Chapter 2

Transcription Activation with Polyamide-Polyproline-Peptide Conjugates

The text of this chapter was taken in part from a manuscript coauthored with Paramjit S. Arora, Aseem Z. Ansari, Professor Mark Ptashne (Sloan-Kettering Institute), and Professor Peter B. Dervan (Caltech).

(Arora, P.S.; Ansari, A.Z.; Best, T.P.; Ptashne, M.; Dervan, P.B. "Design of Artificial Transcriptional Activators with Rigid Poly-*L*-proline Linkers" *J. Am. Chem. Soc*. **2002**, *124*, 13067.)

Abstract

Typical eukaryotic transcriptional activators are composed of distinct functional domains, including a DNA-binding domain and an activating domain. Artificial transcription factors have been designed wherein the DNA-binding domain is a minor groove DNA-binding hairpin polyamide linked by a flexible tether to short activating peptides, typically 16-20 residues in size. In this study, the linker between the polyamide and the peptide was altered in an incremental fashion using rigid oligoproline "molecular rulers" in the 18-45 Å length range. We find that there is an optimal linker length which separates the DNA and the activation region for transcription activation.

Introduction

Transcriptional activators typically bind near a gene and recruit the transcriptional machinery to a nearby promoter, thereby stimulating the expression of the gene. These activators comprise two regions: the DNA binding domain and the activating domain. The former defines the promoter address in the genome where the activating region is to be delivered for recruitment of the transcriptional machinery.[1,2] We have previously reported efforts to replace the natural protein activators with nonnatural components smaller in size than nature's proteins. Artificial transcription activators comprised of three synthetic modules, a hairpin polyamide (PA) DNA-binding domain (DBD) and a short peptide activation domain (AD) (typically 16-20 amino acid residues in size) connected by flexible linkers which vary in length, have been shown to initiate transcription at targeted promoter sites in cell-free systems.[3,4] Modeling studies on these polyamide-peptide conjugates with flexible linkers suggested a distance of 20-40 Å between the DNA and the activating region for transcription activation. We sought to determine whether the length of the linker that projected the activating region away from the DNA would have any effect on the degree of activation (Figure 2.1). In the present study, we replaced the flexible linkers between the DBD and AD with rigid oligoprolines of varying incremental lengths (18-45 Å). Our results suggest that there is an optimal window within which activating regions of the type used here can function efficiently.

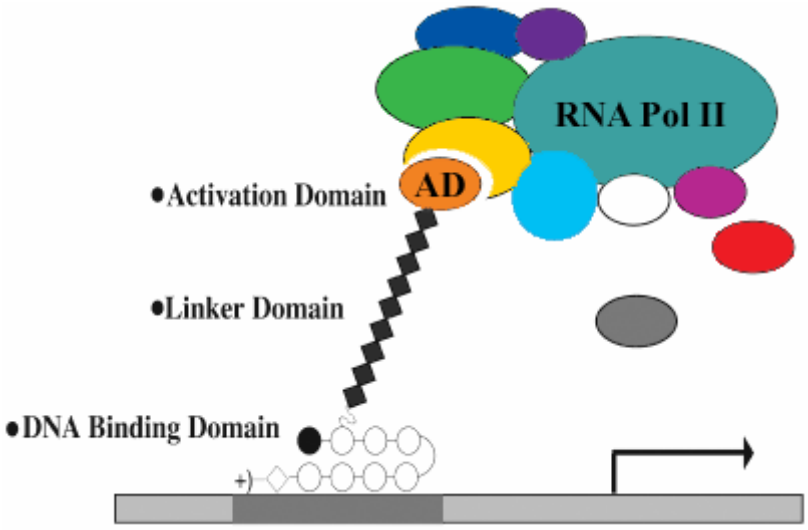
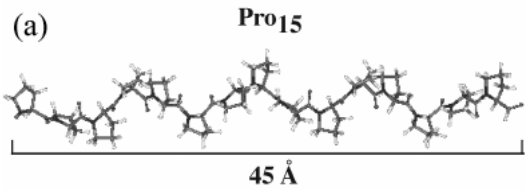


Figure 2.1 Activation of gene transcription by artificial transcription factors. The artificial activator is composed of three separate functional domains. The DNA binding domain consists of minor groove-binding pyrrole/imidazole polyamides. The DNA-binding domain is tethered to the activation domain (AD), a peptide, by a linker domain.

Results

In this study, we replaced the flexible linker between the activating peptide and the hairpin polyamides with 6, 9, 12, and 15 *L*-proline residues (Pro_6-Pro_{15}) (Figure 2.2). Poly-*L*-proline linkers were chosen as the "molecular rulers" for the present study since a stretch of proline residues forms a stable helical structure (the polyproline II helix). Addition of each proline residue increases the length of this helix in a predictable manner, approximately 3 Å per proline residue. Thus, the oligoproline linker projects the activating region peptide away from the DNA-binding polyamide in an incremental manner spanning 18, 27, 36, and 45 Å (Figure 2.2).[5] For the activating region we used two peptides, AH and VP2, that have been previously shown to function efficiently when tethered to a hairpin polyamide (Figure 2.3a).[4] Therefore, three series were synthesized:



(b)

Pro_n	predicted polyproline helix size (Å)
6	18
9	27
12	36
15	45

Figure 2.2 Poly-*L*-proline-based molecular rulers. (a) Structure of a poly-*L*-proline helix. (b) Predicted length of polyproline helices containing the indicated numbers of proline residues.

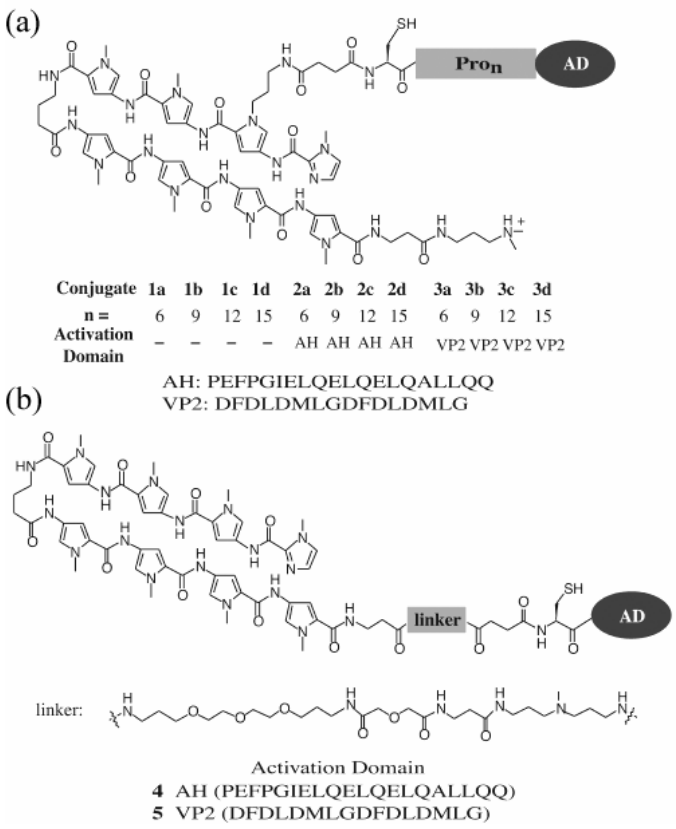


Figure 2.3 Structures of polyamide-activation peptide conjugates **1**-**5** prepared to explore the effect of linker length and flexibility on activation potential. (a) Polyamide-Pro_n-peptide conjugates for which the linker domain consisting of poly-*L*-proline helices serves as a molecular ruler. (b) Polyamide-peptide conjugates bearing a flexible linker.

PAPro$_{6\text{-}15}$-AH (**2a-d**), PA-Pro$_{6\text{-}15}$-VP2 (**3a-d**), and PA-Pro$_{6\text{-}15}$ (**1a-d**) lacking the activation peptides as controls. Conjugates with flexible poly(ethylene glycol) linkers, **4** and **5**, previously shown to activate transcription in vitro were included for comparison (Figure 2.3b).[4]

The eight-ring hairpin polyamides which target 5'-WGWWWW-3' (W = A or T) and two peptides, AH and VP2, were synthesized by solid-phase methods.[6] Each activating peptide contained an N-terminal cysteine for subsequent reaction with a thioester, **6**, via the native ligation reaction to afford the desired conjugate (Figure 2.4).[7] The ability of these conjugates, **1a-d**, **2a-d**, and **3a-d**, to activate transcription *in vitro* was tested using yeast nuclear extract on a template bearing three palindromic binding sites upstream of the minimal core promoter driving the expression of a transcript that

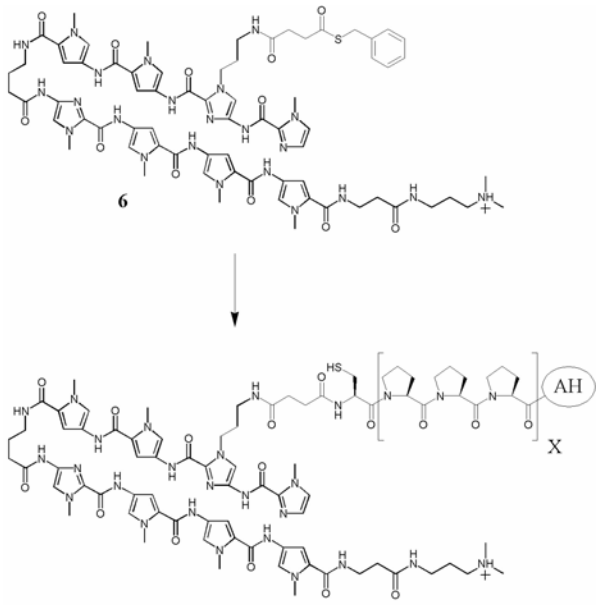


Figure 2.4 Synthesis of polyamide-peptide conjugates. Thiolane **6** was converted to the targeted conjugates via the native ligation reaction.

lacks guanine residues.[3,4] The results show that the linker plays a role in determining the ability of the activating region to stimulate *in vitro* transcription (Figure 2.5). This effect is not a function of the activating peptide, as both activating peptides tested showed similar profiles. Figure 2.5 also compares the two activating peptides attached to the polyamide via different proline linkers to conjugates **4** and **5** bearing the same activating peptides via flexible linkers.

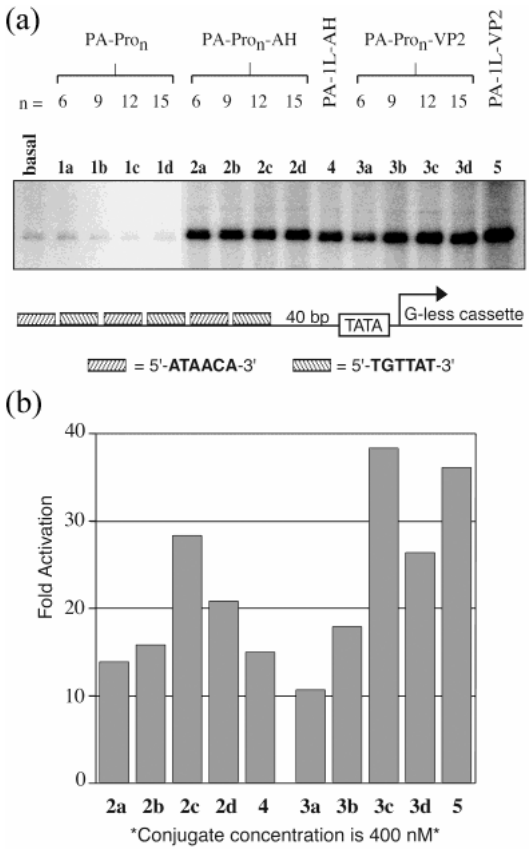


Figure 2.5 *In vitro* transcription reactions with the compounds listed in Figure 2.3; the transcription reaction conditions are described in the Experimental Section. (a) Storage phosphor autoradiogram of the reactions which were performed with a 400 nM concentration of each conjugate. The template configuration showing the sequence and number of the polyamide binding sites is depicted below the gel. (b) Fold activation was determined by comparing the amount of transcription elicited by conjugates **2a-d**, **3a-d**, **4**, and **5** with that of the basal level.

Our results indicate that the strength of the given activating peptide increases with every increment in the spacer length up to 12 residues (36 Å). An increase in the spacer length to 15 residues leads to a decrease in activation. These results suggest an optimal spacing of 36-45 Å between the DNA and the activating regions for efficient transcription. Figure 2.6 shows the relative potency of each compound in comparison to PA-Pro$_6$-AD (**2a**, AD = AH; **3a**, AD = VP2). Although the absolute fold activation mediated by all conjugates varies slightly between different experiments, the overall relationship between linker length and activation potential remains constant over four independent *in vitro* transcription reactions.

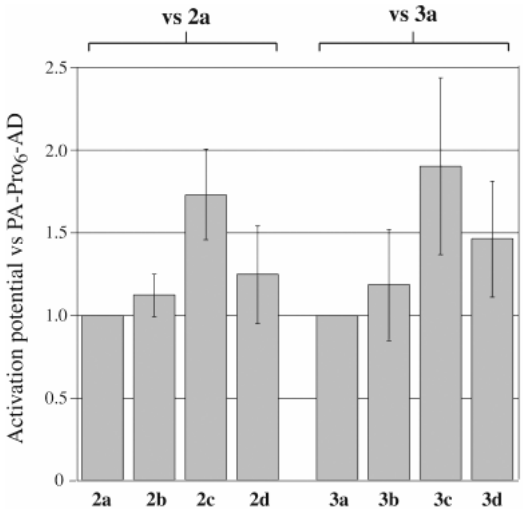


Figure 2.6 Summary of four independent *in vitro* transcription reactions showing the relative potency of each compound in comparison to PA-Pro$_6$-AD (**2a**, AD = AH; **3a**, AD = VP2).

To determine whether any of the effects observed were due to inefficient or inappropriate binding of the conjugate to its site on the promoter, we measured dissociation constants for conjugates **3a-d** using quantitative DNase I footprinting titrations.[8,9] These studies revealed that conjugates bound with similar affinities and specificities to their sites on the promoter DNA (Figure 2.7 and Table 2.1). In control

experiments polyamide-proline conjugates **1a-d** lacking the activation transcription peptides did not activate.

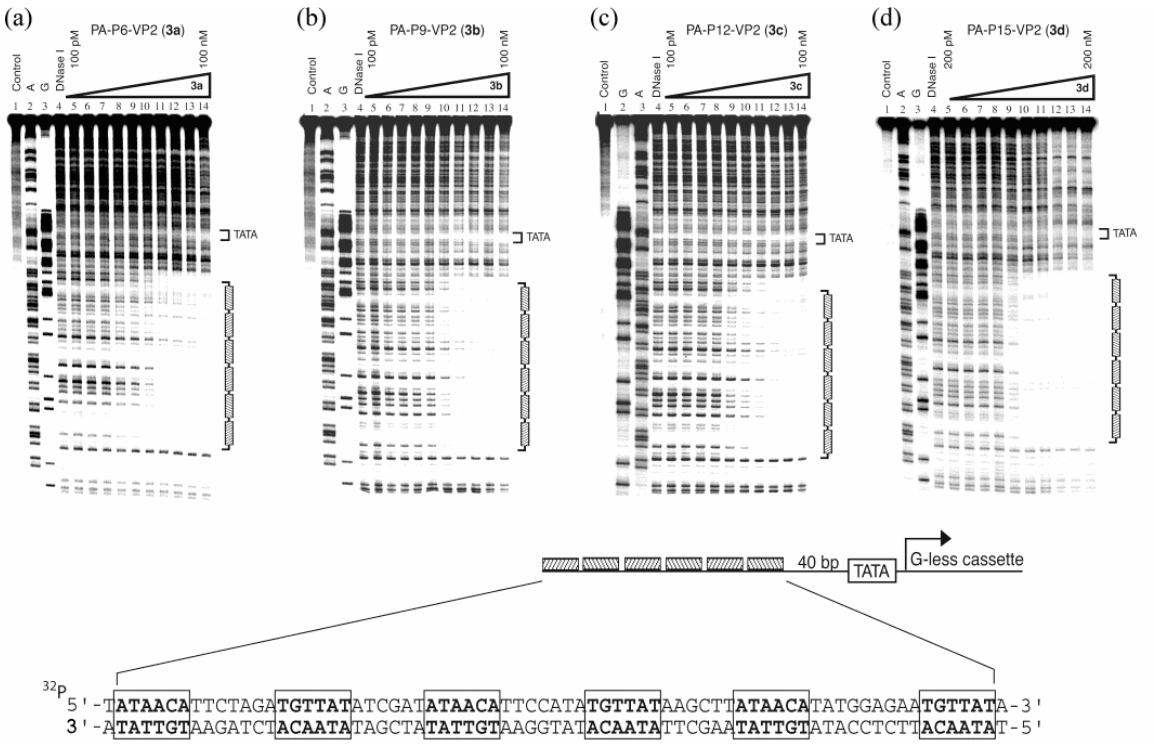


Figure 2.7 Quantitative DNase I footprinting titration of conjugates **3a-d** shows that all conjugates bound to their target sites with similar affinities. (Top) Storage phosphor autoradiogram of a quantitative DNase I footprinting titration of **3a-d** on a 63 bp 5'-^{32}P-labeled PCR fragment containing both the promoter region and 140 bp of the G-less cassette reporter. (a) Lane 1, undigested DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I standard; lanes 5-14, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM **3a**, respectively. (b) Lane 1, undigested DNA; lane 2, A reaction; lane 3, G reaction, lane 4, DNase I standard; lanes 5-14, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM **3b**, respectively. (c) Lane 1, undigested DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-14, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM **3c**, respectively. (d) Lane 1, undigested DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-14, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, and 200 nM **3d**, respectively. (Bottom) Design of the plasmid DNA template used for the DNase I footprinting titration experiments. The promoter region contains six cognate binding sites for the hairpin polyamide upstream of a G-less cassette reporter.

Table 2.1 Equilibrium Dissociation Constants (nM)

conjugate	K_D (nM)	conjugate	K_D (nM)
PA-Pro_6-VP2 (**3a**)	4.8 (±1.4)	PA-Pro_{15}-VP2 (**3d**)	2.3 (±1.3)
PA-Pro_9-VP2 (**3b**)	2.0 (±1.3)	PA-Pro_{12}-AH (**2c**)	7.7 (±1.7)
PA-Pro_{12}-VP2 (**3c**)	3.3 (±1.1)		

The use of poly-*L*-proline linkers as molecular rulers was inspired by the seminal work of Stryer and co-workers, who utilized fluorescence resonance energy transfer (FRET) analysis to show that oligomers composed of up to 12 proline residues can retain a rigid helical structure.[5] To ascertain that an oligomer composed of 15 proline residues does not deviate from the predicted structure, we repeated the FRET analysis on poly-*L*-proline linkers used in our experiments. To perform this analysis, we synthesized poly-*L*-proline linkers bearing Oregon Green (OG) energy donor and tetramethylrhodamine (TAMRA) energy acceptor moieties on each end (conjugates **7a**-**d**, Figure 2.8). In the Stryer study, naphthyl and dansyl groups were used as the energy donor and the energy acceptor moieties, respectively. The different dyes were used in the present study because the Förster radius (the distance at which energy transfer is 50% efficient) for the dansyl-naphthyl pair (R_o = 27 Å) may be too low to allow for efficient FRET when the energy donor and the acceptor are spaced by the Pro_{15} linker (45 Å). The Förster radius (R_o) for the OG-TAMRA pair is predicted to be 55 Å.[10] FRET measurements indicate that the oligomer composed of 15 proline units does retain a rigid structure. The plot of distance versus FRET efficiency provides a linear relationship between the conjugates, and importantly, the experimentally observed FRET efficiency is in good agreement with the predicted r^{-6} dependence (Figure 2.9).[5] Thus, the FRET studies show that the

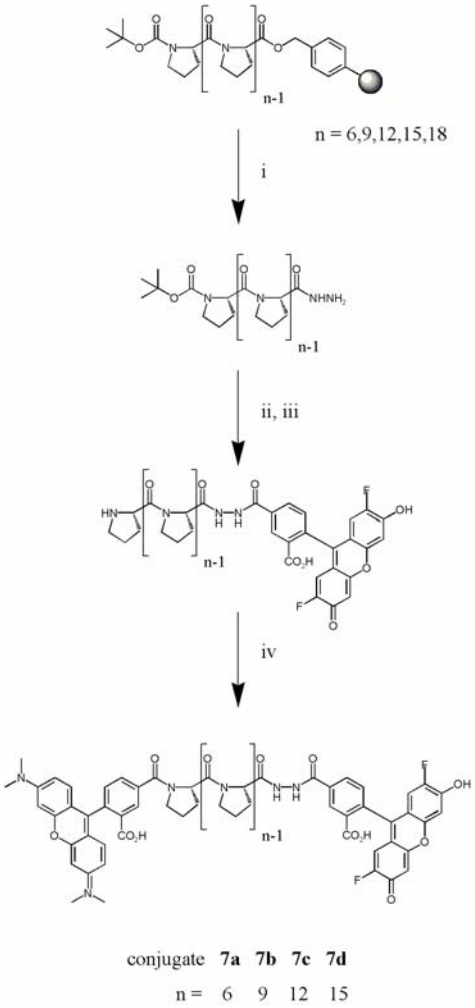


Figure 2.8 Synthesis of polyproline-dye conjugates. (i) 20% hydrazine in ethanol. (ii) Oregon Green-NHS ester, DMF, and DIEA. (iii) 20% TFA in CH_2Cl_2. (iv) TAMRA-NHS ester, DMF, and DIEA.

dependence of transcription activation on the number of *L*-prolines is not due to a change in the linker structure.

Discussion

In this study we examine the role of the distance separating the DBD and AD in our artificial transcription factors. This study was prompted by previous observations in

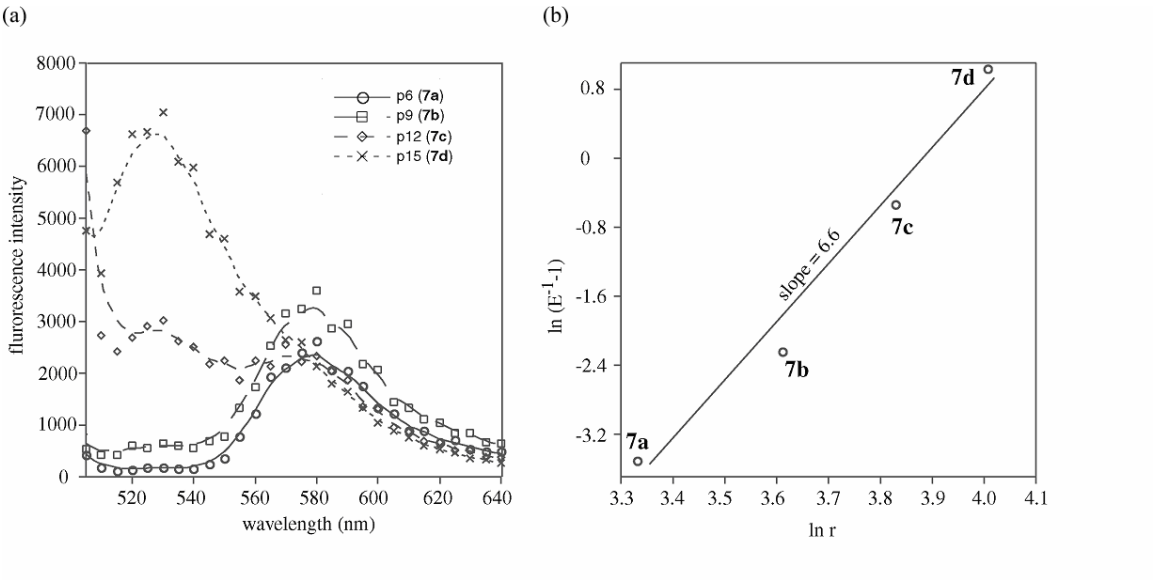


Figure 2.9 FRET data for polyproline helices. (a) The emission spectra of OG-Pro_n-TAMRA conjugates. The conjugates were excited at 495 nm. (b) The dependence of the efficiency of energy transfer on distance is given in this plot of ln (E^{-1} – 1) vs. ln r. The slope is 6.6, in good agreement with the predicted r^{-6} dependence.

which we found that activating regions activated transcription to different degrees on the basis of the point of attachment as well as the nature of the linker separating them from the DNA binding polyamide.[4] This dependence on spatial presentation led us to test the role of linker length in the degree of activation elicited by two acidic activating peptides. The proline linker domain is conjugated to an internal pyrrole of the polyamide, rather than to the C-terminus, which allows the linkers to project out of the minor groove away from the DNA helix. Our data suggest that for the two acidic activating regions (AH and VP2) the strength of activation increases as the linker reaches a length of 36 Å. The increase, while modest, is reproducible and displays a clear trend (Figures 2.5 and 2.6). The effects are muted in part due to three features that were incorporated into the experimental design: the use of multiple DNA binding sites to elicit robust activation, the presence of a flexible hinge tethering the proline linker to the polyamide, and the

unstructured nature of the activating peptides, which themselves may sample a significant amount of solvent space when fully extended.

We did not find a significant contribution of poly-*L*-proline linkers themselves to the level of activation in the absence of tethered activating regions. It has been reported that proline-rich (rather than poly-*L*-proline) activating domains can function to activate transcription even in yeast extracts, though this activation is not as robust as that elicited by acidic activators.[11] Presumably the proline oligomer is not a sufficiently strong activator; its effects are therefore not detected in our studies. An alternative possibility is that the ability of proline-rich activating domains to elicit transcription may be more sensitive to their spatial location.

As the principle of recruitment implies, for a DNA-tethered activator to function efficiently, it must sample sufficient nuclear solvent space to bind and recruit various complexes that participate in transcriptional initiation to a given promoter. However, beyond a certain distance the activating region would not function to recruit, as the local concentration of the machinery at a given promoter would not be tremendously enhanced by binding a distant activating region. It has been shown that eukaryotic activating regions that are not tethered to DNA—even when presented at exceedingly high concentrations such that it would bind to its targets in the transcriptional machinery—do not improve the level of transcription from a given promoter. In fact, as expected, it "squelches" the ability of a DNA-bound activator from functioning—presumably by binding the targets in the machinery.[2,3,12] Thus, for efficient recruitment in the context of our *in vitro* studies with artificial activators on nonchromatinized templates using yeast

extracts, we find that a spacing of 36-45 Å serves as an optimal distance between the DNA and the activating regions to elicit transcriptional activation.

This work is a step forward toward the goal of engineering and integrating at the molecular level the components for functional artificial transcription factors. The field is at an early stage, and one can imagine other components replacing the design reported here, such as triple-helix-forming oligonucleotides or PNAs tethered to activating peptides.[13,14] Long-term goals for the field would be the replacement of the activating peptide with nonpeptide constructs and the activation (or repression) of endogenous genes in cell culture experiments.

Experimental Section

Synthesis of Polyamide-Peptide Conjugates 1-3

Polyamide thioester **6** was transformed into conjugates **1a-d**, **2a-d**, and **3a-d** by previously reported methods.[7] The identities of all conjugates were verified by MALDI-TOF mass spectrometry. Characterization: **1a** (PA-Pro_6), MALDI-TOF $[M+H]^+$ (monoisotopic mass) calcd 2050.6, obsd 2050.0; **1b** (PA-Pro_9), MALDI-TOF $[M+H]^+$ (average mass) calcd 2341.9, obsd 2341.4; **1c** (PA-Pro_{12}), MALDI-TOF $[M+H]^+$ (average mass) calcd 2633.3, obsd 2633.1; **1d** (PA-Pro_{15}), MALDI-TOF $[M+H]^+$ (average mass) calcd 2924.6, obsd 2925.1; **2a** (PAPro_6-AH), MALDI-TOF $[M+H]^+$ (average mass) calcd 4354.5, obsd 4354.8; **2b** (PA-Pro_9-AH), MALDI-TOF $[M+H]^+$ (average mass) calcd 4646.8, obsd 4647.5; **2c** (PA-Pro_{12}-AH), MALDI-TOF $[M-H]^-$ (average mass) calcd 4937.5, obsd 4937.5; **2d** (PA-Pro_{15}-AH), MALDI-TOF $[M+H]^+$ (average mass) calcd 5228.7, obsd 5229.2; **3a** (PA-Pro_6-VP2), MALDI-TOF $[M+H]^+$ (average mass) calcd

3862.3, obsd 3862.7; **3b** (PA-Pro_9-VP2), MALDI-TOF $[M+H]^+$ (average mass) calcd 4153.7, obsd 4154.8; **3c** (PA-Pro_{12}-VP2), MALDI-TOF $[M-H]^{+-}$ (average mass) calcd 4445.0, obsd 4445.3; **3d** (PA-Pro_{15}-VP2), MALDI-TOF $[M+H]^+$ (average mass) calcd 4736.3, obsd 4736.8.

Synthesis of Poly-*L*-proline -Dye Conjugates 7a-d

Poly-*L*-proline peptides of lengths *n* proline residues, where *n* = 6, 9, 12, 15, were individually synthesized by standard *t*-Boc solid phase methods, retaining the n-terminal *t*-Boc protecting group. All peptides were cleaved from resin by treatment with 20% (v/v) anhydrous hydrazine in absolute ethanol for 18 hrs at room temperature. Peptides were purified by reverse-phase HPLC, and lyophilized to provide white powders. Purified peptides (2.5 μmoles) were treated with the 5-*N*-hydrosuccinimidyl ester of Oregon Green 488 (OG, Molecular Probes) (5.0 μmoles), DMF (100 μL), and *N*,*N*-diisopropylethylamine (20 μL), and purified by reverse-phase HPLC to yield C-terminal dye intermediates. The N-terminal *t*-Boc group was removed by treatment with 20% TFA in CH_2Cl_2 and the resulting molecules were treated with the 5-*N*-hydrosuccinimidyl ester of tetramethylrhodamine (TAMRA, Molecular Probes) (5.0 μmoles), DMF (100 μL) and *N*,*N*-diisopropylethylamine (20 μL). Purification by reverse-phase HPLC afforded dye conjugates **7a**-**e** (typical isolated yields: 20-40%). The identity of all conjugates was verified by MALDI-TOF mass spectrometry. Characterization: **7a** (TAMRA-Pro_6-OG): MALDI-TOF $[M+H]^+$ (monoisotopic mass) calcd 1422.5, obsd 1422.6; **7b** (TAMRA-Pro_9-OG): MALDI-TOF $[M+H]^+$ (monoisotopic mass) calcd 1712.7, obsd 1712.5; **7c** (TAMRA-Pro_{12}-OG): MALDI-TOF $[M+H]^+$ (monoisotopic

mass) calcd 2003.8, obsd 2003.9; **7d** (TAMRA-Pro_{15}-OG): MALDI-TOF $[M+H]^+$ (monoisotopic mass) calcd 2295.0, obsd 2295.2.

Fluorescence Resonance Energy Transfer Experiments

Fluorescence emission spectra were collected on an ISS-K2 fluorometer. Solutions (25 nM) of conjugates **7a-d**, Oregon Green, and TAMRA were prepared in 1:1 methanol:bicarbonate buffer (150 mM, pH 8.3). The emission spectra of these solutions were collected from 500-640 nm, with an excitation wavelength of 495 nm.

DNase I Footprinting Titration Experiments

A 363 bp 5' ^{32}P-labeled PCR fragment was generated from template plasmid pAZA812 in accordance with standard protocols and isolated by nondenaturing gel electrophoresis. All DNase I footprinting reactions were carried out in a volume of 400 μL. A polyamide stock solution or water (for reference lanes) was added to TKMC buffer, with final concentrations of 50 mM Tris-HCl, 50 mM KCl, 50 mM $MgCl_2$, and 25 mM $CaCl_2$, pH 7.0, and 15 kcpm 5'-radiolabeled DNA. The solutions were equilibrated for 12-18 h at 22°C. Cleavage was initiated by the addition of 10 μL of a DNase I stock solution and was allowed to proceed for 7 min at 22°C. The reactions were stopped by adding 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 μM base pair calf thymus DNA and then ethanolprecipitated. The cleavage products were resuspended in 100 mM Trisborate-EDTA/80% formamide loading buffer, denatured at 85°C for 10 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 2 h and 15 min.

The gels were dried under vacuum at 80 °C and quantitated using storage phosphor technology.

In Vitro Transcription Assays

Template plasmid pAZA812 was constructed by cloning a 78 bp oligomer bearing three cognate palindromic sequences for conjugates **1a-d**, **2a-d**, and **3a-d** into a *Bgl*2 site 30 bp upstream of the TATA box of pMLΔ53. This plasmid has the AdML TATA box 30 bp upstream of a 277 bp G-less cassette. For each reaction, 20 ng of plasmid (30 fmol of palindromic sites) was preincubated with a 400 nM concentration of the compound for 75 min prior to the addition of 90 ng of yeast nuclear extract in a 25 μL reaction volume. The reactions were performed as previously described and resolved on 8% 30:1 polyacrylamide gels containing 8 M urea. The gels were dried and exposed to photostimulatable phosphorimaging plates (Fuji Photo Film Co.). The data were visualized using a Fuji phosphorimager followed by quantitation using MacBAS software (Fuji Photo Film Co.).

Acknowledgement

We are grateful to the National Institutes of Health for research support, the American Cancer Society for a postdoctoral fellowship to P.S.A., and the Helen Hay Whitney Foundation for a fellowship to A.Z.A.

References

1) Ptashne, M.; Gann, A. *Nature* **1997**, *386*, 569.

2) Ptashne, M.; Gann, A. *Genes and Signals*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

3) Mapp, A.K.; Ansari, A.Z.; Ptashne, M.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3930.

4) Ansari, A.Z.; Mapp, A.K.; Nguyen, D.; Dervan, P.B.; Ptashne, M. *Chem. Biol*. **2001**, *8*, 583.

5) Stryer, L.; Haugland, R.P.; *Biochemistry* **1967**, *6*, 719.

6) Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc*. **1996**, *118*, 6141.

7) Mapp, A.K.; Dervan, P.B. *Tetrahedron Lett*. **2000**, *41*, 9451.

8) Senear, D.F.; Brenowitz, M.; Shea, M.A.; Ackers, G.K. *Biochemistry* **1986**, *25*, 7344.

9) Brenowitz, M; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol*. **1986**, *130*, 132.

10) Haugland, R.P. *Handbook of Fluorescent Probes and Research Chemicals*; 6th ed.; Molecular Probes Inc.: Eugene, OR, 1996.

11) Kim, T.K.; Roeder, R.G. *J. Biol. Chem*. **1993**, *268*, 20866.

12) Gill, G.; Ptashne, M. *Nature* **1988**, *334*, 721.

13) (a) Kuznetsova, S.; Ait-Si-Ali, S.; Nagibneva, I.; Troalan, F.; Le Villain, J.-P.; Harel-Bellan, A.; Svinarchuk, F. *Nucleic Acids Res*. **1999**, *27*, 3995. (b) Stanojevic, D.; Young, R.A. *Biochemistry* **2002**, *41*, 7209.

14) (a) Denison, C.; Kodadek, T. *Chem. Biol*. **1998**, *5*, 129. (b) Liu, B.; Han, Y.; Corey, D.R.; Kodadek, T. *J. Am. Chem. Soc*. **2002**, *124*, 1838.