doi:10.1016/j.jmb.2004.03.044 J. Mol. Biol. (2004) 338, 907–920

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Selection and Characterization of Small Random Transmembrane Proteins that Bind and Activate the Platelet-derived Growth Factor β Receptor

Lisa L. Freeman-Cook[1], Ann M. Dixon[2], Jennifer B. Frank[1], Yu Xia[2] Lara Ely[1], Mark Gerstein[2], Donald M. Engelman[2] and Daniel DiMaio[1]*

[1]*Department of Genetics, Yale University School of Medicine New Haven, CT 06510, USA*

[2]*Department of Molecular Biophysics and Biochemistry Yale University School of Medicine, New Haven, CT 06510, USA*

**Corresponding author*

Growth factor receptors are typically activated by the binding of soluble ligands to the extracellular domain of the receptor, but certain viral transmembrane proteins can induce growth factor receptor activation by binding to the receptor transmembrane domain. For example, homodimers of the transmembrane 44-amino acid bovine papillomavirus E5 protein bind the transmembrane region of the PDGF β receptor tyrosine kinase, causing receptor dimerization, phosphorylation, and cell transformation. To determine whether it is possible to select novel biologically active transmembrane proteins that can activate growth factor receptors, we constructed and identified small proteins with random hydrophobic transmembrane domains that can bind and activate the PDGF β receptor. Remarkably, cell transformation was induced by approximately 10% of the clones in a library in which 15 transmembrane amino acid residues of the E5 protein were replaced with random hydrophobic sequences. The transformation-competent transmembrane proteins formed dimers and stably bound and activated the PDGF β receptor. Genetic studies demonstrated that the biological activity of the transformation-competent proteins depended on specific interactions with the transmembrane domain of the PDGF β receptor. A consensus sequence distinct from the wild-type E5 sequence was identified that restored transforming activity to a non-transforming poly-leucine transmembrane sequence, indicating that divergent transmembrane sequence motifs can activate the PDGF β receptor. Molecular modeling suggested that diverse transforming sequences shared similar protein structure, including the same homodimer interface as the wild-type E5 protein. These experiments have identified novel proteins with transmembrane sequences distinct from the E5 protein that can activate the PDGF β receptor and transform cells. More generally, this approach may allow the creation and identification of small proteins that modulate the activity of a variety of cellular transmembrane proteins.

Keywords: cell transformation; oncogene; TOXCAT; bovine papillomavirus; E5 protein

Introduction

Receptor tyrosine kinases are normally activated by soluble ligands that bind to the extracellular domain of the receptor and induce receptor dimerization. Certain viral transmembrane proteins that do not resemble these ligands can also activate endogenous receptor signaling pathways and induce tumorigenic transformation by interacting with the transmembrane domain of growth factor receptors.[1] For example, the gp55 protein of the Friend leukemia virus binds and activates the erythropoietin receptor, and the E5 protein of

Supplementary data associated with this article can be found at doi: 10.1016/j.jmb.2004.03.044

Abbreviations used: PDGF, platelet-derived growth factor; CAT, chloramphenicol acetyltransferase.

E-mail address of the corresponding author: daniel.dimaio@yale.edu

bovine papillomavirus (BPV) binds and activates the platelet-derived growth factor β (PDGF β) receptor.[1] In both of these cases the interactions driving receptor activation occur primarily within the lipid bilayer. To test the possibility that side-by-side protein–protein interactions involving diverse transmembrane segments can induce growth factor receptor activation, we devised a strategy to identify novel transmembrane domains that can recognize the transmembrane domain of the PDGF β receptor and induce receptor activation and cell transformation.

We used the 44-amino acid BPV E5 protein as a scaffold to identify small transmembrane proteins that activate the PDGF β receptor. The E5 protein, which induces transformation of cultured rodent fibroblasts,[2] has a very hydrophobic transmembrane domain that is interrupted only by a glutamine at position 17.[3] In transformed cells, disulfide-linked homodimers of the E5 protein appear to be localized largely to the membranes of the endoplasmic reticulum and Golgi apparatus in a type II transmembrane orientation.[3–5] Transformation is inhibited by mutations at Gln17, Asp33, and the carboxy-terminal cysteine residues required for homodimer formation, but the transforming activity of the E5 protein is not impaired by a variety of single hydrophobic substitution mutations in the central hydrophobic domain.[6–12] In limited surveys, a small number of E5 mutants with apparently random transmembrane domains have been identified that can transform cells, but the mechanism of transformation by these mutants has not been explored.[9,13,14]

Although the wild-type E5 protein interacts with several cellular proteins,[10,15–17] its primary transforming function appears to be mediated by activation of the cellular PDGF β receptor, a type I transmembrane receptor tyrosine kinase.[8,18–21] The E5 dimer is present in a stable complex with two tyrosine-phosphorylated forms of the PDGF β receptor in transformed cells, a mature form and an intracellular form with immature carbohydrates.[16,21–23] Complex formation results in the ligand-independent dimerization of two molecules of the PDGF β receptor, *trans*-phosphorylation of each receptor in the dimeric complex, and constitutive association with SH2 domain-containing signaling proteins.[8,18,21,23–25] The interaction between the E5 protein and the PDGF β receptor is mediated by transmembrane and juxtamembrane interactions and is quite specific, in that the E5 protein can bind and activate the wild-type PDGF β receptor but not PDGF β receptor mutants with a foreign transmembrane domain or other receptor tyrosine kinases.[18,19,22,26–28] Furthermore, the E5 protein is able to bind and activate mutant PDGF β receptors that lack the extracellular ligand binding domain, and it can bind the transmembrane segment of the PDGF β receptor linked to very little intracellular or extracellular sequence.[18,27]

We proposed that the E5 homodimer transits the membrane as a symmetric left-handed coiled-coil with an interface consisting of Leu14, Gln17, Leu21, Leu24, and Phe28,[5,11] although an alternative model has been proposed.[29] Gln17 stabilizes the E5 dimer, presumably by making interhelical hydrogen bonds.[5,8] Extensive mutagenesis experiments also suggest the existence of a hydrogen bond between Gln17 of the E5 protein and Thr513 of the receptor, and a salt bridge between Asp33 of the E5 protein and Lys499 in the PDGF β receptor.[7–10,12,28] Although Thr513 and Lys499 are both required for the interaction between the E5 protein and the PDGF β receptor, many other point mutations in the PDGF β receptor transmembrane domain also impair transformation, implying that there are additional contacts between the PDGF β receptor and the E5 protein[28,30] (and L.E., D. Mattoon & D.D., unpublished results).

The experiments summarized above revealed that receptor tyrosine kinases can be activated by transmembrane proteins unrelated to their normal ligand. In order to isolate novel proteins that can recognize and bind cellular transmembrane proteins, we developed a method to select small randomized transmembrane proteins that can activate the PDGF β receptor. We identified numerous transmembrane sequences quite distinct from the E5 protein that can activate the PDGF β receptor, suggesting an unexpectedly wide range of sequences capable of participating in productive interactions with the transmembrane domain of a growth factor receptor.

Results

Experimental strategy

In order to isolate novel small transmembrane proteins that activated the PDGF β receptor, we constructed a library in which 15 transmembrane amino acid residues of the 44-amino acid E5 protein were replaced with random hydrophobic sequences. The transmembrane glutamine at position 17 was retained in this library because it plays a role both in homodimerization of E5 monomers and in the interaction between the E5 protein and the PDGF β receptor. Degenerate oligonucleotides were synthesized that encoded a random mixture of Leu, Val, Ile, Met, and a low level of Ser, Thr, Pro, and Ala. The leucine-rich composition of the encoded amino acids was chosen to reflect the abundance of leucine in transmembrane domains, and particularly in papillomavirus E5 proteins. In addition, we excluded phenylalanine from the library, even though there are conserved transmembrane phenylalanine residues in the E5 proteins from various fibropapillomaviruses,[31] in an attempt to identify novel sequence motifs able to interact with the PDGF β receptor transmembrane domain. The

oligonucleotides were converted to double-stranded form, amplified by PCR, and used to replace the central portion of the wild-type E5 gene in a retroviral vector.

Murine C127 fibroblasts, which express endogenous PDGF β receptor, were infected with the resulting library, and infected cells were incubated at confluence to allow focus formation (Figure 1(b)). We estimate that 1×10^5 viruses were screened for transforming activity, which represent a small fraction of possible transmembrane sequences, given the number of positions randomized. Individual transformed foci were picked and expanded, and transforming virus was recovered by infecting transformed cells with a helper virus to supply viral replication functions. The rescued virus was used in a second round of infection of naïve C127 cells, and transformed cells were again selected. The retroviral inserts were amplified from genomic DNA isolated from transformed cells, subcloned, sequenced, and reintroduced into naïve C127 cells to confirm their ability to induce focus formation.

Sequence analysis of the library clones

We sequenced 45 randomly sampled, unselected clones from the library, which were shown to be non-transforming. These unselected clones had the expected amino acid composition and appeared to have a random distribution of the possible amino acids at each position (see Figure 3). Approximately 10% of the randomized library clones induced foci. We recovered 37 transforming clones from the library screen, and four additional transforming clones were identified within the set of unselected clones. Figure 2 lists representative transforming and non-transforming clones, and their relative focus-forming activity. Figure 1S in the Supplementary Material shows the sequences of all of the transforming and non-transforming clones.

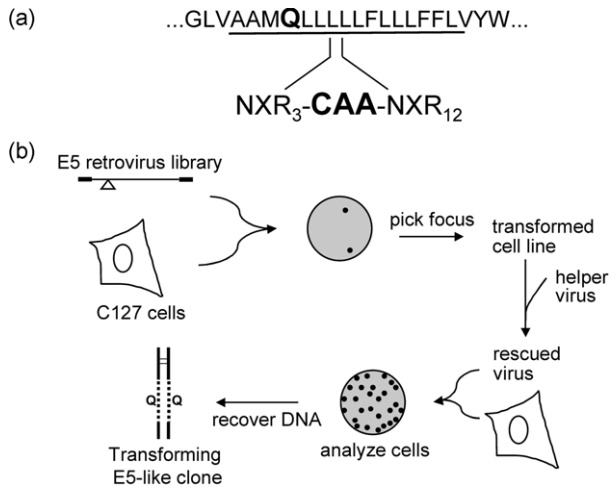

Figure 1. (a) Library design. Fifteen transmembrane amino acids of the BPV E5 protein were replaced with random, predominantly hydrophobic sequences encoded by NXR codons as described in Materials and Methods. A CAA codon encoding glutamine 17 was maintained. (b) Diagram depicting selection strategy. Rare foci arising after infection of C127 fibroblasts with the randomized library were picked and expanded. A helper virus was used to rescue the transforming virus, which was then used to re-infect naïve C127 cells. Genomic DNA was recovered from resulting transformed cell lines, and the retroviral inserts were amplified, subcloned, and sequenced.

The great majority of randomized sequences did not support transformation, but a wide variety of small transmembrane proteins were capable of transforming cells. Inspection of the sequences of the transforming clones revealed that transformation did not require the wild-type E5 amino acid at any randomized position. In fact, nearly any tested hydrophobic amino acid was tolerated at each position within the transmembrane domain. We used Sequence Logo to graphically represent the transmembrane sequences of the non-transforming and transforming proteins (Figure 3). In the Sequence Logo graphs, the frequencies of the amino acids at each position are represented by the relative heights of the letters. For the non-transforming proteins, the composition at each position was quite similar and reflected the expected frequency of each amino acid based on the design of the library. The transforming proteins were strikingly different, with a bias toward a particular amino acid at several positions. For example, position 24 was strongly biased toward valine, and position 25 was strongly biased toward leucine.

We used a χ^2 statistical test to quantify the difference in amino acid distribution at each position between the transforming and non-transforming proteins. Figure 4 shows the difference score calculated for each position (see Materials and Methods for a definition of difference score). At positions 18, 20, 24, and 28, the difference score is greater than 1.3, corresponding to a greater than 95% probability that the distribution of amino acids at these positions is different between the transforming and non-transforming proteins. When these positions were plotted on a helical wheel with 3.5 residue periodicity, consistent with previous evidence that the E5 dimer transits the membrane as a left-handed coiled-coil,[5,11] they all fell on one side of the helix, overlapping with the proposed E5 homodimer interface (Figure 5).

Identification and testing of a consensus sequence in the transforming proteins

We used a computational algorithm to identify a consensus sequence that best discriminated between the sequences of the transforming small transmembrane proteins and the sequences of the non-transforming proteins. The optimal six-amino acid residue sequence that met this criterion was Leu18, Val20, Val24, Leu25, Val28, and Leu29. This consensus sequence has a *t*-score of 8.44,

Transforming Clones		CS	relative activity
LFC1-T1	LLLQMVMLVLVLVLVL	4	++++
LFC1-T2	LVLQLSLMLLAVLLLL	2	++++
LFC1-T4	LLMQLAVAILLLLLML	4	+
LFC1-T8	MLVQLVVMLMILLVVV	4	+++
LFC1-T9	VILQLMVLVVMLLMVL	5	++++
LFC1-T13	VLLQLVVLVLVLLIMM	4	+++
LFC1-T14	LVLQLIMLVLVLLVVL	5	++++
LFC1-T16	MVMQVVVLLVMLLLLM	2	+++
LFC1-T20	LLLQLMMLLMVLLLVL	5	+++
LFC1-T36	LVLQLLLLLLLLLLVV	3	+

Non-Transforming Clones			
LFC1-NT1	LLLQVVLLVVVLLLML	3	-
LFC1-NT2	LLVQVVVLVLLMLLLM	1	-
LFC1-NT10	LMLQVVLVVVVMLLLL	2	-
LFC1-NT12	LVMQTILVLLVLVLLV	2	-
LFC1-NT15	LLVQVLIMLLMVVVLV	0	-
LFC1-NT16	MMLQMILLLMLLMIVI	2	-
LFC1-NT17	ILMQLLVLVLVMLIVV	4	-
LFC1-NT18	VLLQLLLMLIVLMVLV	3	-
LFC1-NT22	VVMQLLLVLLVLLVLL	4	-
LFC1-NT32	MLVQVLMLVLLIMLLL	1	-

Figure 2. Transmembrane sequences of ten representative transforming (top) and non-transforming proteins (bottom). The far right column lists the relative focus-forming activity of the proteins. ++++ Corresponds to >130% of wild-type E5 activity, +++ to 70–130%, ++ to 30–70%, + to 10–30%, and − to <10% of wild-type E5 activity. For each protein, the number of amino acid residues that match the six-position consensus sequence is shown in the column labeled CS.

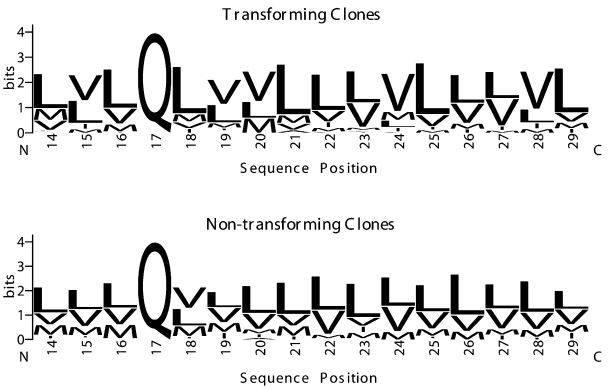

Figure 3. Sequence logo[46] representation of the transmembrane sequence of 41 transforming (top) and 45 non-transforming (bottom) proteins. The height of each letter is proportional to the frequency of the amino acid at that position.

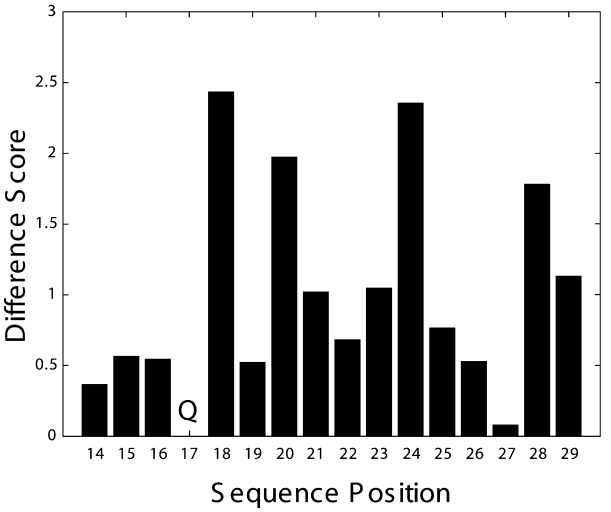

Figure 4. A graphical representation of the position-specific differences in amino acid distribution between the transforming and non-transforming E5-like proteins. At each position, the difference score is defined as $-\log_{10}(P_i)$, where P_i is the probability that the two distributions are identical, as determined from the χ^2 value.

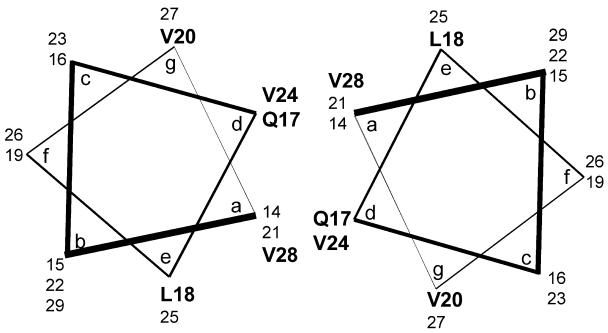

Figure 5. Helical wheel diagram of transmembrane protein dimers. This representation assumes the symmetric left-handed coiled-coil structure predicted for the wild-type E5 protein. The positions in bold, 18, 20, 24, and 28, are those positions where there is a greater than 95% chance that the distribution of amino acids is different between the transforming and non-transforming proteins. The most abundant amino acid in the transforming proteins is shown at these positions. Also shown in bold is the glutamine at position 17. The residues at vertices a and d form the proposed homodimer interface.

corresponding to a probability of less than 10^{-15} that the transforming and non-transforming proteins have identical distributions of amino acids at these six positions. Figure 6 is a histogram enumerating the number of positions at which each transforming and non-transforming protein matches the consensus. Greater than 70% of the transforming proteins matched four or five of the six consensus positions, and the matched positions varied from clone to clone. In contrast, less than 10% of the non-transforming proteins matched four positions, and none matched more than four. There was no obvious correlation between the number of matches to the consensus and focus-forming activity in C127 cells (Figure 2 and Supplemental Figure 1S).

The consensus sequence identified above is different from the sequence of the wild-type E5 protein (Figure 7). In the wild-type sequence, there are phenylalanine residues at positions 23, 27, and 28, but the consensus sequence included valine residues at positions 20, 24, and 28. Not only are two of the three positions of the non-leucine residues different in the phenylalanine- and valine-based motifs, but the chemical properties of phenylalanine and valine are also quite different. To determine if a simple transmembrane protein containing the consensus sequence was capable of inducing transformation, the transmembrane domain of the E5 protein was replaced with three sequences: (1) poly-leucine containing glutamine at position 17 (pL-Q), (2) poly-leucine with glutamine 17 and phenylalanine 23, 27, and 28 (pL-QF), similar to the wild-type E5 sequence, and (3) poly-leucine with glutamine 17 and valine 20, 24, and 28 (pL-QV), representing the consensus sequence derived from the analysis of the active proteins recovered from the library (Figure 7). The pL-Q clone did not induce focus formation. In contrast, either the phenylalanine or valine motif restored efficient focus formation to the poly-leucine protein. Thus, both the phenylalanine motif and the consensus valine motif supported efficient cell transformation.

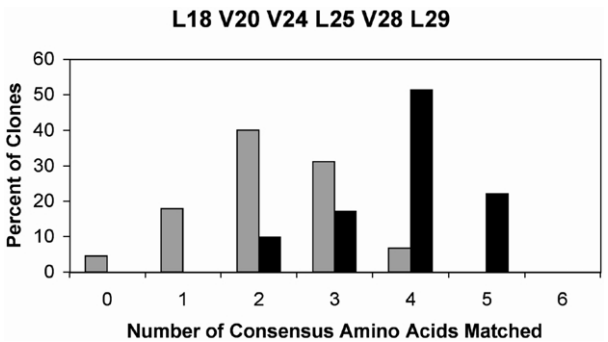

Figure 6. Analysis of the consensus sequence. The histogram shows the percentage of non-transforming (gray) and transforming (black) proteins that match the indicated number of positions in the consensus sequence shown at the top.

Productive interaction between the active transmembrane proteins and the PDGF β receptor

To determine whether the transformation-competent small transmembrane proteins transformed cells by activating the PDGF β receptor, stable C127 cell lines expressing proteins recovered from the library or the poly-leucine proteins were analyzed. All cell lines expressed similar levels of the endogenous PDGF β receptor (Figure 8(a) and (b), top panel). Cell extracts were also immunoprecipitated with an antibody specific for the PDGF β receptor and then immunoblotted with a monoclonal antibody that recognizes phosphotyrosine (Figure 8(a), second panel). Wild-type E5 protein (far right) induced marked phosphorylation of both the mature and precursor forms of the receptor, whereas the empty vector control (C) showed weak background phosphorylation of the mature form only. The mature form of the PDGF β receptor was phosphorylated significantly more than the vector control in cells transformed by the small transmembrane proteins, but there was no significant increase in receptor phosphorylation above background levels in the three non-transformed cell lines tested. The pL-Q protein induced weak phosphorylation of the receptor, the pL-QV protein induced significant phosphorylation, predominantly on the mature form of the receptor, and the pL-QF protein induced dramatic phosphorylation of both the mature and precursor forms of the receptor (Figure 8(b), second panel).

To determine if the small transmembrane proteins formed a stable complex with the PDGF β receptor, cell extracts were immunoprecipitated with an antibody that recognizes the C terminus of the E5 protein, a segment that was not

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	Focus formation
E5	**A**	**M**	**Q**	L	L	L	L	L	**F**	L	L	L	**F**	**F**	L	+++
pL-Q	L	L	**Q**	L	L	L	L	L	L	L	L	L	L	L	L	-
pL-QF	L	L	**Q**	L	L	L	L	L	**F**	L	L	L	**F**	**F**	L	++
pL-QV	L	L	**Q**	L	L	**V**	L	L	L	**V**	L	L	L	**V**	L	++

Figure 7. Amino acid sequence and transforming activity of poly-leucine-based proteins. The amino acid sequence of the transmembrane domain of the wild-type E5 protein, the poly-leucine protein containing Gln17 (pL-Q), the poly-leucine protein with Gln17 and the phenylalanine motif (pL-QF), and the poly-leucine protein with Gln17 and the valine motif (pL-QV) are shown. Non-leucine residues are bold. The column on the right lists the relative focus formation activity of each of the proteins, scored as in Figure 2.

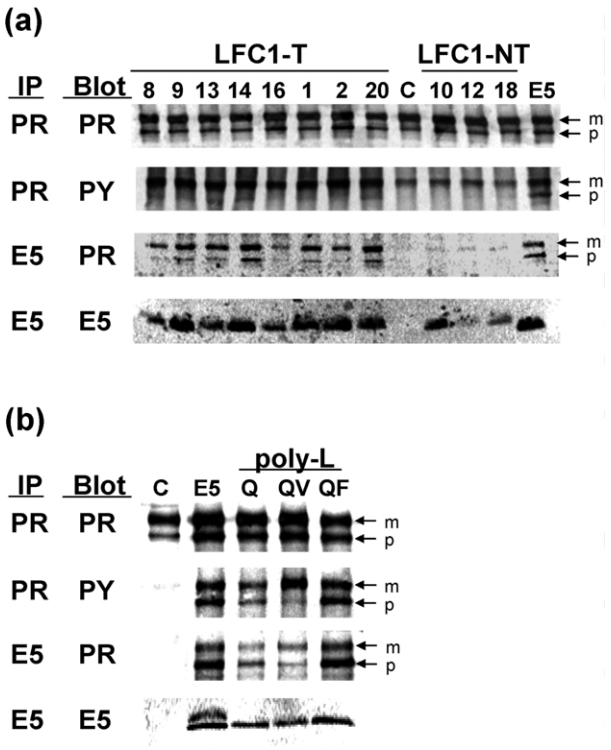

Figure 8. Biochemical analysis of the small transmembrane proteins. Cell extracts were immunoprecipitated and blotted with the indicated antibodies in order to detect expression of the endogenous PDGF β receptor (top), receptor tyrosine phosphorylation (second panel), association between the small transmembrane proteins and the PDGF β receptor (third panel), and expression of the small transmembrane proteins (bottom). The antibodies used for immunoprecipitation (IP) and Western blotting (blot) are indicated: E5, E5 C terminus; PR, PDGF β receptor; PY, phosphotyrosine. Bands corresponding to the mature (m) and precursor (p) forms of the receptor are indicated by the arrows on the right. (a) Analysis of cells expressing the indicated transforming proteins (LFC1-T), an empty vector control (C), the indicated non-transforming proteins (LFC1-NT), and the wild-type E5 protein (E5). (b) Analysis of cells expressing the controls in (a) or the poly-leucine proteins with glutamine alone (pL-Q), or the valine (pL-QV) or phenylalanine (pL-QF) motif.

randomized in the library construction. The immunoprecipitates were immunoblotted with an antibody specific for the PDGF β receptor. The wild-type E5 protein and all tested transforming proteins, including the active poly-leucine proteins, formed a stable complex with both the mature and immature forms of the PDGF β receptor, although the strength of the signal and the relative abundance of the two receptor forms differed among the various transformed cell lines (Figure 8(a) and (b), third panel). However, the non-transforming proteins recovered from the library failed to associate significantly with the receptor. Interestingly, although the pL-Q protein was not capable of inducing transformation, it was able to interact with the PDGF β receptor at a level similar to the transforming pL-QV protein. It is possible that the interaction between pL-Q and the PDGF β receptor was aberrant, causing phosphorylation to occur at a set of sites that were not sufficient to activate the downstream signaling cascades required for cellular transformation.

To test whether the non-transforming small transmembrane proteins fail to activate or associate with the receptor because they are poorly expressed, cell extracts were immunoprecipitated and immunoblotted with the E5 antibody. The wild-type E5 protein and the transforming proteins were highly expressed, as were the non-transforming LFC1-NT10 and LFC1-NT16 proteins (Figure 8(a), bottom panel, and data not shown). All three poly-leucine proteins were highly expressed (Figure 8(b), bottom panel), including pL-Q, which is transformation-defective. The non-transforming proteins LFC1-NT12 and -NT18 were expressed at a lower but detectable level, which presumably contributes to our inability to detect receptor activation and complex formation.

Role of the PDGF β receptor in transformation by the small transmembrane proteins

As a further test of the mechanism of transformation, stable C127 cell lines expressing the wild-type E5 protein or one of several transforming small transmembrane proteins were grown with and without the specific PDGF receptor kinase inhibitor AG1295 (Figure 9, and data not shown). In the absence of AG1295, the wild-type E5 protein

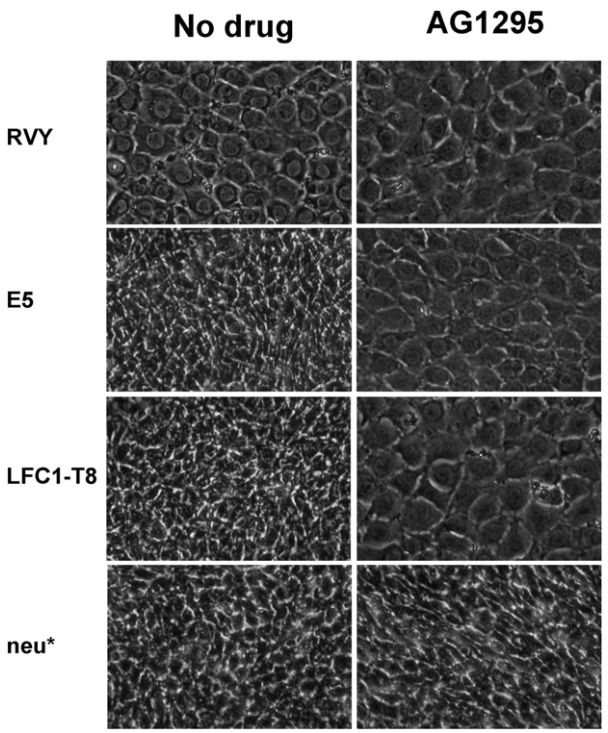

Figure 9. Effect of the PDGF receptor kinase inhibitor AG1295 on transformation of C127 cells. Photomicrographs of C127 cells stably expressing an empty RVY vector, the wild-type E5 protein, the LFC1-T8 transforming protein, or the *neu* oncoprotein are shown in the absence of AG1295 (left) and after four days of treatment with the drug (right).

and each of six transforming proteins tested grew to much higher density and acquired a transformed appearance (LFC1-T8 is shown as a representative example). The transformed cells reverted to a non-transformed morphology in the presence of AG1295, indicating that the transformed phenotype was due to the activity of the PDGF β receptor. As expected, the drug had no effect on the morphology of cell lines expressing an empty retroviral vector (top panels), non-transforming small transmembrane proteins, or cells that were transformed with the unrelated oncogenic *neu* receptor tyrosine kinase (bottom panels, and data not shown). AG1295 also reverted the transformed phenotype induced by the poly-leucine-QV and poly-leucine-QF proteins (data not shown).

We used a transient reporter gene assay to measure the ability of the small transmembrane proteins to signal through the PDGF β receptor, and to determine whether the active proteins interacted specifically with the PDGF β receptor transmembrane domain. CV1 monkey kidney cells, which do not express endogenous PDGF β receptor, were co-transfected with four plasmids. All cells received a constitutively expressed *Renilla* luciferase plasmid to control for transfection efficiency and a reporter plasmid containing three tandem copies of the GAS element driving a firefly luciferase reporter gene. PDGF β receptor activation initiates the STAT signaling cascade, which activates the GAS elements, resulting in increased expression of firefly luciferase. In addition, cells received an empty vector, RVY, or an RVY-based plasmid encoding either a small transmembrane protein from the library, the wild-type E5 protein, or the v-*sis* PDGF homolog, as well as a plasmid encoding either the wild-type PDGF β receptor or a PDGF β receptor mutant in which the transmembrane threonine 513 has been replaced with a leucine. This mutation blocks the interaction with the wild-type E5 protein.[28] Normalized firefly luciferase activity was measured two days after introduction of each combination of plasmids and was expressed as a percentage of activity induced by co-expression of the wild-type E5 gene and the wild-type PDGF β receptor.

As shown in Figure 10, there was low basal activity in cells co-transfected with RVY and a plasmid encoding either receptor. As expected, the wild-type E5 protein activated the wild-type receptor but not the T513L mutant, whereas v-*sis* activated both receptors. Similarly, the transformation-competent proteins LFC1-T1 and -T2 activated the wild-type receptor but not the T513L mutant receptor, indicating that they interacted with the PDGF β receptor *via* its transmembrane domain to initiate signaling. The non-transforming proteins LFC1-NT10 and -NT18 did not activate either receptor, confirming their lack of biological activity. Taken together, the results presented in this and the preceding section indicate that most, if not all, of the active small transmembrane proteins transformed cells by activating the PDGF β receptor *via* transmembrane interactions, whereas the non-transforming proteins did not interact productively with the PDGF β receptor.

Small transmembrane segments dimerize in membranes

The wild-type E5 protein must form a dimer to interact with the PDGF β receptor and cause cellular transformation. We used the TOXCAT assay to assess the ability of the transmembrane segments recovered from the library to dimerize in membranes.[32] This assay employs a chimeric fusion protein in which the transmembrane domain of interest is inserted between the N-terminal DNA binding domain of ToxR, a transcriptional activator that functions in an oligomerization-dependent fashion, and maltose-binding protein, which acts as a periplasmic anchor protein. The fusion protein is expressed in *Escherichia coli* together with a chloramphenicol acetyltransferase (CAT) reporter gene under the control of a ToxR-responsive *ctx* promoter. Oligomerization of the transmembrane domains within the bacterial inner membrane results in oligomerization of the linked ToxR domain, transcriptional activation of the *ctx* promoter, and CAT expression. We chose this approach to monitor

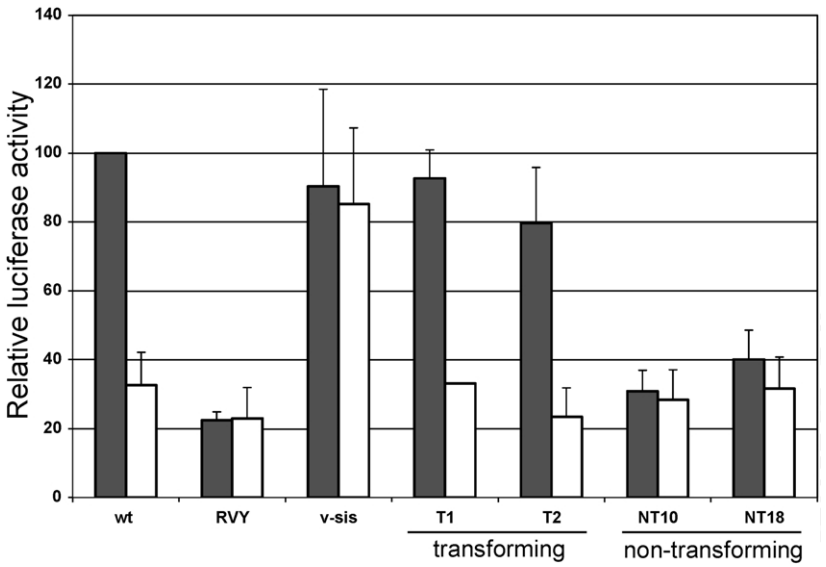

Figure 10. Activity of small transmembrane proteins requires the PDGF β receptor transmembrane domain. The graph shows the relative firefly luciferase activity induced by the wild-type E5 protein (wt), an empty vector control (RVY), a viral PDGF homolog (v-*sis*), and two transforming and non-transforming library proteins co-expressed with the wild-type PDGF β receptor (gray) or the T513L mutant receptor (white). The activity of each combination is expressed as a percentage of the activity induced by co-transfection of the wild-type E5 gene and the wild-type PDGF β receptor. For each experiment, three replicates of each transfection combination were performed. The average of two or three such experiments is shown, with error bars representing the standard deviation.

dimer formation because the level of CAT expression is proportional to the strength of the transmembrane interaction. In addition, this approach allowed us to assess the intrinsic ability of the transmembrane segments to oligomerize in the absence of disulfide bond formation.

The transmembrane domains of the wild-type E5 protein, a Q17L mutant which lacks transforming activity and displays impaired dimerization,[8] two transforming proteins recovered from the library (LFC1-T8 and LFC1-T14), and three non-transforming proteins (LFC1-NT10, LFC1-NT12, and LFC1-NT18) were cloned into the TOXCAT fusion construct. TOXCAT constructs containing transmembrane domains of wild-type dimerization-competent glycophorin A (GpA) and its dimerization-defective mutant, G83I, were used as controls, and the CAT activities for all constructs were normalized to the values obtained for GpA. The TOXCAT chimera containing the transmembrane domain of the wild-type E5 protein resulted in high CAT activity (Figure 11), consistent with published results,[33] indicating that the E5 transmembrane domain is capable of strong self-association in this system. Mutating Gln17 to leucine resulted in a 60% reduction in CAT activity, confirming the importance of Gln17 in E5 dimer formation. The chimeras containing the transmembrane domains of the two transforming and three non-transforming proteins all yielded similar CAT activity, comparable to that induced by the wild-type E5 transmembrane domain. These results indicate that the transmembrane domains of both the transforming and the non-transforming proteins are capable of forming strong homo-oligomers, presumably dimers, in membranes.

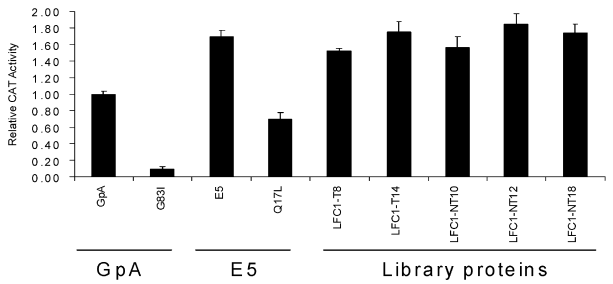

Figure 11. TOXCAT assay to determine the oligomerization capacity of the small transmembrane proteins in membranes. The relative CAT activity is shown for the wild-type E5 transmembrane domain as well as the transmembrane domain of the dimerization-impaired E5 Q17L mutant, the indicated transforming proteins (LFC1-T), and the indicated non-transforming proteins (LFC1-NT). As controls, the CAT activity was determined for the glycophorin A (GpA) protein and its dimerization-defective G83I mutant, and the CAT activities for all constructs were normalized to the values obtained for GpA. The average of four or more independent experiments is shown with error bars.

Structural models of transmembrane dimers

To determine whether the proteins recovered from the library were capable of forming dimers with structures similar to that proposed for the wild-type E5 protein, we generated structural models of the wild-type E5 protein and several of the recovered proteins. Structural calculations were carried out using the CNS searching of helix interactions (CHI) method from the crystallographic and NMR system (CNS).[34–36]

CHI simulations were performed on two parallel α-helices containing the transmembrane sequence of interest. The dimer models indicated that all of the proteins analyzed can form plausible left-handed coiled-coils. A model of the wild-type E5 protein, in good agreement with previously

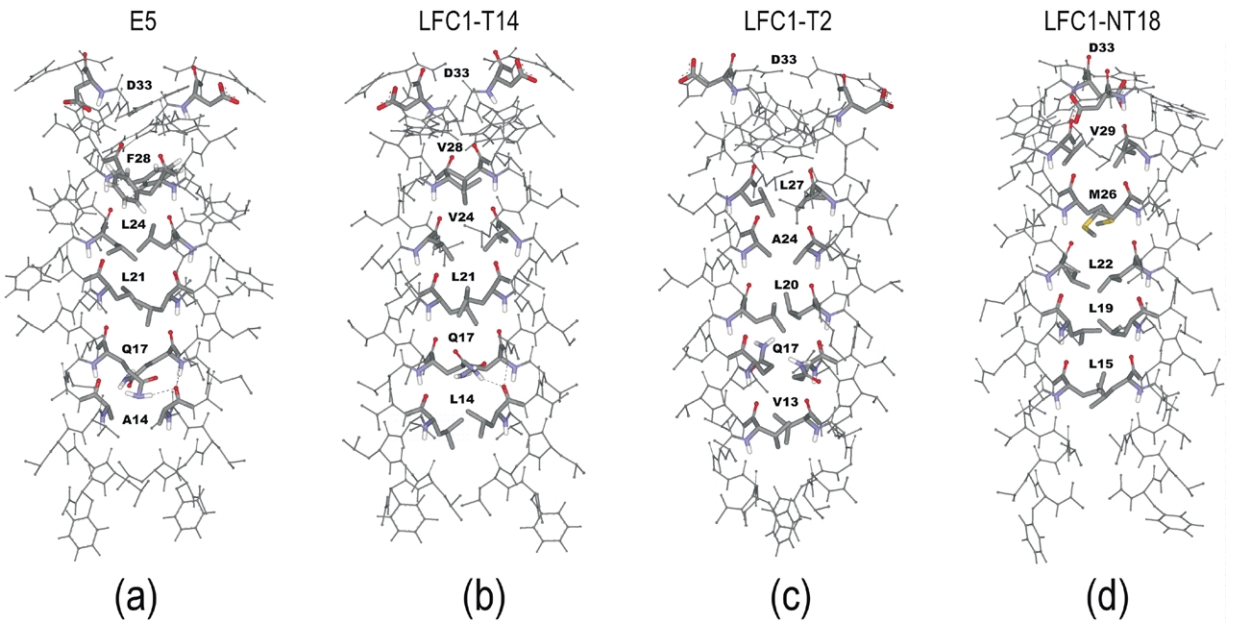

Figure 12. Structural models of transmembrane dimers. CHI structural models of protein dimers are shown for the wild-type E5 protein, the LFC1-T14 and LFC1-T2 transforming proteins, and the LFC1-NT18 non-transforming protein. The residues predicted to lie in the interface of the protein dimers are colored and labeled, as is Asp33, and likely interhelical hydrogen bonds are designated by dotted lines.

published results,[5] is shown together with models of the small transmembrane proteins in Figure 12. There are striking similarities between the models of the wild-type E5 protein (a), the small transforming proteins that closely conform to the consensus sequence, namely LFC1-T14 (b) and LFC1-T8 (not shown), and the active poly-leucine proteins (not shown). In all of these models, the dimer interface consisted of well-packed residues at positions 14, 21, 24 and 28 as well as the glutamine at position 17. Based on this analysis, all of the interfacial residues contributed to stabilization of the dimer, even though many of these residues in the small transmembrane proteins differed from the corresponding residues in the wild-type E5 protein. Figure 12 also shows that the active proteins contained likely interhelical hydrogen bonds involving glutamine 17, defined in this analysis whenever a hydrogen bond donor and acceptor were within 3 Å of one another. A model of an LFC1-T2 dimer was also generated (Figure 12(c)), because this transforming protein poorly matched the consensus sequence. Interestingly, the LFC1-T2 dimer had a different dimer interface, containing residues at positions 13, 20, 24 and 27, and yet still contained the critical glutamine 17 residue. Although glutamine 17 was present at the dimer interface in this model, it did not appear to participate in interhelical hydrogen bonding.

Models were also generated for three non-transforming proteins, LFC1-NT10 and -NT12 (not shown) and LFC1-NT18 (Figure 12(d)). The LFC1-NT10 and -NT12 proteins dimerized with an interface very similar to the wild-type E5 protein, while the interface of the LFC1-NT18 dimer was completely divergent, lacking even glutamine 17. Furthermore, neither LFC1-NT10 nor -NT18 displayed evidence of interhelical hydrogen bonds.

We have proposed that aspartic acid 33 of the E5 protein forms an essential salt-bridge with lysine 499 in the PDGF β receptor.[7] In the wild-type E5 protein and all of the transforming proteins modeled, the side-chains of both aspartic acid 33 residues pointed away from the dimer interface and were presumably accessible to interact with the lysine of the receptor (Figure 12(a)–(c)). In contrast, the dimer model of the non-transforming LFC1-NT18 protein shows aspartic acid 33 pointed toward the dimer interface, preventing any favorable interactions with the PDGF β receptor lysine 499 (Figure 12(d)). Without the crucial interactions between aspartic acid 33 and lysine 499, binding of the LFC1-NT18 protein to the PDGF β receptor may be greatly reduced, consistent with its inability to co-immunoprecipitate with the PDGF β receptor.

Discussion

Here, we selected and characterized novel, transformation-competent proteins from a library of small proteins with random transmembrane domains. The inability of 90% of the random clones to induce transformation indicates that there are important sequence constraints on activity. Nevertheless, it is striking that so many different sequences were compatible with transformation, even though fully one-third of the E5 protein was randomized. In fact, nearly any tested amino acid was tolerated at each of the 15 randomized positions within the transmembrane domain. Despite

the diversity of active transmembrane sequences, all tested transforming proteins bound and activated the PDGF β receptor *via* transmembrane interactions and required receptor activity for transformation and signaling. This sequence degeneracy suggests a remarkable flexibility and few absolute sequence requirements for recognition and activation of the PDGF β receptor.

Glutamine 17, aspartic acid 33, and the C-terminal cysteine residues were fixed in the library. Thus, the specific residues known to be required for E5 dimerization and PDGF β receptor binding were present in all of the library clones, biasing any potential interactions toward the PDGF β receptor. We have not ruled out the possibility that a subset of the small transmembrane proteins interacts with other cellular transmembrane proteins.

Statistical analysis revealed significant differences between the amino acid residues present at specific positions in the transforming compared to the non-transforming transmembrane segments. These positions fall on one side of a left-handed coiled-coil helical wheel, overlapping with the putative dimer interface, and they may form a second interface on the transmembrane dimer that is critical for an essential aspect of transformation, such as interaction with the PDGF β receptor. Our analysis also identified a consensus sequence that discriminates strongly between the transforming and non-transforming proteins. However, some of the small transmembrane proteins transformed poorly or not at all even though they conformed well to the consensus, implying that there are additional features of the transmembrane domain that modulate transforming activity. The position and identity of non-leucine residues in the consensus sequence were different from those of the wild-type E5 protein, as demanded by our omission of phenylalanine residues from the random segment. Nevertheless, a similar level of transforming activity was conferred on a poly-leucine protein by both the valine motif identified as the consensus sequence and the phenylalanine motif present in the wild-type E5 protein. Furthermore, we recovered rare transforming proteins that differed from both the wild-type E5 protein and the consensus sequence. Thus, multiple, different transmembrane sequence motifs are capable of inducing transformation *via* activation of the PDGF β receptor.

Despite their divergent sequences, the transmembrane domains of the random proteins oligomerize in cell membranes. Dimerization is presumably largely due to the presence of glutamine 17 in the dimer interface, because studies with model peptides show that polar amino acids such as glutamine are able to induce dimerization in hydrophobic environments, due to the formation of strong interhelical hydrogen bonds.[37–40] In addition, the C-terminal cysteine residues presumably favor disulfide bond-mediated dimer formation in mammalian cells and tend to align the transmembrane monomers in the same orientation as in the wild-type E5 dimer.

The CHI program was used to model the dimeric structure of the novel transmembrane segments. Two transforming proteins that were a good match to the consensus sequence (LFC1-T8 and LFC1-T14) as well as the active poly-leucine proteins were predicted to form dimers with the same interfacial positions as the wild-type E5 protein.[5,11] This interface allows glutamine 17 to form interhelical hydrogen bonds within the E5 dimer as well as to the PDGF β receptor, and it orients aspartic acid 33 away from the homodimer interface so that it is available to interact with lysine 499 of the PDGF β receptor. In contrast, the predicted dimer interface of the transformation-competent LFC1-T2 protein, which is a poor match to the consensus sequence, is shifted by one rotational position relative to the wild-type E5 interface. Although the active homodimers appear to adopt similar overall structures, different amino acids constitute the homodimer interface in the transforming proteins. A variety of amino acid sequences are also tolerated outside of the homodimer interface, and some of these amino acids must contact the PDGF β receptor. These results suggest that both the homodimer interface and the interface with the PDGF β receptor can accommodate diverse sequences, implying that there is considerable sequence flexibility in these interactions. In fact, the nature of the interaction between some of the transforming proteins and the PDGF β receptor may be very different from the interaction with the wild-type E5 protein. For example, a single monomer of the LFC1-T2 protein contributes both glutamine 17 and aspartic acid 33 to the same face of the dimer (A.M.D., unpublished results), whereas each of the two intertwined monomers of the wild-type E5 coiled-coil contributes one of these residues to each face of the dimer.[5]

In order to transform cells, the small transmembrane proteins must accumulate in cells, form stable homodimers, localize properly, and bind to the PDGF β receptor to induce receptor dimerization in the proper orientation. It has been shown for other receptor tyrosine kinases that the proper orientation of receptor monomers within the dimer is essential for *trans*-phosphorylation and initiation of the signaling cascade.[41,42] The transmembrane domains of the non-transforming small transmembrane proteins may lack specific residues that directly interact with the PDGF β receptor. Alternatively, some of the non-transforming proteins may not be expressed at a sufficient level or may not form stable homodimers in mammalian cells able to interact productively with the receptor.

In summary, we have developed a genetic method to identify small transmembrane proteins that bind and activate the PDGF β receptor, resulting in cell transformation. A surprisingly large fraction of transmembrane sequences are active in this assay. This approach may not only help to define the rules of engagement with the

transmembrane domain of the PDGF β receptor, but may also provide an approach to modulate the activity of a variety of transmembrane proteins. For example, variations of this method may be useful in constructing and identifying small transmembrane proteins that activate other receptor tyrosine kinases. Receptor tyrosine kinases are essential for growth and development, and aberrant activation of these kinases is associated with many human diseases, including cancers. Thus, a general strategy to modulate receptor activity and the oligomerization state of transmembrane proteins is of great interest.

Materials and Methods

Library construction

A PCR-based approach was used to construct a library of small transmembrane proteins. The upstream 5′ degenerate oligonucleotide had a fixed 5′ end derived from the E5 sequence, including a *Spe*I restriction enzyme site, followed by three NXR codons where N is an equal mix of A, T, C, and G, X is a 5 : 0.1 mixture of T : C, and R is an equal mix of A and G; a CAA codon encoding glutamine; 12 more NXR codons; and finally a fixed 3′end derived from the E5 sequence (Figure 1(a)). The downstream oligonucleotide was non-degenerate and contained a *Bam*HI restriction enzyme site at the 5′ end followed by the antisense sequence of the 3′ end of the wild-type E5 gene. The 3′ end of this primer was complementary to the 3′ end of the upstream primer. The two oligonucleotides were annealed and extended to create double-stranded products with a randomized central sequence. Two short primers that annealed to the extreme fixed ends of the double-stranded products were then used to amplify the library for 30 cycles. The PCR products were purified, digested with *Spe*I and *Bam*HI, and subcloned into the RVY retroviral vector to generate a library of genes in which 15 transmembrane codons of the E5 gene were replaced with a random mix of L : V : I : M : S : T : A : P codons in the following expected ratio 40 : 20 : 10 : 10 : 2 : 2 : 2 : 2. Cloning details are available from the authors upon request.

Cell lines and tissue culture

The retroviral packaging cell line, helper virus producer line, C127, and CV1 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum and antibiotics (DMEM-10). Retroviral stocks were prepared in GP2-293 cells as described.[43] To assay focus-forming activity, C127 cells were infected with retrovirus, and the number of foci was counted three weeks after infection. In parallel, infected cells were selected in medium containing 300 μg/ml of hygromycin B. Focus formation efficiency was calculated by normalizing the number of foci to the number of hygromycin B-resistant colonies arising in the same infection. Cell lines were established from pools of drug-resistant colonies.

Helper virus rescue of transforming library clones

Helper virus was harvested from NIH 3T3 TK-cells transfected with a pMLV-K plasmid construct to produce Moloney MLV (obtained from A. D. Miller, Fred Hutchinson Cancer Research Center).[44] Individual C127 foci were isolated using cloning cylinders, transferred into 24 well plates, and after 24 hours were infected with helper virus in medium with 4 μg/ml of Polybrene. The media was changed to DMEM-10 with 4 μg/ml of polybrene after four to five hours, and approximately 16 hours later the infected cells were passed into T25 flasks in 5 ml DMEM-10 with 4 μg/ml of polybrene. After four days, the supernatant was harvested, filtered through a 0.45 μm syringe filter, and used to infect naïve C127 cells as described.[8]

Recovery of transforming library clones

Hygromycin-resistant colonies that grew after infection of C127 cells with the rescued virus were pooled and expanded, and genomic DNA was harvested by using a DNeasy kit (Qiagen). A Hot-Start PCR protocol with rTthXL polymerase (Applied Biosystems) was used essentially as described[45] to amplify the retroviral insert. Briefly, the PCR mixture containing the polymerase, 1 μg of genomic DNA, and primers flanking the randomized segment was preheated to 80 °C for 60 seconds prior to addition of magnesium acetate at a final concentration of 1.0 mM. We used 75 cycles of amplification, rapidly cooling from the 93 °C denaturing step to 70 °C, before slowly cooling to the extension temperature of 60 °C. The cycling parameters and preheating reduced non-specific priming. In some cases, nested PCR was performed using primers internal to the primers used in the Hot-Start reaction. Detailed PCR protocols are available from the authors upon request. The PCR products were digested with *Spe*I and *Bam*HI and subcloned back into a retroviral vector to generate individual genes encoding the random transmembrane sequences. The sequence of the complete DNA insert was obtained for each clone.

PDGF β receptor inhibitor studies

C127 cells were plated at ~70% confluence in 12 well plates and grown in the presence or absence of 50 μM of AG1295 (Calbiochem) in DMEM-10. Cells were photographed after four days in the presence of the PDGF receptor inhibitor or in its absence.

Biochemical analysis

Cell lines at near confluence were serum-starved for 24 hours. Cell extracts were made using modified RIPA buffer containing protease and phosphatase inhibitors as described.[18] Immunoprecipitation and immunoblotting were carried out essentially as described.[8,18] For anti-phosphotyrosine blots, a 1 : 500 dilution of P-Tyr-100 (Cell Signaling Technology) in 5% (w/v) BSA in TBST was used.

Transient assay for PDGF receptor signaling

2×10^5 CV1 cells were plated into six well plates. Approximately 16 hours later, the cells were transfected with 10 ng of the pRL-SV40 *Renilla* luciferase plasmid, 750 ng of the pGAS-luciferase reporter plasmid (obtained from X. -Y. Fu, Yale University), 0.5 μg of an LXSN-based plasmid encoding the wild-type or mutant PDGF β receptor, and 0.5 μg of an RVY-based plasmid encoding the E5 protein, a small transmembrane protein from the

library, or v-*sis*. The DNA was added to 791 μl of Opti-MEM (Gibco) and 9 μl of 2 mg/ml lipofectamine (Invitrogen). The transfection mix was incubated at room temperature for 20 minutes before adding it to the cells that had been washed twice with 1 ml of PBS and once with 1 ml of Opti-MEM. After five hours, the medium was changed to 3 ml DMEM-10. After 24 hours, the cells were incubated in medium without serum supplementation in order to decrease background activation of the PDGF receptor, and the cells were harvested 24 hours later. A dual-luciferase assay was carried out using the dual-luciferase reporter assay system (Promega), according to the manufacturer's protocol. Samples were read using a Turner Designs TD-20/20 luminometer. The transfections were done in triplicate, and the results were normalized to the expression of the *Renilla* luciferase transfection control.

Computational analysis of E5-like sequences

The sequence logo[46] representation was created using WebLogo†.[47] At each position, the size of each letter is proportional to the frequency of the corresponding amino acid in that position. The total height of all the letters in the position is proportional to the conservation of the position.

To quantify the difference in amino acid distributions between the transforming and non-transforming clones, we calculated a χ^2 value for each position after grouping the amino acid residues into five categories: L (Leu), M (Met), V (Val), I (Ile), and O (Others).[48] We then performed a chi-square test with four degrees of freedom (one less than the number of amino acid categories), and we derived a difference score for each position, where the difference score is the -log (probability of the distributions being identical). Details are given in the Supplementary Material. The ranking of all positions was the same when the difference score was computed by using other methods based on information theory, such as the Jensen-Shannon divergence (data not shown). We also developed an automated procedure to identify an optimal consensus sequence. For each trial consensus sequence, a similarity score was computed for each clone by counting the number of matches between the consensus sequence and the clone. This resulted in two sets of similarity scores: $\{x_A\}$ for transforming clones and $\{x_B\}$ for non-transforming clones. The performance of the trial consensus sequence was then evaluated by the *t*-score, a measure of the separation of the means between these two sets of scores.[48] An iterated simulated annealing procedure was then used to find consensus sequences with optimal performance.[49] The details of the method are presented in the Supplementary Material.

Construction and characterization of TOXCAT chimera and CAT assays

Double-stranded oligonucleotides encoding the transmembrane domains (residues L8–H34) of the wild-type E5 protein, the E5 Q17L mutant, and the library proteins were ligated into the *Nhe*I and *Bam*HI sites in the pccKAN plasmid.[32] The resulting plasmids were expressed in *E. coli*, and equivalent expression levels of the chimeric proteins were verified by Western blotting. Membrane insertion of the chimeric proteins was verified using 0.1 M NaOH washes.[50] Whole cell, membrane, and soluble fractions were electrophoresed on SDS-PAGE, followed by Western analysis using antibodies against maltose binding protein (MBP) (New England Biolabs). The correct orientation of the TOXCAT constructs in the membrane was confirmed through protease sensitivity in a spheroplast assay, described elsewhere.[32] Acetylation of chloramphenicol by bacterial lysates was measured in an enzymatic assay by biochemical fractionation and scintillation counting (see Supplementary Material for details).

Computational searches using CHI

Details of the CHI computational search method have been described.[34,35,36] All structural calculations were carried out in vacuo on a Silicon Graphics O2 workstation. Using CHI, two canonical α-helices containing residues 8–34 of BPV E5 and its mutants were built. The starting geometries incorporated both right-handed (−25°) and left-handed (+25°) crossing angles and a distance between the two helices ranging from 10.5–11.0 Å. In a full search of all pairwise interactions, the two helices were independently rotated about their central axis in 45° increments from 0° to 360°. Molecular dynamics (MD) simulations were performed using simulated annealing of atomic coordinates at each pair of rotations. Four different MD simulations were performed for each starting geometry. Energy minimization of structures was carried out both before and after MD simulations, and a total of 512 structures were obtained from each search. Groups of ten or more final structures with a backbone RMSD of 1 Å or less were placed into clusters, followed by calculation of an average structure for each cluster and energy minimization.

† http://weblogo.berkeley.edu

Acknowledgements

We thank Joshua Drumm and Kristin Yates for assistance recovering and testing library clones; Inigo San Gil of the Yale core facility for bioinformatics for help with initial sequence analysis; Louis Alexander for assistance with PCR protocols; Dusty Miller and Xin-Yuan Fu for essential reagents; Dawn Mattoon for critical reading of the manuscript; and Jan Zulkeski for assistance in preparation of the manuscript. L.L.F.-C. was supported by an Anna Fuller postdoctoral fellowship and an NIH postdoctoral training grant (CA09159). A.M.D. was supported by a postdoctoral fellowship from the American Cancer Society (PFCSM-0301701). J.B.F. was supported by an NIH predoctoral training grant (GM07499). D.M.E., Y.X. and M.G. were supported by a grant from the NIH (P01 GM054160). This work was supported by a grant from the National Cancer Institute to D.D. (CA37157).

References

1. DiMaio, D., Lai, C. C. & Klein, O. (1998). Virocrine transformation: the intersection between viral

transforming proteins and cellular signal transduction pathways. *Annu. Rev. Microbiol.* **52**, 397–421.
2. DiMaio, D., Guralski, D. & Schiller, J. T. (1986). Translation of open reading frame E5 of bovine papillomavirus is required for its transforming activity. *Proc. Natl Acad. Sci. USA*, **83**, 1797–1801.
3. Schlegel, R., Wade-Glass, M., Rabson, M. S. & Yang, Y.-C. (1986). The E5 transforming gene of bovine papillomavirus encodes a small hydrophobic protein. *Science*, **233**, 464–467.
4. Burkhardt, A., Willingham, M., Gay, C., Jeang, K.-T. & Schlegel, R. (1989). The E5 oncoprotein of bovine papillomavirus is oriented asymmetrically in Golgi and plasma membranes. *Virology*, **170**, 334–339.
5. Surti, T., Klein, O., Aschheim, K., DiMaio, D. & Smith, S. O. (1998). Structural models of the bovine papillomavirus E5 protein. *Proteins: Struct. Funct. Genet.* **33**, 601–612.
6. Horwitz, B. H., Burkhardt, A. L., Schlegel, R. & DiMaio, D. (1988). 44-amino-acid E5 transforming protein of bovine papillomavirus requires a hydrophobic core and specific carboxyl-terminal amino acids. *Mol. Cell. Biol.* **8**, 4071–4078.
7. Klein, O., Kegler-Ebo, D., Su, J., Smith, S. & DiMaio, D. (1999). The bovine papillomavirus E5 protein requires a juxtamembrane negative charge for activation of the platelet-derived growth factor beta receptor and transformation of C127 cells. *J. Virol.* **73**, 3264–3272.
8. Klein, O., Polack, G. W., Surti, T., Kegler-Ebo, D., Smith, S. O. & DiMaio, D. (1998). Role of glutamine 17 of the bovine papillomavirus E5 protein in platelet-derived growth factor beta receptor activation and cell transformation. *J. Virol.* **72**, 8921–8932.
9. Meyer, A. N., Xu, Y.-F., Webster, M. K., Smith, A. S. & Donoghue, D. J. (1994). Cellular transformation by a transmembrane peptide: structural requirements for the bovine papillomavirus E5 oncoprotein. *Proc. Natl Acad. Sci. USA*, **91**, 4634–4638.
10. Sparkowski, J., Mense, M., Anders, M. & Schlegel, R. (1996). E5 oncoprotein transmembrane mutants dissociate fibroblast transforming activity from 16-kilodalton protein binding and platelet-derived growth factor receptor binding and phosphorylation. *J. Virol.* **70**, 2420–2430.
11. Mattoon, D., Gupta, K., Doyon, J., Loll, P. J. & DiMaio, D. (2001). Identification of the transmembrane dimer interface of the bovine papillomavirus E5 protein. *Oncogene*, **20**, 3824–3834.
12. Nilson, L. A., Gottlieb, R. L., Polack, G. W. & DiMaio, D. (1995). Mutational analysis of the interaction between the bovine papillomavirus E5 transforming protein and the endogenous beta receptor for platelet-derived growth factor in mouse C127 cells. *J. Virol.* **69**, 5869–5874.
13. Horwitz, B. H., Weinstat, D. L. & DiMaio, D. (1989). Transforming activity of a 16-amino-acid segment of the bovine papillomavirus E5 protein linked to random sequences of hydrophobic amino acids. *J. Virol.* **63**, 4515–4519.
14. Kulke, R., Horwitz, B. H., Zibello, T. & DiMaio, D. (1992). The central hydrophobic domain of the bovine papillomavirus E5 transforming protein can be functionally replaced by many hydrophobic amino acid sequences containing a glutamine. *J. Virol.* **66**, 505–511.
15. Cohen, B. D., Lowy, D. R. & Schiller, J. T. (1993). The conserved C terminal domain of the bovine papillomavirus E5 oncoprotein can associate with an alpha-adaptin-like molecule: a possible link between growth factor receptors and viral transformation. *Mol. Cell. Biol.* **13**, 6462–6468.
16. Goldstein, D. J., Andresson, T., Sparkowski, J. J. & Schlegel, R. (1992). The BPV-1 E5 protein, the 16 kDa membrane pore-forming protein and the PDGF receptor exist in a complex that is dependent on hydrophobic transmembrane interactions. *EMBO J.* **11**, 4851–4859.
17. Suprynowicz, F. A., Baege, A., Sunitha, I. & Schlegel, R. (2002). c-Src activation by the E5 oncoprotein enables transformation independently of PDGF receptor activation. *Oncogene*, **21**, 1695–1706.
18. Drummond-Barbosa, D. A., Vaillancourt, R. R., 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**, 2570–2581.
19. Goldstein, D. J., Li, W., Wang, L.-M., Heidaran, M. A., Aaronson, S. A., Shinn, R. *et al.* (1994). The bovine papillomavirus type 1 E5 transforming protein specifically binds and activates the β-type receptor for platelet-derived growth factor but not other tyrosine kinase-containing receptors to induce cellular transformation. *J. Virol.* **68**, 4432–4441.
20. Nilson, L. A. & DiMaio, D. (1993). Platelet-derived growth factor receptor can mediate tumorigenic transformation by the bovine papillomavirus E5 protein. *Mol. Cell. Biol.* **13**, 4137–4145.
21. Petti, L., Nilson, L. A. & DiMaio, D. (1991). Activation of the platelet-derived growth factor receptor by the bovine papillomavirus E5 transforming protein. *EMBO J.* **10**, 845–855.
22. Petti, L. & DiMaio, D. (1994). Specific interaction between the bovine papillomavirus E5 transforming protein and the beta receptor for platelet-derived growth factor in stably transformed and acutely transfected cells. *J. Virol.* **68**, 3582–3592.
23. 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**, 6736–6740.
24. Lai, C. C., 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**, 15241–15246.
25. Lai, C. C., Henningson, C. & DiMaio, D. (2000). Bovine papillomavirus E5 protein induces the formation of signal transduction complexes containing dimeric activated PDGF β receptor and associated signaling proteins. *J. Biol. Chem.* **275**, 9832–9840.
26. Staebler, A., Pierce, J. H., Brazinski, S., Heidaran, M. A., Li, W., Schlegel, R. & Goldstein, D. J. (1995). Mutational analysis of the β-type platelet-derived growth factor receptor defines the site of interaction with the bovine papillomavirus type 1 E5 transforming protein. *J. Virol.* **69**, 6507–6517.
27. Nappi, V. M., Schaefer, J. A. & Petti, L. M. (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**, 47149–47159.
28. Petti, L. M., Reddy, V., Smith, S. O. & 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**, 7318–7327.
29. Adduci, A. J. & Schlegel, R. (1999). The transmembrane domain of the E5 oncoprotein contains functionally discrete helical faces. *J. Biol. Chem.* **274**, 10249–10258.
30. Nappi, V. M. & Petti, L. M. (2002). Multiple transmembrane amino acid requirements suggest a highly specific interaction between the bovine papillomavirus E5 oncoprotein and the platelet-derived growth factor beta receptor. *J. Virol.* **76**, 7976–7986.
31. Kulke, R. & DiMaio, D. (1991). Biological properties of the deer papillomavirus E5 gene in mouse C127 cells: growth transformation, induction of DNA synthesis, and activation of the platelet-derived growth factor receptor. *J. Virol.* **65**, 4943–4949.
32. Russ, W. P. & Engelman, D. M. (1999). TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proc. Natl Acad. Sci. USA*, **96**, 863–868.
33. Gurezka, R., Laage, R., Brosig, B. & Langosch, D. (1999). A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *J. Biol. Chem.* **274**, 9265–9270.
34. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta. Crystallog. D. Biol. Crystallog.* **54**, 905–921.
35. Adams, P. D., Engelman, D. M. & Brunger, A. T. (1996). Improved prediction of the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins: Struct. Funct. Genet.* **26**, 257–261.
36. Adams, P. D., Arkin, I. T., Engelman, D. M. & Brunger, A. T. (1995). Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nature Struct. Biol.* **2**, 154–162.
37. Choma, C., Gratkowski, H., Lear, J. D. & DeGrado, W. F. (2000). Asparagine-mediated self-association of a model transmembrane helix. *Nature Struct. Biol.* **7**, 161–166.
38. Gratkowski, H., Lear, J. D. & DeGrado, W. F. (2001). Polar side chains drive the association of model transmembrane peptides. *Proc. Natl Acad. Sci. USA*, **98**, 880–885.
39. Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T. & Engelman, D. M. (2000). Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nature Struct. Biol.* **7**, 154–160.
40. Zhou, F. X., Merianos, H. J., Brunger, A. T. & Engelman, D. M. (2001). Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl Acad. Sci. USA*, **98**, 2250–2255.
41. Bell, C. A., Tynan, J. A., Hart, K. C., Meyer, A. N., Robertson, S. C. & Donoghue, D. J. (2000). Rotational coupling of the transmembrane and kinase domains of the Neu receptor tyrosine kinase. *Mol. Biol. Cell.* **11**, 3589–3599.
42. Burke, C. L., Lemmon, M. A., Coren, B. A., Engelman, D. M. & Stern, D. F. (1997). Dimerization of the p185neu transmembrane domain is necessary but not sufficient for transformation. *Oncogene*, **14**, 687–696.
43. DeFilippis, R. A., Goodwin, E. C., Wu, L. & DiMaio, D. (2003). Endogenous human papillomavirus E6 and E7 proteins differentially regulate proliferation, senescence, and apoptosis in HeLa cervical carcinoma cells. *J. Virol.* **77**, 1551–1563.
44. Miller, A. D. & Verma, I. M. (1984). Two base changes restore infectivity to a noninfectious molecular clone of Moloney murine leukemia virus (pMLV-1). *J. Virol.* **49**, 214–222.
45. Alexander, L., Weiskopf, E., Greenough, T. C., Gaddis, N. C., Auerbach, M. R., Malim, M. H. *et al.* (2000). Unusual polymorphisms in human immunodeficiency virus type 1 associated with nonprogressive infection. *J. Virol.* **74**, 4361–4376.
46. Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucl. Acids Res.* **18**, 6097–6100.
47. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* In the press.
48. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical recipes in C: the art of scientific computing*, 2nd edit., Cambridge University Press, Cambridge, UK.
49. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, **220**, 671–680.
50. Chen, H. & Kendall, D. A. (1995). Artificial transmembrane segments. Requirements for stop transfer and polypeptide orientation. *J. Biol. Chem.* **270**, 14115–14122.

Edited by G. von Heijne

(Received 6 February 2004; received in revised form 13 March 2004; accepted 13 March 2004)

Supplementary Material comprising analysis of clones, CAT assays and one Figure is available on Science Direct