor Chip SA to capture the biotinylated PNT domain "ligand". The ligand was diluted to 50 μg/ml in HBS-EP+ buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% v/v Tween-20, pH 7.4) and immobilized on the chip using the Biacore X100 control software immobilization wizard. The regeneration scouting software wizard was used to determine the regeneration step involving flowing 0.2% SDS over the chip for 30 s.

The association (k_{on}) and dissociation (k_{off}) rate constants and the equilibrium dissociation constant (K_D) for binding of the analyte with the immobilized ligand were determined using the Biacore X100 kinetics/affinity assay software wizard. The positive and negative binding controls were analytes with the complementary wild-type PNT domain interface or with the same monomerizing substitution, respectively. For the experimental runs, the analyte protein sample was initially diluted in HBS-EP+ buffer to 0.2, 2, 20, 40, and 60 nM. If weakened binding ($K_D > 60$ nM) was observed, the analyte protein sample was re-run in HBS-EP+ buffer at 20, 200, 2000, 4000, and 6000 nM.

Structural determination of A93D-V112E-PNT domain by X-ray crystallography

A construct of ETV6 spanning residues 40 to 125 with the A93D and V112 E substitutions (A93D-V112E-PNT domain) was purified as described above. Crystals were grown by sitting drop vapor diffusion at room temperature with 2 μl drops prepared with a 1:1 mixture of ~9.3 mg/ml protein (20 mM MOPS, 50 mM NaCl, and 0.5 mM EDTA, pH 7.0) and reservoir solutions from the Hampton Index reagent crystallization screen (Hampton Research). Several conditions yielded protein crystals and those grown in 2.8 M sodium acetate (pH 7.0) were used for data collection. These crystals were cryoprotected by brief soaking in reservoir buffer

supplemented with 30% (v/v) glycerol, followed by flash freezing in liquid nitrogen.

Diffraction data were collected at the CLS (Canadian Light Source) on beamline 08B1-1 (51). The data were cut-off at 1.85 Å based on the CC1/2 metric and processed and scaled using XDS (52). Crystals were of space group $P6_522$ with two protein molecules in the asymmetric unit. Phase determination using molecular replacement was performed with a PNT domain monomer from PDB: 1LKY using the AutoSol program in Phenix (53). Model building was performed in Coot, and refinement was executed using the Phenix software suite (54). Eight of 85 amino acid residues of the construct (4 at the N-terminus and 2 at the C-terminus) were disordered in the crystal and not modeled.

Molecular dynamics simulations on monomeric and heterodimeric PNT domains

MD simulations were based on coordinate files, encompassing residues S47 to Q123, from the crystal structure of the monomeric A93D-V112E-PNT domain (determined herein) and a dimer of the A93D-PNT domain and V112R-PNT domain from PDB: 1LKY.

The MD simulations were performed on the University of Bristol High Performance Computer BlueCrystal using GROMACS (Version 5.1.2) (55). The systems were solvated with TIP3P waters in an orthorhombic box 2 nm larger than the longest dimension of the protein. Sodium and chloride ions were included at 50 mM to emulate experimental conditions, while neutralizing the monomerizing mutations to have no overall net charge. The amber99sb-ildn forcefield was used to parameterize the protein simulations (56). Non-bonded long-range electrostatic interactions were calculated using the Particle Mesh Ewald method with a 1.2 nm cut-off. Bonds were constrained using the LINCS algorithm allowing the use of a 2 fs timestep for the MD integration. The energy of the system was minimized over 5000 steps of the steepest descent energy minimization. The system then underwent a position-restraint simulation over 200 ps where the protein was restrained to its initial position while heating the system to 310 K and introducing pressure at 1 bar using the Berendsen barostat (55). The full unconstrained MD simulations were run over 1 μs with integration step sizes of 2 fs using the leap-frog algorithm, and trajectory files were recorded every 100 ps. The temperature was maintained at 310 K using v-rescale modified Berendsen thermostat and at 1 bar with the Parinello-Rahman barostat. Trajectories were analyzed and processed, including principal component analysis, utilizing GROMACS tools. The simulations were visualized with VMD (57), gnuplot, and PyMol (58).

Data availability

The X-ray structure (coordinates and structure factor files) has been submitted to the PDB under accession number 7JU2. The NMR chemical shifts have been submitted to the BMRB under accession numbers 50430 (Monomeric V112E-PNT domain), 50431 (Complexed V112E-PNT domain), 50432

(Complexed A93D-PNT domain), and 50433 (Monomeric A93D-PNT domain). All other data described here are available within the manuscript and supporting information.

Conflict of interest—The authors declared no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Abbreviations—The abbreviations used are: CSP, chemical shift perturbation; EN, ETV6-NTRK3 fusion oncoprotein; ETS, E26 transformation specific; HX, hydrogen exchange; MD, molecular dynamics; PF, protection factor; PC, principal component; PTK, protein tyrosine kinase; RCI-S^2, random coil index-squared order parameter; RMSD, root mean square deviation; RMSF, root mean square fluctuation; SPR, surface plasmon resonance.

References

1. Lopez, R. G., Carron, C., Oury, C., Gardellin, P., Bernard, O., and Ghysdael, J. (1999) TEL is a sequence-specific transcriptional repressor. *J. Biol. Chem.* **274**, 30132–30138
2. Green, S. M., Coyne, H. J., McIntosh, L. P., and Graves, B. J. (2010) DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J. Biol. Chem.* **285**, 18496–18504
3. Hollenhorst, P. C., McIntosh, L. P., and Graves, B. J. (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu. Rev. Biochem.* **80**, 437–471
4. Kim, C. A., and Bowie, J. U. (2003) SAM domains: Uniform structure, diversity of function. *Trends Biochem. Sci.* **28**, 625–628
5. Tran, H. H., Kim, C. A., Faham, S., Siddall, M., and Bowie, J. U. (2002) Native interface of the SAM domain polymer of TEL. *BMC Struct. Biol.* **2**, 5
6. Qiao, F., Song, H., Kim, C. A., Sawaya, M. R., Hunter, J. B., Gingery, M., Rebay, I., Courey, A. J., and Bowie, J. U. (2004) Derepression by depoly-merization. *Cell* **118**, 163–173
7. Potter, M. D., Buijs, A., Kreider, B., van Rompaey, L., and Grosveld, G. C. (2000) Identification and characterization of a new human ETS-family transcription factor, TEL2, that is expressed in hematopoietic tissues and can associate with TEL1/ETV6. *Blood* **95**, 3341–3348
8. Vivekanand, P., and Rebay, I. (2012) The SAM domain of human TEL2 can abrogate transcriptional output from TEL1 (ETV-6) and ETS1/ETS2. *PLoS One* **7**, 5–10
9. Knight, M. J., Leettola, C., Gingery, M., Li, H., and Bowie, J. U. (2011) A human sterile alpha motif domain polymerizome. *Protein Sci.* **20**, 1697–1706
10. Mackereth, C. D., Schärpf, M., Gentile, L. N., MacIntosh, S. E., Slupsky, C. M., and McIntosh, L. P. (2004) Diversity in structure and function of the Ets family PNT domains. *J. Mol. Biol.* **342**, 1249–1264
11. Hock, H., Meade, E., Medeiros, S., Schindler, J. W., Valk, P. J. M., Fujiwara, Y., and Orkin, S. H. (2004) Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* **18**, 2336–2341
12. Wang, L. C., Kuo, F., Fujiwara, Y., Gilliland, D. G., Golub, T. R., and Orkin, S. H. (1997) Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J.* **16**, 4374–4383
13. Kim, C. A., Phillips, M. L., Kim, W., Gingery, M., Tran, H. H., Robinson, M. A., Faham, S., and Bowie, J. U. (2001) Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J.* **20**, 4173–4182
14. Bohlander, S. K. (2005) ETV6: A versatile player in leukemogenesis. *Semin. Cancer Biol.* **15**, 162–174
15. De Braekeleer, E., Douet-Guilbert, N., Morel, F., Le Bris, M.-J., Basinko, A., and De Braekeleer, M. (2012) ETV6 fusion genes in hematological malignancies: A review. *Leuk. Res.* **36**, 945–961
16. Wai, D., Knezevich, S., Lucas, T., Jansen, B., Kay, R., and Sorensen, P. H. B. (2000) The ETV6-NTRK3 gene fusion encodes a chimeric protein tyrosine kinase that transforms NIH3T3 cells. *Oncogene* **19**, 906–915
17. Gerak, C. A. N., Zhang, S. M., Balgi, A. D., Sadowski, I. J., Sessions, R. B., McIntosh, L. P., and Roberge, M. (2020) A multipronged screening approach targeting inhibition of ETV6 PNT domain polymerization. *SLAS Discov. Adv. Sci. Drug Discov.* https://doi.org/10.1177/2472555220979599
18. Tognon, C. E., Mackereth, C. D., Somasiri, A. M., McIntosh, L. P., and Sorensen, P. H. B. (2004) Mutations in the SAM domain of the ETV6-NTRK3 chimeric tyrosine kinase block polymerization and trans-formation activity. *Mol. Cell. Biol.* **24**, 4636–4650
19. Cetinbas, N., Huang-Hobbs, H., Tognon, C., Leprivier, G., An, J., McKinney, S., Bowden, M., Chow, C., Gleave, M., McIntosh, L. P., and Sorensen, P. H. (2013) Mutation of the salt bridge-forming residues in the ETV6-SAM domain interface blocks ETV6-NTRK3-induced cellular transformation. *J. Biol. Chem.* **288**, 27940–27950
20. Williamson, M. P. (2013) Using chemical shift perturbation to charac-terise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16
21. Nauli, S., Farr, S., Lee, Y.-J., Kim, H.-Y., Faham, S., and Bowie, J. U. (2007) Polymer-driven crystallization. *Protein Sci.* **16**, 2542–2551
22. Poulos, S., Agah, S., Jallah, N., and Faham, S. (2017) Symmetry based assembly of a 2 dimensional protein lattice. *PLoS One* **12**, e0174485
23. Shen, Y., and Bax, A. (2012) Identification of helix capping and β-turn motifs from NMR chemical shifts. *J. Biomol. NMR.* **52**, 211–232
24. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., and Thornton, J. M. (2018) PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134
25. Berjanskii, M. V., and Wishart, D. S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971
26. Holm, L., and Sander, C. (1995) DALI: A network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480
27. Paterson, Y., Englander, S. W., and Roder, H. (1990) An antibody binding site on cytochrome c defined by hydrogen exchange and two-dimensional NMR. *Science* **249**, 755–759
28. Dyson, H. J., Kostic, M., Liu, J., and Martinez-Yamout, M. A. (2008) Hydrogen-deuterium exchange strategy for delineation of contact sites in protein complexes. *FEBS Lett.* **582**, 1495–1500
29. Mandell, J. G., Falick, A. M., and Komives, E. A. (1998) Identification of protein – protein interfaces by decreased amide solvent accessibility. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14705–14710
30. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E., and Englander, S. W. (2012) Protein hydrogen exchange: Testing current models. *Protein Sci.* **21**, 987–995

31. Krishna, M. M. G., Hoang, L., Lin, Y., and Englander, S. W. (2004) Hydrogen exchange methods to study protein folding. *Methods* **34**, 51–64

32. Sessions, R. B., Gibbs, N., and Dempsey, C. E. (1998) Hydrogen bonding in helical polypeptides from molecular dynamics simulations and amide hydrogen exchange analysis: Alamethicin and melittin in methanol. *Biophys. J.* **74**, 138–152

33. Keskin, O., Ma, B., and Nussinov, R. (2005) Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **345**, 1281–1294

34. McIntosh-Smith, S., Price, J., Sessions, R. B., and Ibarra, A. A. (2015) High performance in silico virtual drug screening on many-core processors. *Int. J. High Perform. Comput. Appl.* **29**, 119–134

35. Arkin, M. R., Randal, M., DeLano, W. L., Hyde, J., Luong, T. N., Oslob, J. D., Raphael, D. R., Taylor, L., Wang, J., McDowell, R. S., Wells, J. A., and Braisted, A. C. (2003) Binding of small molecules to an adaptive protein-protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1603–1608

36. Wilson, A. J. (2009) Inhibition of protein-protein interactions using designed molecules. *Chem. Soc. Rev.* **38**, 3289–3300

37. Mercurio, F. A., Di Natale, C., Pirone, L., Iannitti, R., Marasco, D., Pedone, E. M., Palumbo, R., and Leone, M. (2017) The Sam-Sam interaction between Ship2 and the EphA2 receptor: Design and analysis of peptide inhibitors. *Sci. Rep.* **7**, 1–11

38. Huang-Hobbs, H. (2013) *Dissecting the Mechanism of ETV6 Polymerization.* M.Sc. thesis, University of British Columbia

39. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788

40. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* **6**, 277–293

41. Goddard, T. D., and Kneller, D. G. (1999) Sparky 3, University of California, San Fransisco, CA

42. Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327

43. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158

44. Bahrami, A., Assadi, A. H., Markley, J. L., and Eghbalnia, H. R. (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *Plos Comput. Biol.* **5**. e1000307, 1–15

45. Zhang, Y. Z. (1995) *Protein and Peptide Structure and Interactions Studied by Hydrogen Exchange and NMR.* Ph.D. thesis, University of Pennsylvania

46. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86

47. Connelly, G. P., Bai, Y., Jeng, M.-F., and Englander, S. W. (1993) Isotope effects in peptide group hydrogen exchange. *Proteins Struct. Funct. Bioinforma* **17**, 87–92

48. Klock, H. E., and Lesley, S. A. (2009) The polymerase incomplete primer extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. In: Doyle, S. A., ed. *High Throughput Protein Expression and Purification: Methods and Protocols*, Humana Press, Totowa, NJ: 91–103

49. Ashraf, S. S., Benson, R. E., Payne, E. S., Halbleib, C. M., and Grøn, H. (2004) A novel multi-affinity tag system to produce high levels of soluble and biotinylated proteins in *Escherichia coli*. *Protein Expr. Purif.* **33**, 238–245

50. Porollo, A., and Meller, J. (2007) Prediction-based fingerprints of protein-protein interactions. *Proteins Struct. Funct. Bioinforma* **66**, 630–645

51. Fodje, M., Grochulski, P., Janzen, K., Labiuk, S., Gorin, J., and Berg, R. (2014) 08B1-1: An automated beamline for macromolecular crystallography experiments at the Canadian Light Source. *J. Synchrotron Radiat.* **21**, 633–637

52. Kabsch, W. (2010) Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132

53. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221

54. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132

55. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56

56. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958

57. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38

58. *The PyMOL Molecular Graphics System*, Schrödinger, LLC.