Biologically active LIL proteins built with minimal chemical diversity

Erin N. Heim[a], Jez L. Marston[a,1], Ross S. Federman[b,1], Anne P. B. Edwards[a], Alexander G. Karabadzhak[c], Lisa M. Petti[a], Donald M. Engelman[c,d,2], and Daniel DiMaio[a,c,d,e,2]

[a]Department of Genetics, Yale School of Medicine, New Haven, CT 06520-8005; [b]Department of Immunobiology, Yale School of Medicine, New Haven, CT 06520-8011; [c]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8024; [d]Yale Cancer Center, New Haven, CT 06520-8028; and [e]Department of Therapeutic Radiology, Yale School of Medicine, New Haven, CT 06520-8040

Contributed by Donald M. Engelman, July 21, 2015 (sent for review April 15, 2015; reviewed by Alessandro Senes and Bill Sugden)

We have constructed 26-amino acid transmembrane proteins that specifically transform cells but consist of only two different amino acids. Most proteins are long polymers of amino acids with 20 or more chemically distinct side-chains. The artificial transmembrane proteins reported here are the simplest known proteins with specific biological activity, consisting solely of an initiating methionine followed by specific sequences of leucines and isoleucines, two hydrophobic amino acids that differ only by the position of a methyl group. We designate these proteins containing leucine (L) and isoleucine (I) as LIL proteins. These proteins functionally interact with the transmembrane domain of the platelet-derived growth factor β-receptor and specifically activate the receptor to transform cells. Complete mutagenesis of these proteins identified individual amino acids required for activity, and a protein consisting solely of leucines, except for a single isoleucine at a particular position, transformed cells. These surprisingly simple proteins define the minimal chemical diversity sufficient to construct proteins with specific biological activity and change our view of what can constitute an active protein in a cellular context.

synthetic biology | oncogene | traptamer | E5 protein | PDGF receptor

The chemical diversity of the 20 standard amino acid side-chains found in proteins supports myriad biochemical activities essential for life, and additional chemical diversity is generated by posttranslational amino acid modifications. The number of potential protein sequences is enormous: for proteins only 300 amino acids long, ~10^{400} possible sequences exist, a number larger by many orders-of-magnitude than the number of atoms in the known universe. This immense number of sequences and the chemical and conformational complexity of naturally occurring proteins hinder our ability to understand protein structure, folding, and function, and complicate protein engineering efforts. In addition, many different related amino acid sequences can fold into nearly identical structures, and even proteins with quite divergent sequences or protein folds can use similar chemistry to carry out related functions or display the same catalytic activity (e.g., refs. 1–3). These considerations further complicate attempts to clearly identify and understand the key structural features of proteins. Thus, the isolation of biologically active proteins with drastically reduced chemical complexity would represent a significant advance in protein science by facilitating the correlation of specific structural features with function.

Up to 30% of all proteins contain ~20–25-amino acid, primarily hydrophobic segments that span cell membranes (4). These transmembrane domains often play critical roles in cellular processes by engaging in highly specific protein–protein and protein–lipid interactions required for the proper folding, oligomerization, and function of transmembrane proteins (5–8). For example, receptor tyrosine kinases (RTKs) such as the PDGF-β receptor (PDGFβR) usually consist of an extracellular ligand binding domain, a single transmembrane domain, and a cytoplasmic tyrosine kinase domain. RTKs typically exist as catalytically inactive monomers, and ligand binding induces RTK dimerization, kinase activation, autophosphorylation of the receptor on cytoplasmic tyrosine residues, and mitogenic signaling. Mutations in the transmembrane domains of some RTKs, such as ErbB2, can cause constitutive activation of the receptor (9).

Some naturally occurring proteins consist of only one or a few helical transmembrane domains with little additional folded structure, and computational methods have been used to design artificial transmembrane peptides with biochemical or biological activity (10–13). The dimeric 44-amino acid bovine papillomavirus (BPV) E5 protein is the smallest known naturally occurring oncoprotein (14–16). This very hydrophobic protein, essentially a free-standing transmembrane domain with short extramembraneous segments, binds specifically to the transmembrane domain of the PDGFβR, resulting in constitutive activation of this receptor (14, 16–21). Sustained mitogenic signaling by the activated PDGFβR confers the transformed phenotype on cells, including loss of contact inhibition, morphologic transformation, focus formation, growth factor independence, and tumorigenicity in animals. Coimmunoprecipitation studies in detergent extracts of transformed cells demonstrated that the E5 protein and PDGFβR exist in a stable complex, and extensive mutational analysis, isolation of compensatory mutants between the E5 protein and the PDGFβR, and molecular modeling suggest that the E5 protein interacts directly with the transmembrane domain of the PDGFβR. This transmembrane interaction appears to be mediated by interactions between specific hydrophilic residues in the two proteins as well as by packing interactions.

Significance

Most proteins are long polymers of amino acids with 20 or more chemically distinct side-chains, whereas transmembrane domains are short membrane-spanning protein segments with mainly hydrophobic amino acids. Here, we have defined the minimal chemical diversity sufficient for a protein to display specific biological activity by isolating artificial 26-aa-long transmembrane proteins consisting of random sequences of only two hydrophobic amino acids, leucine and isoleucine. A small fraction of proteins with this composition interact with the transmembrane domain of a growth factor receptor to specifically activate the receptor, resulting in growth transformation. These findings change our view of what can constitute an active protein and have important implications for protein evolution, protein engineering, and synthetic biology.

Author contributions: E.N.H., J.L.M., R.S.F., A.P.B.E., A.G.K., L.M.P., D.M.E., and D.D. designed research; E.N.H., J.L.M., R.S.F., A.P.B.E., A.G.K., and L.M.P. performed research; E.N.H., J.L.M., R.S.F., A.P.B.E., A.G.K., L.M.P., D.M.E., and D.D. analyzed data; and E.N.H., A.G.K., D.M.E., and D.D. wrote the paper.

Reviewers: A.S., University of Wisconsin–Madison; and B.S., University of Wisconsin.

The authors declare no conflict of interest.

[1]J.L.M. and R.S.F. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: daniel.dimaio@yale.edu or donald.engelman@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1514230112/-/DCSupplemental.

We have developed genetic methods to select short biologically active transmembrane proteins modeled on the E5 protein from libraries expressing many thousands of artificial proteins with randomized hydrophobic segments (named transmembrane protein aptamers or "traptamers") (22–25). Because of the relative simplicity of transmembrane domains, we reasoned that this approach could be used to define the minimal chemical complexity sufficient to construct biologically active proteins. Here, we isolate short transmembrane oncoproteins that are the simplest proteins ever described with specific biological activity. These proteins change our understanding of the chemical diversity required for protein function.

Results

To isolate ultrasimple oncoproteins, we constructed a retroviral library (designated UDv6) expressing traptamers in which a methionine initiation codon was followed by 25 codons encoding equimolar leucine and isoleucine in random order (Fig. 1 *A* and

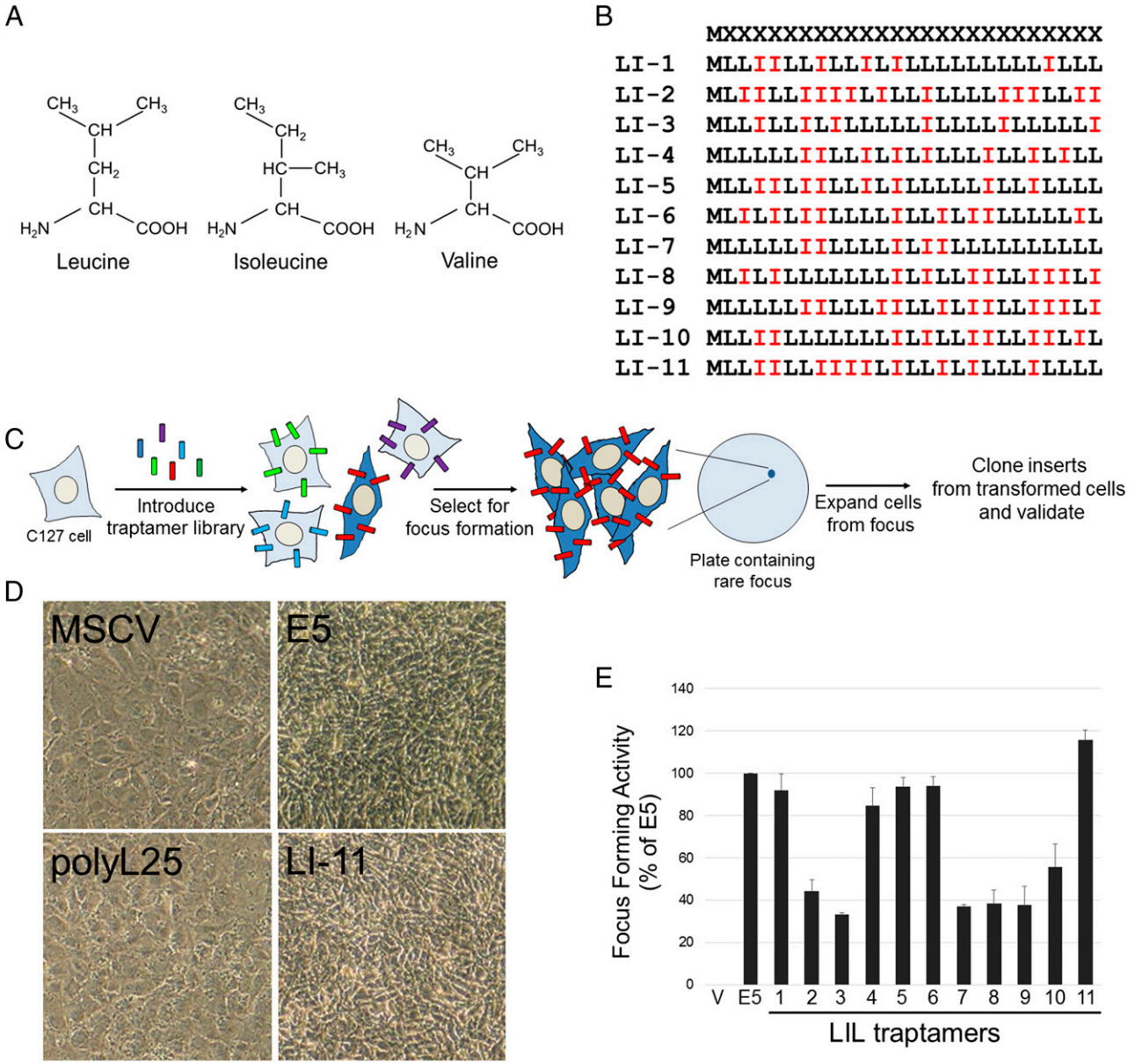

Fig. 1. Proteins composed of only leucine and isoleucine transform cells. (*A*) Isoleucine and leucine side-chains differ in the position of a methyl group. Valine is identical to isoleucine except for the absence of a terminal methyl group. (*B*) Sequences of the UDv6 library and the active traptamers. The first line shows the design of the library, where M represents methionine and X represents positions randomized to be equimolar leucine (L) and isoleucine (I). The other lines show the sequences of the 11 active traptamers. (*C*) Scheme to isolate traptamers with transforming activity. C127 mouse fibroblasts were infected with a retrovirus library expressing traptamers (the different colored cylinders represent different traptamer sequences) and incubated at confluence for 3 wk. The red traptamer induces transformation (indicated by the dark blue cells) and focus formation. Genes encoding the traptamers were cloned from genomic DNA isolated from cells expanded from individual transformed foci. Traptamer activity was validated by reintroduction into naïve C127 cells. (*D*) C127 cells were infected with high-titer stocks of empty vector (MSCV), or MSCV expressing BPV E5, polyL25, or traptamer LI-11. Photomicrographs were taken at 100× magnification 10 d later. Transformed cells are refractile and irregularly pile on each other after loss of contact inhibition. (*E*) C127 cells were infected with low-titer stocks of the viruses expressing traptamers listed in *B*, and foci of transformed cells were counted 3 wk later. Focus-forming activity (number of foci formed normalized to viral titer) of traptamers is shown relative to E5 activity (set at 100%). V, cells infected with empty MSCV vector. The results show the average of three independent experiments with SE.

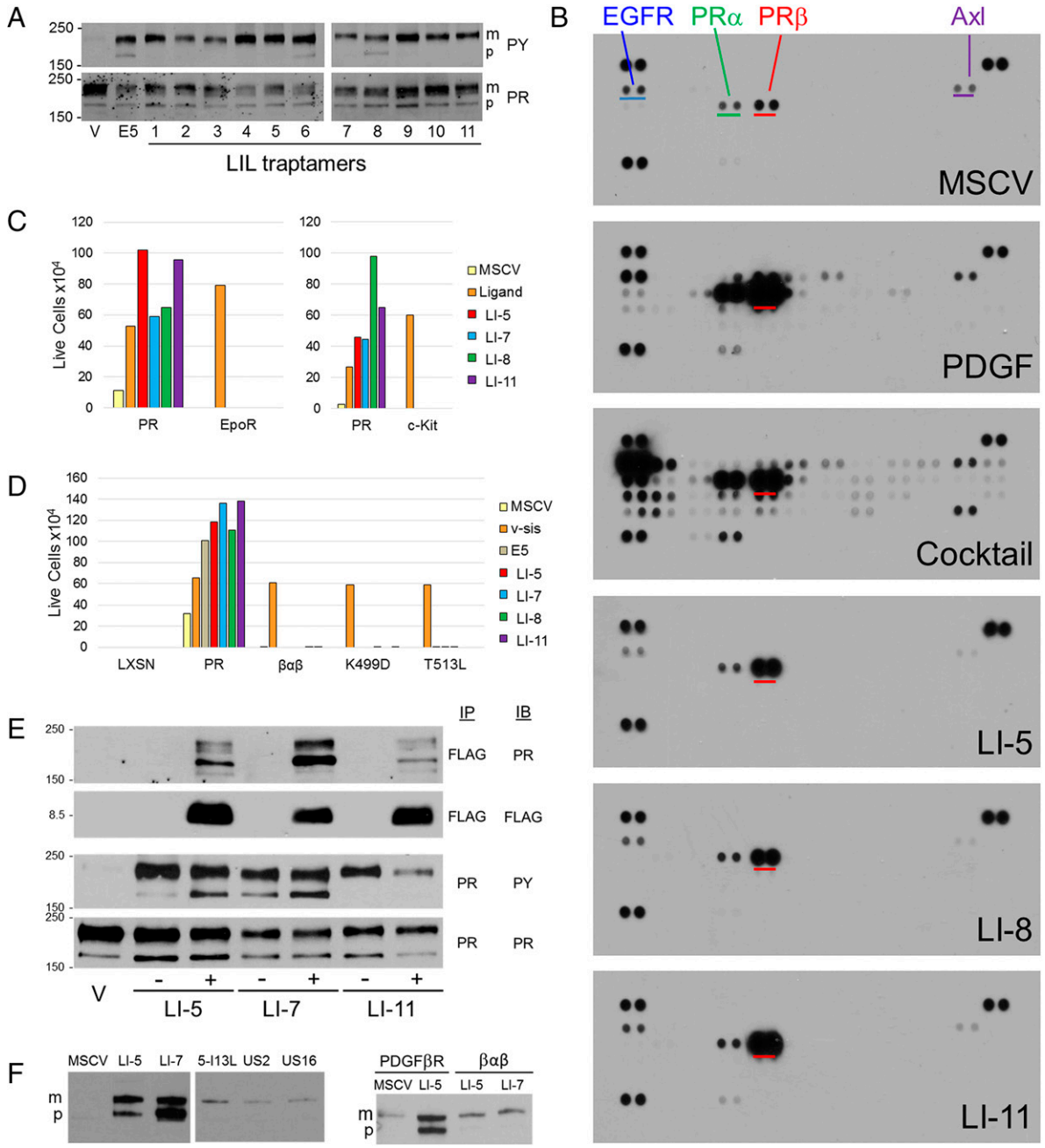

Fig. 2. Traptamers specifically bind and activate the PDGFβR. (*A*) Extracts prepared from C127 cells expressing MSCV vector (V), the E5 protein, or the indicated traptamer were immunoprecipitated with anti-PDGF receptor (PR) antibody and blotted for phosphotyrosine (PY, *Upper*) or PDGFβR (*Lower*). m, mature PDGFβR; p, precursor form of the receptor with immature carbohydrates activated by the E5 protein and some of the traptamers. (*B*) Extracts were prepared from human diploid fibroblasts that were untreated (MSCV), acutely treated with PDGF-DD (PDGF) or with a mixture of growth factors, or stably transformed by traptamer LI-5, LI-8, and LI-11. phospho-RTK arrays were incubated with cell extracts, processed, and exposed to film for the same length of time. The spots representing PDGFβR are underlined in red on all arrays. EGF receptor, PDGFαR, and Axl are underlined in blue, green, and purple, respectively, on the array probed with the extract of control cells. The dark pairs of spots at the corners are standards. (*C*) BaF3 cells expressing the PDGFβR (PR), the erythropoietin receptor (EpoR) (*Left*), or the c-kit (*Right*) were infected with MSCV or MSCV expressing the indicated traptamer or v-sis (for PR cells) or treated with EPO (for EpoR cells) or stem cell factor (SCF, for c-kit cells). The number of live cells was counted 5 d after IL-3 removal. Each experiment was performed three times. Although the inherent variability of this assay precluded a rigorous statistical analysis, the same trends were observed in each experiment. The graphs show the results of a representative experiment. (*D*) BaF3 cells expressing the LXSN empty vector, the PDGFβR, a chimeric PDGFβR containing the transmembrane domain of the PDGFαR (βαβ), or a mutant PDGFβR were infected with MSCV or MSCV expressing v-sis, the E5 protein, or the indicated traptamer. The number of live cells was counted 5 d after IL-3 removal. The experiment was performed three times and the same trends were observed in each experiment. The graph shows the results of a representative experiment. (*E*) Detergent extracts were prepared from untransformed C127 cells harboring MSCV (V) or cells transformed by the indicated FLAG-tagged (+) or untagged (–) traptamer. Extracts were immunoprecipitated (IP) and immunoblotted (IB) with the indicated antibody. (*F*, *Left*) Detergent extracts were prepared from BaF3 cells expressing the PDGFβR and MSCV or the indicated FLAG-tagged traptamer. US2 and US16 are inactive unselected clones. Extracts were imunoprecipitated with anti-FLAG and immunoblotted with anti-PDGF receptor. Both panels were from the same exposure from the same gel; an irrelevant lane was removed. (*Right*) Detergent extracts were prepared from BaF3 cells expressing the PDGFβR or the βαβ chimeric receptor together with MSCV or FLAG-tagged LI-5 or LI-7. Extracts were immunoprecipitated with anti-FLAG and immunoblotted with anti-PDGF receptor.

B). The theoretical complexity of this library is 2^{25} = ~33.5 million. For the experiments reported here, we pooled ~5.5 million bacterial colonies to generate the library, setting an upper limit on the number of independent sequences in the library. Normal C127 mouse fibroblasts were infected at low multiplicity with ~160,000 clones from the library, and rare foci of transformed cells were isolated and expanded into cell lines (Fig. 1*C*). Individual traptamer genes recovered from the DNA of these cells were tested for transforming activity. Eleven recovered traptamers, each with a different sequence (Fig. 1*B*), induced morphologic transformation in C127 cells and in human primary foreskin fibroblasts (HFFs) (Fig. 1*D* and Fig. S1). Several of these traptamers strongly induced focus formation in C127 cells (approximately 40 transformed foci per 1,000 infectious units, similar to the BPV E5 protein) (Fig. 1*E*), although we cannot normalize the transforming activities of the various traptamers to their expression because the abundance of these untagged proteins cannot be measured. In contrast, a clone consisting of methionine followed by 25 leucines (designated polyL25) did not induce transformation (Fig. 1*D*), nor did traptamers randomly chosen from the library without biological selection (Fig. S2). Codon-optimized versions of traptamers LI-7 and LI-11, which contained numerous nucleotide changes that did not affect the amino acid sequence, transformed cells, whereas introduction of a single translation stop codon abolished activity (Fig. S3). Thus, the encoded proteins and not the RNA structures are responsible for transformation.

We hypothesized that the active traptamers induced transformation by interacting with a cellular target protein and activating a mitogenic signaling cascade. To determine if these traptamers, like the E5 protein, activate the PDGFβR, lysates of transformed C127 cells were immunoprecipitated with anti-PDGFβR and immunoblotted with anti-phosphotyrosine antibody. As shown in Fig. 2*A*, *Upper*, basal tyrosine phosphorylation of the PDGFβR in untransformed cells was very low. In contrast, substantial tyrosine phosphorylation of the PDGFβR was detected in cells expressing the E5 protein or any of the active traptamers, indicating that the traptamers activate the PDGFβR. Like the E5 protein, some of the traptamers activated not only the mature form of the receptor, but also a rapidly migrating precursor form of the PDGFβR containing immature carbohydrates. This might result from the localization of different traptamers (and hence the activated PDGFβR) in different intracellular compartments.

To examine the specificity of the traptamers, we used phospho-RTK arrays to assess activation of numerous RTKs. Extracts were prepared from serum-starved HFFs, from HFFs acutely treated with PDGF-DD [which nominally activates only the PDGFβR (26)] or a mixture of growth factors, or from HFFs stably transformed by traptamer LI-5, LI-8, or LI-11. These extracts were incubated with nitrocellulose membranes containing duplicate spots of antibodies that captured 49 individual RTKs, after which the membranes were probed with an antiphosphotyrosine antibody. The intensity of a spot corresponds to the extent of tyrosine phosphorylation of the target RTK. As shown in Fig. 2*B*, the PDGFβR, PDGF-α receptor (PDGFαR), EGF receptor, and Axl displayed a low basal level of constitutive tyrosine phosphorylation in control cells. PDGF-DD caused a dramatic increase in tyrosine phosphorylation of the PDGFβR and lesser increases in several other RTKs, whereas treatment with the growth factor mixture resulted in tyrosine phosphorylation of many RTKs. Strikingly, for all three tested traptamers, the only significant difference compared with control cells was markedly increased tyrosine phosphorylation of the PDGFβR (Fig. 2*B*), demonstrating that the traptamers are highly specific in their ability to activate the PDGFβR. Indeed, in this assay the traptamers were more specific for the PDGFβR than was PDGF itself. Minor differences in the intensity of some faint spots in response to a given traptamer were not reproducible.

We used murine BaF3 cells to determine if traptamer activity required the PDGFβR. BaF3 cells lack endogenous PDGFβR and require IL-3 for growth, but become IL-3–independent if an exogenous growth factor receptor is expressed and activated (Fig. 2*C*) (27, 28). Cells expressing the PDGFβR alone or the traptamer alone grew poorly in the absence of IL-3. However, when PDGFβR was coexpressed with v-sis, the viral homolog of PDGF, or any of the four tested traptamers, the cells proliferated in the absence of IL-3 (Fig. 2*C*), demonstrating that the traptamers require PDGFβR for activity. We also tested the traptamers in BaF3 cells expressing a different RTK, c-kit, or a cytokine receptor, the erythropoietin receptor (EPOR). Both of these receptors conferred IL-3 independence when stimulated with their cognate ligands but did not confer IL-3 independence in response to the traptamers (Fig. 2*C*). Thus, the ability of the traptamers to confer growth factor independence was specific for the PDGFβR. We also used this system to map the domain of the PDGFβR required for productive interaction. We first analyzed BaF3 cells expressing a chimeric receptor in which the transmembrane domain of the PDGFβR was replaced with the transmembrane domain of the PDGFαR (termed βαβ) (Fig. 2*D*) (29, 30). Cells expressing βαβ grew when they coexpressed v-sis, which binds the extracellular domain of the PDGFβR. In contrast, neither E5 nor the traptamers conferred IL-3 independence in cells expressing βαβ. Similarly, v-sis but not the traptamers cooperated with a PDGFβR mutant containing a Lys499 to aspartic acid or a Thr513 to leucine substitution in the transmembrane region. These data demonstrate that these four traptamers require the PDGFβR transmembrane domain for activity and that activity is inhibited by single amino acid substitutions in this segment of the receptor.

Next we wished to investigate whether the traptamers interacted with the PDGFβR, but the very hydrophobic nature of the traptamers complicates biophysical analysis, which traditionally has been used to study the interaction of natural and artificial transmembrane domains with far greater chemical diversity (10, 31, 32). Therefore, we used coimmunoprecipitation to determine if the traptamers were present in a stable complex with the PDGFβR in transformed cells. We first added a FLAG epitope tag to the N terminus of traptamers LI-5, LI-7, and LI-11. Addition of the FLAG-tag caused no more than a ~50% increase or decrease in focus-forming activity of these traptamers (Fig. S4), and all tagged traptamers induced tyrosine phosphorylation of the PDGFβR (Fig. 2*E*, third row). Detergent extracts of transformed C127 cells were immunoprecipitated with anti-FLAG antibody and subjected to immunoblotting. The anti-FLAG antibody detected a ~8.5-kDa band in cells transduced with a FLAG-tagged traptamer gene (Fig. 2*E*, second row), confirming that the tagged traptamers were expressed.

Immunoblotting with anti-PDGFβR showed that anti-FLAG coimmunoprecipitated the mature and precursor PDGFβR from cells expressing the FLAG-tagged traptamers, but not from cells expressing their untagged counterparts, indicating that the traptamers were present in a stable complex with the PDGFβR (Fig. 2*E*, *Top*). We further showed that the defective LI-5 isoleucine 13 to leucine mutant (see next paragraph) as well as two unselected traptamers associated poorly with the PDGFβR in a coimmunoprecipitation experiment in comparison with the active traptamers, LI-5 and LI-7 (Fig. 2*F*). Furthermore, LI-5 and LI-7, which showed the most robust coimmunoprecipitation with wild-type PDGFβR, failed to coimmunoprecipitate with the βαβ receptor chimera, which contains an ectopic transmembrane domain and fails to cooperate with these traptamers in the BaF3 cell growth factor independence assay (Fig. 2*F*). This experiment shows that stable complex formation between the active traptamers and the PDGFβR correlates with transforming activity and requires the transmembrane domain of the receptor.

To determine if specific amino acids in the traptamers were required for activity, we conducted complete mutagenesis of LI-5 and LI-11, in which each leucine was changed individually to isoleucine and vice versa. Mutations at several positions dra-

matically inhibited focus formation (Fig. S5). LI-11 required isoleucines at positions 8, 13, and 22 and leucines at positions 2, 6, 14, 17, 20, 21, and 25 for high focus-forming activity (Fig. 3*A*). Leucine or isoleucine, as appropriate, was enriched at these positions in the active traptamers compared with the unselected traptamers (Fig. S6). LI-5 required a subset of these residues, L6, I13, L14, L17, and L21 (Fig. 3*B*). Thus, the requirements for LI-5 and LI-11 were similar but not identical. These mutagenesis results are summarized in Fig. 3*C*. Interestingly, none of the amino acids that differ between the two traptamers (shown in green in Fig. 3*C*) are required for activity. We then mutated the required isoleucines in LI-5 and LI-11 to valine, another hydrophobic amino acid (Fig. 1*A*). Substitution of a valine at position 13 abolished or severely inhibited the activity of both traptamers, but valine was tolerated at position 8 or 22 in LI-11 (Fig. 3*D*). These data suggest that isoleucine 13 is particularly important for traptamer activity and that the required isoleucines at positions 8 and 22 in LI-11 play a qualitatively different role than isoleucine 13 in transforming activity.

To ask if the inactivating mutations are likely to cause global disruption of traptamer structure, we used all-atom molecular dynamics simulations in explicit POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid bilayers in excess water to generate models of wild-type LI-11 and its defective mutants, 11-I13L and 11-I13V (see *SI Materials and Methods* for the modeling and references). After initial equilibration, the simulations were run for 150 ns. The models of all three traptamers converged on stable, extended α-helical backbone structures roughly perpendicular to the lipid bilayer (tilt ~10° to 25°), with position 13 in the middle of the bilayer (Fig. 3*E*). Notably, the backbone models of wild-type LI-11 and its two mutants were very similar (the average peptide backbone RMSD of wild-type LI-11 compared with either mutant was <1.7 Å). As expected for an α-helix, the radially displayed amino acid side-chains did not interact with each other or with the hydrophilic groups in the protein backbone and the arrangement of side-chain carbon atoms at position 13 in wild-type LI-11 was markedly different from their arrangement in the mutants (Fig. S7). These results imply that the inactivating amino acid substitutions at position 13 do not cause global alterations in the structure of the protein, but rather cause a local structural perturbation at the site of the mutation.

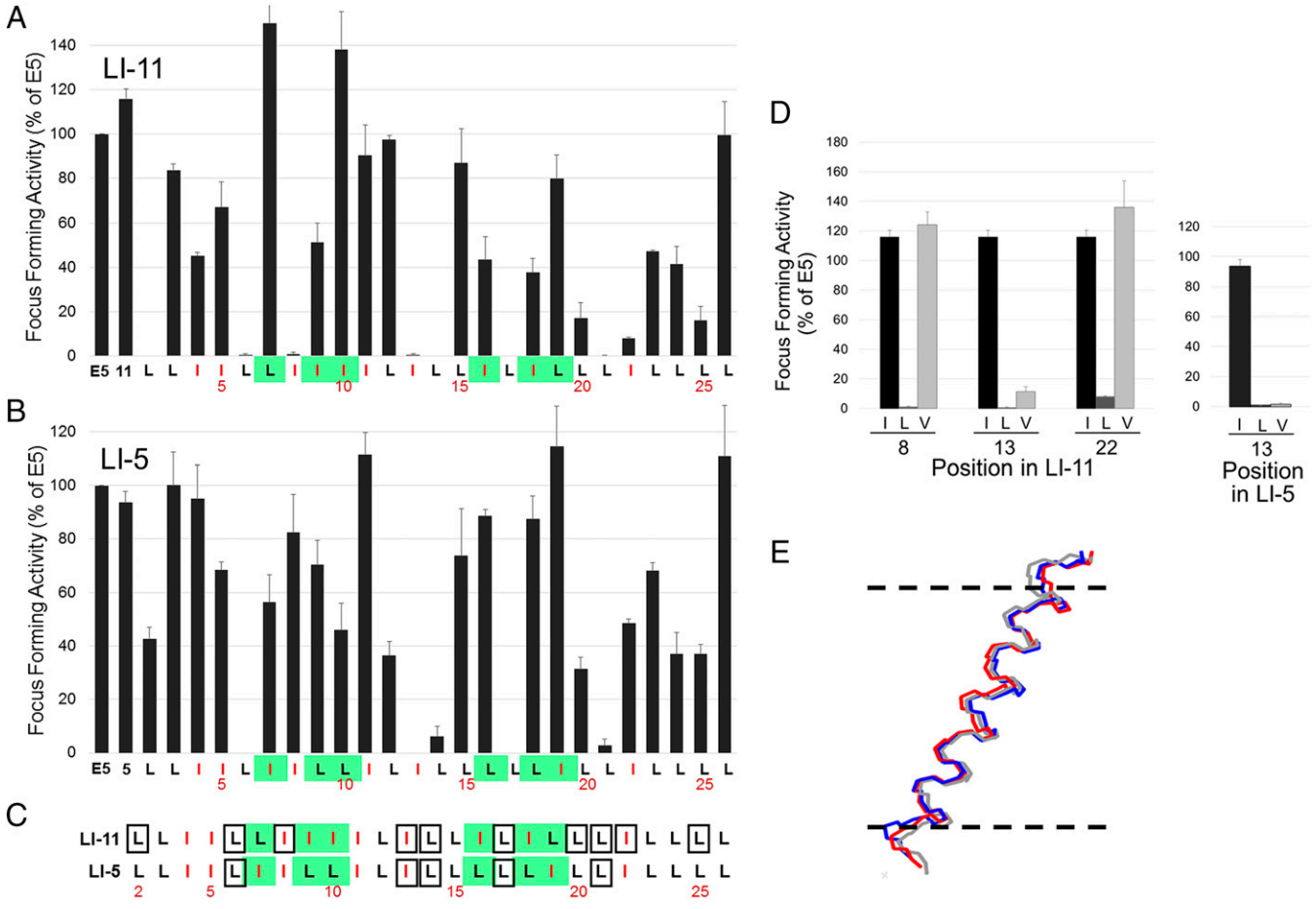

Fig. 3. Comprehensive mutational analysis of traptamers. (*A*) C127 cells were infected with low-titer retrovirus expressing the E5 protein, wild-type LI-11, or an LI-11 mutant in which a leucine was changed to isoleucine or an isoleucine was changed to leucine. Sequence of wild-type LI-11 starting at position 2 is shown at the bottom, with isoleucines shown in red and positions that differ between LI-5 and LI-11 highlighted in green. The bars above the sequence show the activity of the mutant with the substitution at that position. Transformed foci were counted after infected cells were incubated at confluence for 3 wk. Focus-forming activity displayed as in Fig. 1*E*. (*B*) Focus-forming activity of LI-5 mutants tested and displayed as in *A*. (*C*) Sequences of LI-5 and LI-11 are displayed as described in *A*. Residues that when mutated to leucine or isoleucine result in 80% or greater reduction in focus-forming activity are shown in boxes. (*D*) Focus-forming activity was measured and displayed as described in *A* for traptamers containing isoleucine (black), leucine (dark gray), or valine (light gray) at the indicated position in traptamer LI-11 (*Left*) or at position 13 in LI-5 (*Right*). (*E*) Molecular dynamics simulation indicates that the helical backbones of LI-11, LI-11-I13L, and LI-11-I13V are very similar. The diagram shows an overlay of lateral views of the backbones of three aligned helices. The horizontal dashed lines represent the approximate extent of the membrane bilayer hydrophobic region; the amino termini of the proteins are at the top.

Previous work in other laboratories showed that insertion of defined dimerization motifs was sufficient to drive homodimerization of transmembrane peptides (33–35). Because our results suggested that the mutations interfere with the transforming activity of the traptamers by acting locally on the structure of the traptamer, we tested whether introduction of isoleucines at crucial positions in a simple sequence was sufficient for biological activity. We inserted isoleucines into polyL25 at positions 8, 13, and 22 singly and in combination and tested the ability of these mutants to transform cells (Fig. 4*A*). All polyL25 mutants containing isoleucine at position 13 displayed focus-forming activity, whereas mutants lacking Ile13 were inactive, as was polyL25 containing valines at positions 8, 13, and 22 (Fig. 4*B* and Fig. S8). Most strikingly, polyL25 containing a single isoleucine at position 13 (polyL25-I13) induced focus formation and morphologic transformation (Fig. 4 *B* and *C* and Fig. S9). Moreover, polyL25-I8,13,22, and polyL25-I13 induced specific tyrosine phosphorylation of the PDGFβR in transformed HFFs (Fig. 4 *D* and *E*). We also constructed polyL25 mutants containing a single isoleucine at positions 11 through 15 and scored their ability to induce morphologic transformation and focus formation in C127 cells. Although isoleucine at position 13 conferred activity, mutants with an isoleucine at other positions were defective in these assays, demonstrating that the position of the isoleucine was important for transforming activity. (Fig. 4*B* and Fig. S9). We also constructed FLAG-tagged versions of these traptamers and tested their activity in C127 cells. As expected, the isoleucine at position 13 conferred relatively high focus-forming activity (Fig. 4*F*). In addition, an isoleucine at flanking position 12 or 14 conferred ∼10-fold lower activity, whereas traptamers with isoleucine at position 11 or 15 were inactive (Fig. 4*F*), even though all traptamers were expressed at similar levels (Fig. 4*G*). Thus, insertion of a single isoleucine at a particular position in a stretch of leucines is sufficient to specifically activate the PDGFβR and transform cells, and isoleucine at an immediately adjacent position conferred weak transforming activity.

Discussion

We have used a genetic selection to isolate exceptionally simple 26-amino acid proteins with specific biological activity. Remarkably, these proteins contain only leucine and isoleucine and appear to consist of only a single transmembrane helix. Leucine and isoleucine side-chains differ only in the position of a methyl group, lack chemically reactive groups, and are not subject to posttranslational modification. Despite their chemical simplicity, these proteins specifically activate the endogenous PDGFβR but no other receptors in transformed cells, and they confer growth factor independence in cooperation with wild-type PDGFβR but not with other receptors or mutant PDGFβRs containing substitutions in the transmembrane region. In addition, conservative amino acid substitutions at specific positions in the traptamers inhibit activity without affecting the overall structure of the helix, and repositioning a single side-chain methyl group at position 13 from the γ-carbon to the β-carbon (thereby converting leucine to isoleucine)—but not at other positions—in an otherwise monotonous inactive protein is sufficient to generate biological activity.

Our results strongly suggest that traptamers consisting of only leucine and isoleucine can induce cell transformation by selectively interacting in a sequence-specific manner with the transmembrane domain of a cellular protein, most likely the PDGFβR itself, resulting in specific activation of the receptor. Indeed, we showed that the active traptamers but not a defective mutant are present in a stable complex with the wild-type PDGFβR but not with a chimeric PDGFβR containing a foreign transmembrane domain. The specificity of the traptamers suggests that they are unlikely to transform cells by forming aggregates that cause nonspecific receptor clustering. Thus, they differ from repeat-associated non-ATG (RAN) translation products, simple naturally occurring proteins translated in multiple reading frames from trinucleotide or hexanucleotide gene expansions, which cause neurologic disease by nonspecific aggregation and induction of cell toxicity (36). Determining how traptamers with such a limited chemical repertoire specifically recognize a target transmembrane domain will provide new insights into the structural basis for protein–protein recognition. Because leucine and isoleucine side-chains can participate in packing interactions only, further study of these traptamers will provide a focused view of this important class of interactions in the absence of alternative binding solutions provided by chemically diverse side-chains.

These short, simple proteins can be subjected to a level of mutational analysis that would be difficult if not impossible with conventional proteins. A leucine at certain positions in traptamers LI-5 and LI-11 is required for activity, and an isoleucine is required at certain other positions. Whether or not an amino acid is required at a particular position is determined in part by its context. For example, the isoleucines at position 8 and 22 are required in LI-11 but not in LI-5, even though both traptamers contain isoleucine at these positions. Interestingly, none of the positions that differ between LI-5 and LI-11 are required for high activity of either traptamer (Fig. 3*C*). Thus, the amino acids at nonessential positions can dictate whether an amino acid at a shared position is required for optimal activity, presumably by affecting the packing of the helices.

Our mutational analysis also indicated that not all required amino acids are equivalent: the isoleucines at position 8 and 22 in LI-11 cannot be replaced by leucine but can be replaced by valine, which like isoleucine is a β-branched amino acid, suggesting that rotational restriction of a β-branched side-chain at these positions, rather than the isoleucine side-chain per se, is important for LI-11 activity. In contrast, the isoleucine at position 13 in LI-5 and LI-11 cannot be replaced with valine or leucine without loss of activity, implying that the isoleucine side-chain at position 13 makes specific packing contacts essential for transforming activity. The very similar structures of valine, leucine, and isoleucine imply that the terminal methyl group of Ile13 makes an essential packing contact that is missing from the valine mutant. Furthermore, the second terminal methyl group in leucine may sterically clash with its binding partner, or an essential methyl group on the β-carbon may be missing in the leucine mutant.

Single amino acid mutations can have dramatic effects on transmembrane domain interactions. For example, a leucine to isoleucine mutation at position 512 in the transmembrane domain of the PDFGFβR disrupts the direct interaction between the PDGFβR and the E5 protein, preventing transformation (17). Binding energy contributions from contacts at transmembrane helical interfaces have been measured, for example by Fleming et al. (37), who studied the homodimeric interaction of the GlycophorinA transmembrane domain. This study showed the free energy of dissociation of the wild-type dimer is 9 kcal/mol, and that the disruptive L75A and I76A substitutions decrease affinity by 1.1 kcal/mol and 1.7 kcal/mol, respectively. These changes were attributed to changes in van der Waals interactions between the helices. Although it has been suggested that these energies, measured with peptides in detergent micelles, overestimate energies in membranes (38), a change of only ∼1.3 kcal/mol is sufficient to change a "strong" to a "weakly associated" transmembrane dimer (33), probably as a result of constraints imposed by the helical conformation in the lipid bilayer. The limited rotation of isoleucine compared with leucine may make an additional, entropic contribution to binding energy.

It is remarkable that polyL25 is activated by isoleucine at position 13, activated poorly by isoleucine at position 12 or 14, and not activated by isoleucine at position 11 or 15, because in all cases the single isoleucine is embedded in a long stretch of leucines. Thus, an isoleucine near the middle of the lipid bilayer is required for traptamer activity, and activity drops the farther the isoleucine is from this central position. The match between

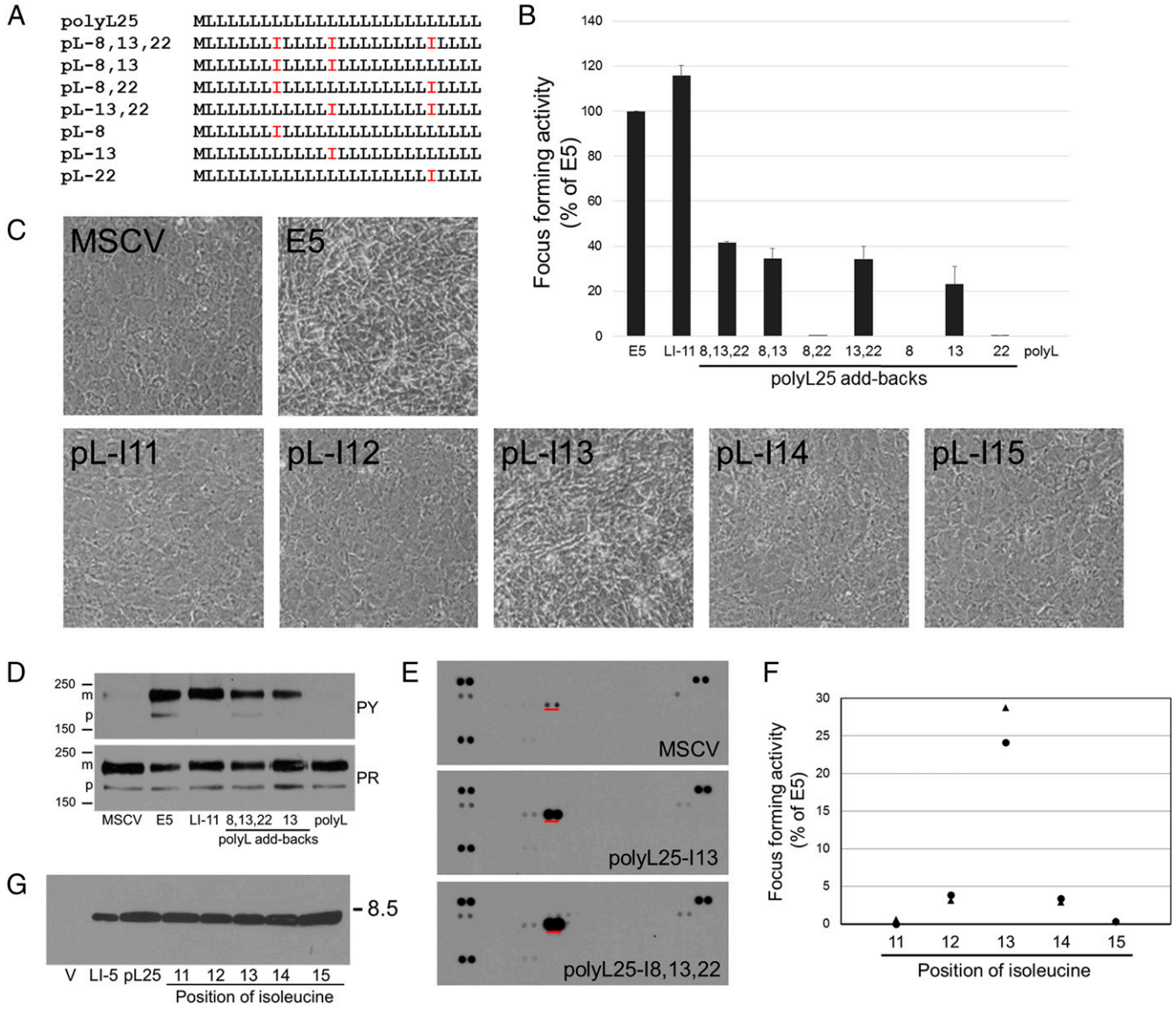

Fig. 4. Insertion of isoleucine into polyL25 restores activity. (*A*) Sequences of add-back mutants in which isoleucines were inserted at the indicated positions in polyL25. (*B*) Traptamers listed in *A* were tested for focus forming activity as in Fig. 1*E*. The results show the average of three independent experiments with SE indicated. (*C*) C127 cells were infected with high-titer stocks of empty MSCV vector or MSCV expressing the E5 protein or a polyL25 mutant containing a single isoleucine at the indicated position. Photomicrographs were taken at 100× magnification 10 d later. The titer of all virus stocks was similar. (*D*) Extracts were prepared from C127 cells harboring empty MSCV or MSCV expressing the E5 protein, LI-11, polyL25-I8,13,22, polyL25-I13, or polyL25 (polyL). Extracts were immunoprecipitated with anti-PDGFβR antibody, subjected to SDS/PAGE, and immunoblotted with antibody recognizing phosphotyrosine (*Upper*) or PDGFβR (*Lower*). (*E*) Extracts were prepared from serum-starved human diploid fibroblasts (MSCV) or cells stably transformed by polyL25-I18,13,22, or polyL25-I13. phospho-RTK arrays were processed as described in Fig. 2*B*. The tyrosine phosphorylated PDGFβR spots are underlined in red. (*F*) C127 cells were infected with low-titer retrovirus stocks of MSCV expressing BPV E5 or a FLAG-tagged polyL25 add-back clone containing isoleucine at the indicated position. Foci were counted 2 to 3 wk after infection, normalized for virus titer, and expressed relative to the activity of E5. Circles and triangles show independent experiments. Each symbol is the average of two plates of foci from a single infection. (*G*) C127 cells were infected with MSCV vector or MSCV expressing FLAG-tagged LI-5, polyL25, or a polyL25 clone with isoleucine inserted at the indicated position. After selection of puromycin-resistant cells, traptamer expression was determined by immunoprecipitation and immunoblotting with anti-FLAG antibody.

the hydrophobic profile of the helix and the hydrophobic profile of the membrane stabilizes the position of a transmembrane helix across a membrane with respect to the displacement of the helix perpendicular to the plane of the membrane. The pattern of activity of the polyL25 add-back traptamers implies that there is a permitted range of helix displacements across the membrane. A displacement from the free-energy minimum will cost energy, and when the displacement energy exceeds the energy of binding of the critical isoleucine with its target, the productive interaction will be lost. We speculate that the isoleucine side-chain at position 13 makes an essential packing contact with the hydrophilic threonine residue at position 513 in the middle of the PDGFβR transmembrane domain, a residue that is required for traptamer activity (Fig. 2*D*) and is proposed to interact directly with Gln17 of the E5 protein (21). Alternatively, these traptamers may act as homodimers, and Ile13 may allow the self-association of two molecules of the traptamer to generate an active dimer.

Our results demonstrate that proteins consisting of only two different amino acids can display specific biological activity. This

finding raises a number of interesting questions, which are under investigation. Do the active traptamers bind directly to the PDGFβR transmembrane domain or do they merely exist in a stable complex that contains the PDGFβR as well as other transmembrane proteins? Is the PDGFβR uniquely susceptible to activation by these ultrasimple traptamers, or can similar proteins modulate the activity of other cellular target proteins? Can similar approaches be used to isolate simple biologically active soluble proteins? Can such simple proteins serve as models for the design of biologically active peptides and peptidomimetic molecules? Do extant genomes encode similar simple biologically active proteins, or did such proteins exist in primordial times before the full complement of amino acids was available? Our results change the view of what can constitute an active protein.

Materials and Methods

Detailed experimental methods are presented in *SI Materials and Methods*.

Retroviral Library Construction. The UDv6 library was constructed by using degenerate oligonucleotides containing library-specific primer binding sites for amplification, a methionine start codon, and 25 degenerate codons encoding equimolar leucine and isoleucine followed by stop codons. To synthesize the randomized segment, equimolar A and C were used in the first position of each codon, T was used in the second position, and equimolar A, C, and T were used in the third position. After primer extension and amplification, the amplification products were ligated into pMSCVpuro retroviral expression vector and transformed into DH10β *Escherichia coli*. Plasmid DNA was harvested from ~5,500,000 pooled ampicillin-resistant colonies. Further details are in *SI Materials and Methods*.

IL-3 Independence Assay. BaF3 lines stably expressing a receptor were generated by infecting BaF3 cells with retrovirus containing a receptor gene cloned in the pLXSN retroviral vector and selection for G418 resistance. The resulting stable cell lines were infected with empty MSCVpuro, or MSCVpuro expressing a traptamer, E5 or v-sis, and selected with puromycin. To assay for IL-3 independent proliferation, washed cells were incubated in RPMI lacking IL-3. Some samples expressing EpoR or c-kit were treated with human erythropoietin or human stem cell factor, respectively. Further details are in *SI Materials and Methods*.

RTK Array. HFFs were infected with MSCV empty vector or viruses expressing a traptamer and selected in puromycin. Cells were serum-starved overnight before harvest. Some control samples were incubated with PDGF-DD or with a mixture of growth factors consisting of PDGF-BB, epidermal growth factor, fibroblast growth factor, neuregulin 1, stem cell factor, Sros 1, and insulin in 50% (vol/vol) FBS before lysis. All subsequence steps were performed as described in the Human Phospho-RTK Array Kit (Proteome Profiler, R&D Systems). Briefly, the cells were lysed in the provided lysis buffer with protease and phosphatase inhibitors, then incubated with the nitrocellulose membranes overnight. Membranes were then incubated with the anti-Phospho-Tyrosine-HRP detection antibody and visualized with Chemi Reagent Mix. Further details are in *SI Materials and Methods*.

Cloning, Tissue Culture, and Biochemical Analysis. Standard procedures were used for tissue culture, retrovirus production, cloning, immunoprecipitation, and immunoblotting. See *SI Materials and Methods* for details.

ACKNOWLEDGMENTS. We thank Emily Cohen and Kelly Chacon for helpful discussions; the Douglas Tobias (University of California, Irvine) group for guidance with the Molecular Dynamics simulations; and Jan Zulkeski for assistance in preparing this manuscript. This study was supported in part by National Institutes of Health training Grants T32GM007499 (to E.N.H.) and T32GM007223 (to R.S.F.); National Institutes of Health Grants CA037157 (to D.D.) and GM073857 (to D.M.E.); and by generous gifts from Ms. Laurel Schwartz (to D.D.). Computational modeling used the Extreme Science and Engineering Discovery Environment was supported by National Science Foundation Grant ACI-1053575 (to D.M.E.).

1. Rennell D, Bouvier SE, Hardy LW, Poteete AR (1991) Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* 222(1):67–88.
2. Cheek S, Zhang H, Grishin NV (2002) Sequence and structure classification of kinases. *J Mol Biol* 320(4):855–881.
3. Hamburger ZA, Brown MS, Isberg RR, Bjorkman PJ (1999) Crystal structure of invasin: A bacterial integrin-binding protein. *Science* 286(5438):291–295.
4. Lehnert U, et al. (2004) Computational analysis of membrane proteins: Genomic occurrence, structure prediction and helix interactions. *Q Rev Biophys* 37(2):121–146.
5. Mackenzie KR (2006) Folding and stability of alpha-helical integral membrane proteins. *Chem Rev* 106(5):1931–1977.
6. Moore DT, Berger BW, DeGrado WF (2008) Protein-protein interactions in the membrane: Sequence, structural, and biological motifs. *Structure* 16(7):991–1001.
7. Popot JL, Engelman DM (2000) Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* 69:881–922.
8. Schneider D, Finger C, Prodöhl A, Volkmer T (2007) From interactions of single transmembrane helices to folding of alpha-helical membrane proteins: Analyzing transmembrane helix-helix interactions in bacteria. *Curr Protein Pept Sci* 8(1):45–61.
9. Bargmann CI, Hung MC, Weinberg RA (1986) Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. *Cell* 45(5):649–657.
10. Yin H, et al. (2007) Computational design of peptides that target transmembrane helices. *Science* 315(5820):1817–1822.
11. DiMaio D (2014) Viral miniproteins. *Annu Rev Microbiol* 68:21–43.
12. Hofmann MW, et al. (2004) De novo design of conformationally flexible transmembrane peptides driving membrane fusion. *Proc Natl Acad Sci USA* 101(41): 14776–14781.
13. Lear JD, Wasserman ZR, DeGrado WF (1988) Synthetic amphiphilic peptide models for protein ion channels. *Science* 240(4856):1177–1181.
14. Burkhardt A, DiMaio D, Schlegel R (1987) Genetic and biochemical definition of the bovine papillomavirus E5 transforming protein. *EMBO J* 6(8):2381–2385.
15. DiMaio D, Guralski D, Schiller JT (1986) Translation of open reading frame E5 of bovine papillomavirus is required for its transforming activity. *Proc Natl Acad Sci USA* 83(6):1797–1801.
16. Schlegel R, Wade-Glass M, Rabson MS, Yang Y-C (1986) The E5 transforming gene of bovine papillomavirus encodes a small, hydrophobic polypeptide. *Science* 233(4762): 464–467.
17. Edwards AP, Xie Y, Bowers L, DiMaio D (2013) Compensatory mutants of the bovine papillomavirus E5 protein and the platelet-derived growth factor β receptor reveal a complex direct transmembrane interaction. *J Virol* 87(20):10936–10945.
18. Lai CC, Henningson C, DiMaio D (1998) Bovine papillomavirus E5 protein induces oligomerization and trans-phosphorylation of the platelet-derived growth factor beta receptor. *Proc Natl Acad Sci USA* 95(26):15241–15246.
19. Petti L, DiMaio D (1992) Stable association between the bovine papillomavirus E5 transforming protein and activated platelet-derived growth factor receptor in transformed mouse cells. *Proc Natl Acad Sci USA* 89(15):6736–6740.
20. Petti L, Nilson LA, DiMaio D (1991) Activation of the platelet-derived growth factor receptor by the bovine papillomavirus E5 transforming protein. *EMBO J* 10(4): 845–855.
21. Talbert-Slagle K, DiMaio D (2009) The bovine papillomavirus E5 protein and the PDGF β receptor: It takes two to tango. *Virology* 384(2):345–351.
22. Cammett TJ, et al. (2010) Construction and genetic selection of small transmembrane proteins that activate the human erythropoietin receptor. *Proc Natl Acad Sci USA* 107(8):3447–3452.
23. Freeman-Cook LL, DiMaio D (2005) Modulation of cell function by small transmembrane proteins modeled on the bovine papillomavirus E5 protein. *Oncogene* 24(52):7756–7762.
24. Freeman-Cook LL, et al. (2004) Selection and characterization of small random transmembrane proteins that bind and activate the platelet-derived growth factor beta receptor. *J Mol Biol* 338(5):907–920.
25. Chacon KM, et al. (2014) De novo selection of oncogenes. *Proc Natl Acad Sci USA* 111(1):E6–E14.
26. Bergsten E, et al. (2001) PDGF-D is a specific, protease-activated ligand for the PDGF beta-receptor. *Nat Cell Biol* 3(5):512–516.
27. Drummond-Barbosa DA, Vaillancourt RR, Kazlauskas A, DiMaio D (1995) Ligand-independent activation of the platelet-derived growth factor beta receptor: requirements for bovine papillomavirus E5-induced mitogenic signaling. *Mol Cell Biol* 15(5):2570–2581.
28. Palacios R, Steinmetz M (1985) Il-3–dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes in vivo. *Cell* 41(3):727–734.
29. Nappi VM, Schaefer JA, Petti LM (2002) Molecular examination of the transmembrane requirements of the platelet-derived growth factor beta receptor for a productive interaction with the bovine papillomavirus E5 oncoprotein. *J Biol Chem* 277(49):47149–47159.
30. Petti LM, Reddy V, Smith SO, DiMaio D (1997) Identification of amino acids in the transmembrane and juxtamembrane domains of the platelet-derived growth factor receptor required for productive interaction with the bovine papillomavirus E5 protein. *J Virol* 71(10):7318–7327.
31. Call ME, Pyrdol J, Wiedmann M, Wucherpfennig KW (2002) The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell* 111(7):967–979.
32. Wucherpfennig KW, Gagnon E, Call MJ, Huseby ES, Call ME (2010) Structural biology of the T-cell receptor: Insights into receptor assembly, ligand recognition, and initiation of signaling. *Cold Spring Harb Perspect Biol* 2(4):a005140.

33. Nash A, Notman R, Dixon AM (2015) De novo design of transmembrane helix-helix interactions and measurement of stability in a biological membrane. *Biochim Biophys Acta* 1848(5):1248–1257.
34. Orzáez M, Lukovic D, Abad C, Pérez-Payá E, Mingarro I (2005) Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin A dimerisation motif. *FEBS Lett* 579(7):1633–1638.
35. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM (2000) Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat Struct Biol* 7(2):154–160.
36. Cleary JD, Ranum LP (2013) Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum Mol Genet* 22(R1):R45–R51.
37. Fleming KG, Ackerman AL, Engelman DM (1997) The effect of point mutations on the free energy of transmembrane α-helix dimerization. *J Mol Biol* 272(2):266–275.
38. MacKenzie KR, Fleming KG (2008) Association energetics of membrane spanning alpha-helices. *Curr Opin Struct Biol* 18(4):412–419.
39. Naviaux RK, Costanzi E, Haas M, Verma IM (1996) The pCL vector system: Rapid production of helper-free, high-titer, recombinant retroviruses. *J Virol* 70(8):5701–5705.
40. Talbert-Slagle K, et al. (2009) Artificial transmembrane oncoproteins smaller than the bovine papillomavirus E5 protein redefine sequence requirements for activation of the platelet-derived growth factor beta receptor. *J Virol* 83(19):9773–9785.
41. Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys J* 97(1):50–58.
42. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.
43. Yu W, He X, Vanommeslaeghe K, MacKerell AD, Jr (2012) Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem* 33(31):2451–2468.
44. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.
45. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577–8593.
46. Grubmuller H, Heller H, Windemuth A, Schulten K (1991) Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol Simul* 6(3):121–142.
47. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* 103(11):4613–4621.
48. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101(5):4177–4189.
49. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.