

# Activation of gene expression by small molecule transcription factors

Anna K. Mapp\*<sup>†</sup>, Aseem Z. Ansari\*<sup>†</sup>, Mark Ptashne\*<sup>§</sup>, and Peter B. Dervan\*<sup>§</sup>

\*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; and <sup>†</sup>Molecular Biology Program, Sloan-Kettering Institute, New York, NY 10021

Contributed by Peter B. Dervan, February 15, 2000

**Eukaryotic transcriptional activators are minimally comprised of a DNA binding domain and a separable activation domain; most activator proteins also bear a dimerization module. We have replaced these protein modules with synthetic counterparts to create artificial transcription factors. One of these, at 4.2 kDa, mediates high levels of DNA site-specific transcriptional activation *in vitro*. This molecule contains a sequence-specific DNA binding polyamide in place of the typical DNA binding region and a nonprotein linker in place of the usual dimerization peptide. Thus our activating region, a designed peptide, functions outside of the archetypal protein context, as long as it is tethered to DNA. Because synthetic polyamides can, in principle, be designed to recognize any specific sequence, these results represent a key step toward the design of small molecules that can up-regulate any specified gene.**

According to our current picture, an activator binds to its cognate sites in the genome and recruits the transcriptional machinery to nearby promoters; initiation of transcription then follows (1). The yeast activator Gal4, like many eukaryotic activators, comprises three separable modules: DNA binding, dimerization, and activation (1–3). Each of these modules can be “swapped” with corresponding modules from other activator proteins (1, 2, 4, 5). Additional regulatory nuances have in some cases been superimposed on these three essential components (6). Many different peptides, including the 15-residue designed peptide called AH (amphipathic helix), function as activating regions when tethered to DNA (7–9). In all cases reported thus far, however, natural DNA binding and dimerization domains have been used to tether activating regions (natural or designed) to DNA (7–12). It has, therefore, been difficult to exclude the possibility that these natural components play some essential role in gene activation other than to bring the activating region to specific sites on DNA.

We sought to develop a general motif for artificial transcription factors that would incorporate synthetic counterparts of all functional modules of naturally occurring activator proteins and would furthermore be more versatile in their ability to up-regulate designated genes. In the design of our synthetic activator, the protein-DNA binding module was replaced with a hairpin polyamide composed of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids that binds in the minor groove of DNA (Fig. 1A). Hairpin polyamides are capable of targeting predetermined DNA sequences with affinities and specificities comparable to DNA binding proteins in accordance with a simple set of pairing rules dictated by the side-by-side binding of the aromatic amino acids (13–15). These synthetic DNA binding ligands are cell permeable, and one such compound was shown to specifically interfere with gene expression in mammalian cell culture (16, 17). The hairpin polyamide selected for the present study, ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (where  $\gamma$  is  $\gamma$ -aminobutyric acid,  $\beta$  is  $\beta$ -alanine, and Dp is dimethylaminopropylamide), binds the sequence 5'-TGTTAT-3' with a dissociation constant ( $K_D$ )

of 1.1 nM. In our initial experiments, we targeted a palindromic binding site containing this sequence as an inverted repeat separated by 7 bp. We used as a dimerization element a sequence that is known to form a coiled-coil, residues 251–281 of the yeast protein Gcn4 (18). The activation domain used in all of our experiments was AH (PEFPGIELQELQELQALLQQ) (7).

## Methods

**Synthesis of Conjugates 3–5.** Polyamide 1 was prepared according to established protocols (19) and then was combined with 1.2 equivalent (eq) thiolane-2,5-dione (20) and 3 eq *i*-Pr<sub>2</sub>EtN in 1-methyl-2-pyrrolidinone at a final concentration of 10 mM. After 15 min, 1.5 vol of 100 mM NaOAc (pH 3.2) were added followed by 3 eq of benzyl bromide. After an additional 15 min, the reaction mixture was subjected to purification by reversed-phase HPLC, and the appropriate fractions were concentrated to isolate 2 (53%) as a white powder. Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) MS analysis of 2: [M + H] calculated (monoisotopic) 1470.7, observed 1470.7. Polyamide 2 (1  $\mu$ mol) and peptide (0.8–1  $\mu$ mol) were combined in 5% 1-methyl-2-pyrrolidinone in 6 M Gn•HCl, 100 mM potassium phosphate buffer (pH 7.3), and 10% (vol/vol) thiophenol was added to this solution. Reaction progress was monitored by analytical HPLC and upon completion (3–5 d), purification of the mixture by reversed-phase HPLC resulted in isolation of the desired hairpin polyamide-peptide conjugates. Yields and characterization: 3 (PA-Gcn4-AH): 11%, MALDI-TOF [M + H] (average mass) calculated 7465.4, observed 7465.4; 4 (PA-Gcn4): 21%, MALDI-TOF: [M + H] (average mass) calculated 5159.8, observed 5159.9; 5 (PA-AH): 22%, MALDI-TOF: [M + H] (average mass) calculated 3774.2, observed 3774.5.

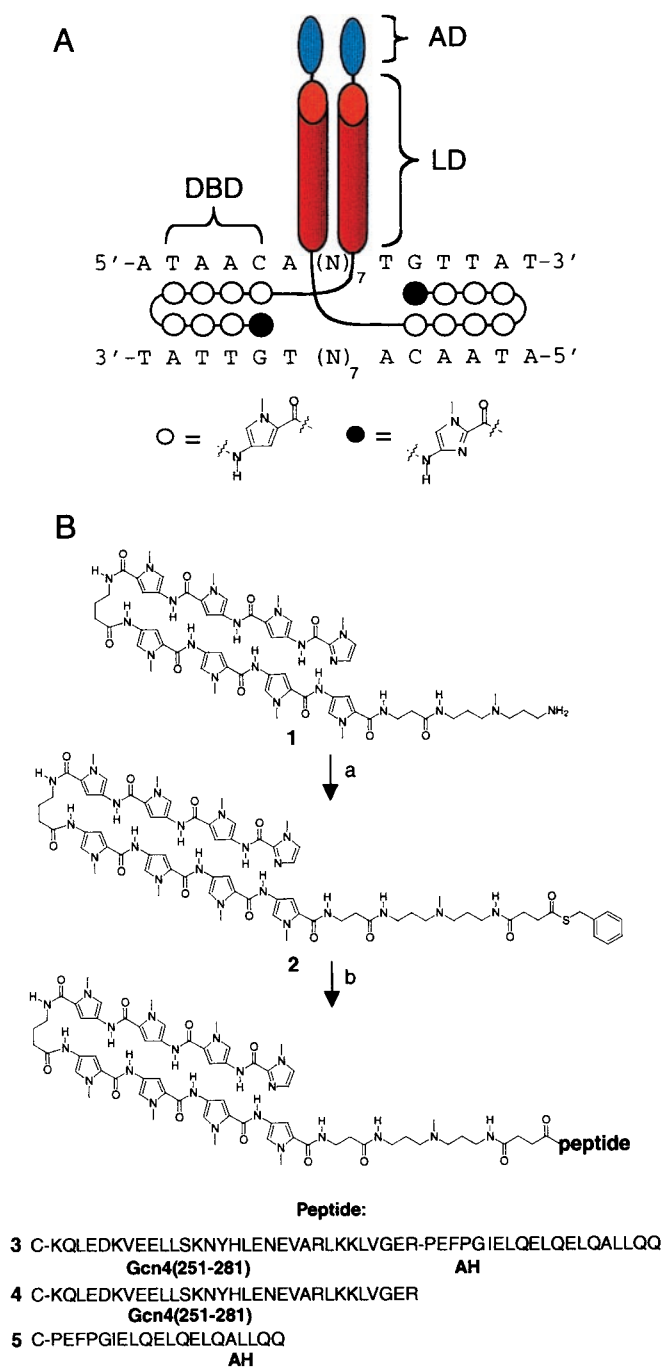
**Synthesis of Conjugates 8–9.** The ethylene glycol-derived linker was prepared as the *N*-*t*-butoxycarbonyl amino acid for use in solid-phase synthesis from 4,7,10-trioxa-1,13-tridecanediamine by monoprotection with *N*-*t*-butoxycarbonyl anhydride followed by reaction with diglycolic anhydride and incorporated into polyamides 6 and 7 according to established protocols (19). Transformation into conjugates 8 (PA-1L-AH) and 9 (PA-2L-AH) was accomplished as outlined above. Yields and characterization: 8: 12%, MALDI-TOF MS [M + H] (average mass) calculated 4164.7, observed 4164.6; 9: 11%, MALDI-TOF MS [M + H] (average mass) calculated 4482.0, observed 4482.2.

Abbreviations: Py, pyrrole; AH, amphipathic helix; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight.

\*A.K.M. and A.Z.A. contributed equally to this work.

<sup>§</sup>To whom reprint requests should be addressed. E-mail: dervan@caltech.edu or m-ptashne@ski.mskcc.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.



**Fig. 1.** Design and synthesis of activators. (A) The synthetic activator consisting of an activation domain (AD), a dimerization or linker domain (LD), and a DNA binding domain (DBD) complexed with the cognate palindromic DNA site of the hairpin polyamide-peptide conjugate. (B) The synthesis of polyamide 1 was accomplished according to established protocols. (a) Treatment of 1 with thiolane-2,5-dione followed by benzyl bromide provided thioester 2 in good yield (53%). (b) Combination of thioester 2 with each peptide in denaturing buffer then provided the targeted conjugates via the native ligation reaction.

**DNase I Footprinting Titrations.** Quantitative DNase I footprinting titrations were carried out in accordance with established protocols on a 3'-<sup>32</sup>P-labeled 271-bp pPT7 *EcoRI/PvuII* restriction fragment (15, 21). The plasmid pPT7 was kindly provided by S. Swalley (California Institute of Technology).

**In Vitro Transcription Assays.** The template plasmid was constructed by cloning a 78-bp oligomer bearing three cognate palindromic sequences into a *BglII* 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. The “mismatch” template was constructed by cloning a 78-bp oligomer containing three palindromic “mismatch” sites into a *BglII* site 30 bp upstream of the TATA box of pMLΔ53 (22). For each reaction, 20 ng of plasmid (30 fmol of palindromic sites) was preincubated with conjugate for 75 min before the addition of 90 ng of yeast nuclear extract in a 25-μl reaction volume under standard conditions (23, 24). The reactions were processed as described (23, 24) and resolved on 8% 30:1 polyacrylamide gels containing 8 M urea. Gels were dried and exposed to photostimulatable phosphorimaging plates (Fuji). Data were visualized by using a Fuji PhosphorImager followed by quantitation using MACBAS software (Fuji).

## Results and Discussion

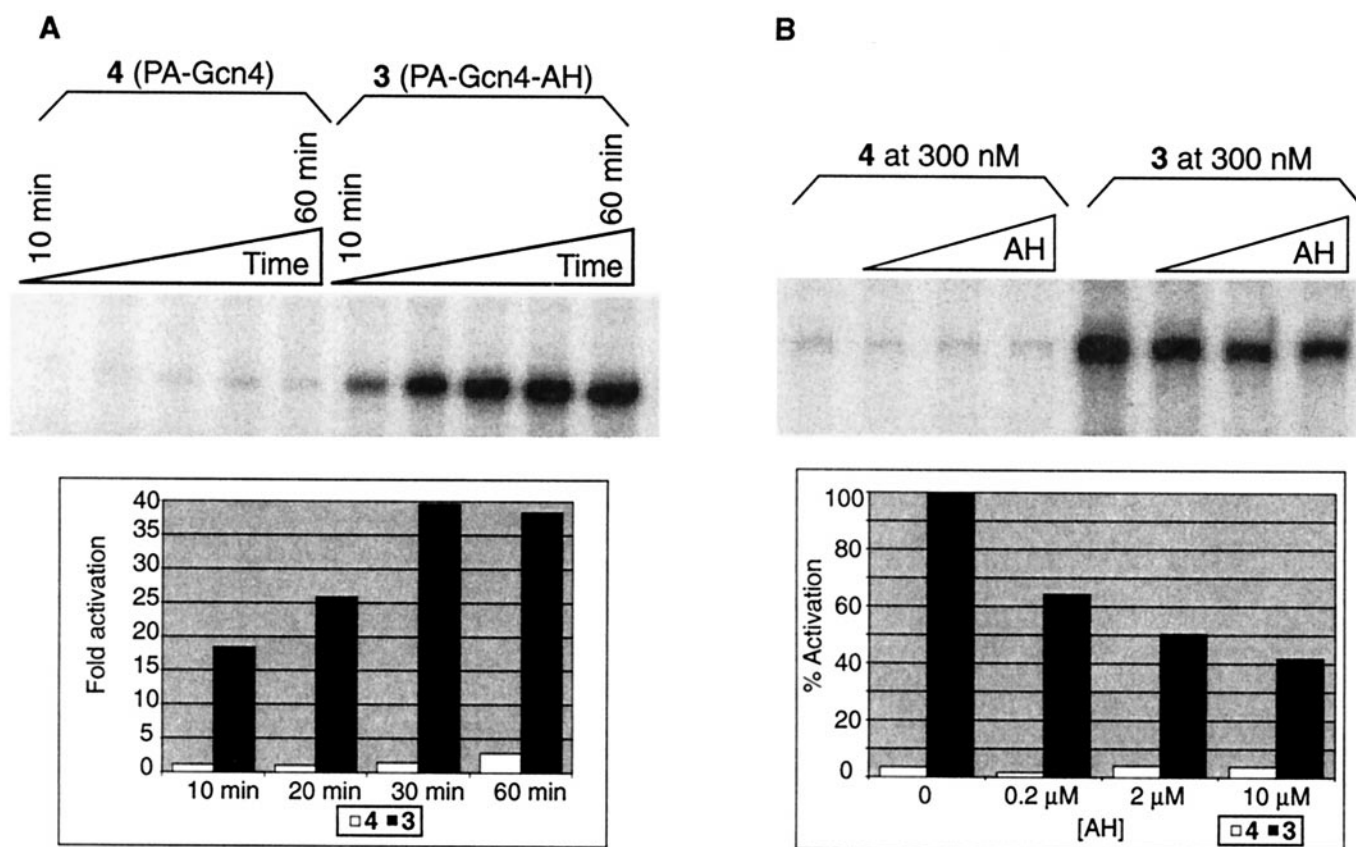
**Synthesis of Artificial Activators.** The hairpin polyamides and peptides were synthesized by solid-phase protocols, and the peptides each contained an N-terminal cysteine for subsequent attachment to the polyamide (19). Polyamide 1 then was treated with thiolane-2,5-dione (20) followed by benzyl bromide to produce thioester 2 functionalized for use in the native ligation procedure described by Kent and colleagues (Fig. 1B) (25). Three polyamide-peptide conjugates were prepared by this sequence. Conjugate 3 (PA-Gcn4-AH) contains the eight-ring hairpin polyamide as the DNA binding module in addition to residues 251–281 of the yeast protein Gcn4 as a dimerization element and the designed peptide AH as the activating region. The two control compounds 4 (PA-Gcn4) and 5 (PA-AH) each lack one or another of the three key functional components.

Model building based on available crystal structures (18, 25) suggested that the polyamide-peptide conjugates would target the palindromic binding site shown in Fig. 1A. The data from a quantitative DNase I footprinting titration between conjugate 3 and DNA containing the 19-bp site were fit by a cooperative binding isotherm ( $K_D = 11$  nM) (Fig. 2A) (21, 27). The decrease in overall binding affinity of conjugate 3 relative to the parent hairpin polyamide is attributed to the attachment of the peptide at the C-terminal position of the polyamide, known to have a deleterious effect.

**Activation of Transcription.** Conjugate 3 (PA-Gcn4-AH) activated transcription in yeast nuclear extracts on a DNA template containing three palindromic binding sites upstream of the start site (Fig. 2B) (23, 24). Thus, inclusion of 3 at 200 nM concentration in the reactions resulted in 13-fold levels of activated transcription over basal levels. In control experiments, polyamide alone (1) or polyamide coupled to the Gcn4 dimerization domain (4) but lacking AH did not stimulate transcription. Furthermore, activation depended on the presence of cognate polyamide binding sites upstream of the transcription initiation site. Thus, on a template with palindromic sites containing a single base pair mismatch at each half site, conjugate 3 failed to activate significantly (Fig. 2C).

**Time Dependence of Activation.** The time course experiment of Fig. 3A reveals that the activation profile of conjugate 3 (PA-Gcn4-AH) was consistent with that previously determined for protein transcriptional activators (28). At 20 min, the level of transcription was 40-fold above the basal level. Fig. 3B shows that at high concentrations of free AH peptide the activation elicited by conjugate 3 was decreased by 50%. AH peptide thus competes with the DNA-bound conjugate 3 for binding to the transcriptional machinery in a phenomenon





**Fig. 3.** Dependence of activation level upon time and activating region. (A) (Upper) Storage phosphor autoradiogram showing an *in vitro* transcription time course experiment with conjugates **3** and **4** present at 300 nM concentration. Aliquots were processed at 10, 20, 30, 40, and 60 min. (Lower) Comparison of the amount of transcript obtained at each time point relative to basal transcription levels in which no conjugate was present (fold activation). (B) (Upper) Storage phosphor autoradiogram showing the effect of increasing concentrations of untethered 21-aa AH peptide on transcription reactions containing either conjugates **4** or **3** at 300 nM concentration. Transcription reactions were performed for 30 min in the presence of 0, 0.2  $\mu$ M, 2  $\mu$ M, and 10  $\mu$ M concentrations of AH peptide. (Lower) For each reaction, the amount of transcript obtained was compared with basal transcription levels to give the respective fold activation value. These values are displayed as percent activation compared with the results from the reaction containing conjugate **3** (lane 5), which is defined as 100%.

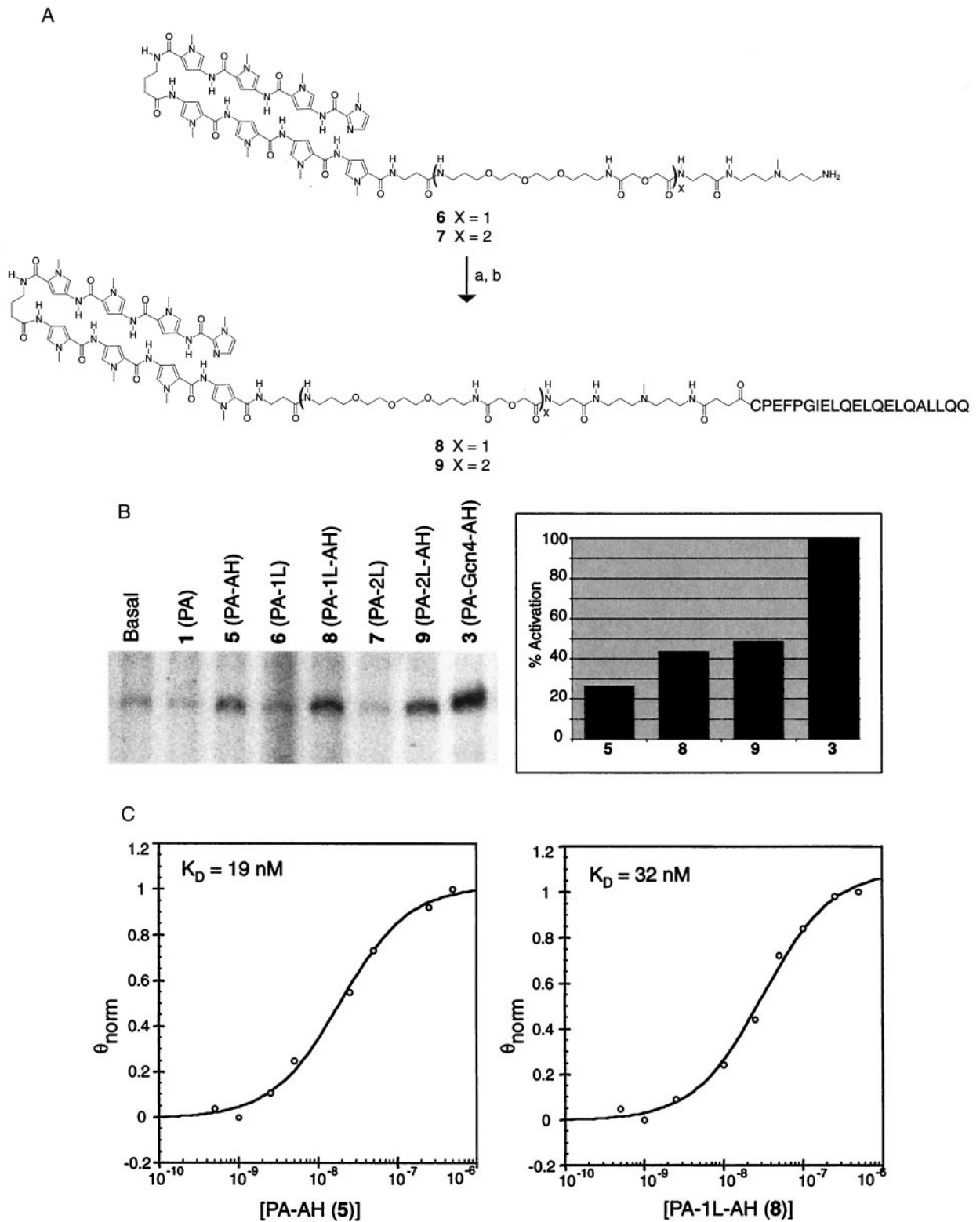
referred to as squelching (29, 30). This reinforces the idea that DNA-tethered AH recruits the transcriptional machinery to the nearby promoter.

**Functional Role of Dimerization Element.** The functional necessity of the dimerization element was investigated by the evaluation of conjugates containing the activator peptide AH separated by flexible straight-chain linkers of 12 atoms (conjugate **5**), 36 atoms (conjugate **8**), and 55 atoms (conjugate **9**) (Fig. 4A). As shown in Fig. 4B, conjugate **8** (PA-1L-AH) activated transcription at approximately 50% the level of conjugate **3** (PA-Gcn4-AH). Increasing the linker length to 55 atoms (conjugate **9**) did not result in a further increase in activation levels; this is likely because of the flexibility of the linker moiety, which may not project AH fully away from DNA. The use of a shorter linker (**5**, PA-AH) provided a conjugate that activated transcription 25% as well as did conjugate **3** (PA-Gcn4-AH), suggesting that spatial separation of the activator module from DNA plays a role in the efficiency of activation. Two observations demonstrate that conjugates **5**, **8**, and **9** do not dimerize. As shown in Fig. 4C, data from quantitative DNase I footprinting titrations were fit by noncooperative isotherms ( $K_D$  for **5** = 19 nM;  $K_D$  for **8** = 32 nM). Furthermore, in contrast to titrations containing conjugate **3**

(PA-Gcn4-AH), DNase I-mediated cleavage was observed at positions between the monomeric binding sites within the palindrome.

Our data demonstrate that each component of a naturally occurring transcription factor can be substituted with a smaller non-natural module to provide an artificial transcription factor with a size of 4.2 kDa. As assayed in our *in vitro* system, dimerization *per se* is not required for activator function, nor does a natural DNA binding domain play a role in activation that cannot be substituted by a hairpin polyamide tethering the activation module to DNA. Finally, because hairpin polyamides composed of Py, imidazole, and hydroxyproline amino acid units can in principle be prepared to target a wide range of predetermined DNA sequences, the class of artificial transcription factors presented here represent a key step toward up-regulation of physiologically relevant genes by small molecules.

We thank X. Lu, S. Swalley, and E. Baird for helpful discussions and S. Malik for the  $\Delta 53pML$  plasmid. We are grateful to the National Institutes of Health for support of this work and for a postdoctoral fellowship to A.K.M. and to the Helen Hay Whitney foundation for a fellowship to A.Z.A.



**Fig. 4.** Substitution of the dimerization module with a flexible ethylene glycol-derived linker. (A) The synthesis of hairpin polyamides **6** and **7** and conjugates **8** and **9** was carried out as described in Fig. 1B. (B) (Left) Storage phosphor autoradiogram showing *in vitro* transcription reactions containing parent hairpin polyamides **1** (lane 2), **6** (lane 4), and **7** (lane 6) or conjugates **5** (lane 3), **8** (lane 5), and **9** (lane 7), which have the AH peptide attached by flexible linkers of increasing length at 500 nM concentration. A higher conjugate concentration relative to the experiments presented in Figs. 2 and 3 was used to accommodate the slightly lower binding affinity (2- to 3-fold) of conjugates **5**, **8**, and **9** relative to conjugate **3** (PA-Gcn4-AH). (Right) The activation levels for conjugates **3**, **5**, **8**, and **9** were determined by comparison with the amount of transcript obtained from reactions containing the relevant parent hairpin polyamides. The fold activation values thus obtained are displayed as percentages relative to the fold activation mediated by conjugate **3**, defined as 100%. (C) Data from DNase I footprinting titrations with **5** and **8**. The curve through each data set is the best-fit Langmuir binding titration isotherm ( $n = 1$ ) obtained from a nonlinear least-squares algorithm.

1. Ptashne, M. & Gann, A. (1997) *Nature (London)* **386**, 569–577.
2. Brent, R. & Ptashne, M. (1985) *Cell* **43**, 729–736.
3. Keegan, L., Gill, G. & Ptashne, M. (1986) *Science* **231**, 699–704.
4. Hope, I. A. & Struhl, K. (1986) *Cell* **46**, 885–894.
5. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. (1988) *Nature (London)* **335**, 563–564.
6. Lefstin, J. A. & Yamamoto, K. A. (1998) *Nature (London)* **392**, 885–888.
7. Giniger, E. & Ptashne, M. (1987) *Nature (London)* **330**, 670–672.
8. Ma, J. & Ptashne, M. (1987) *Cell* **51**, 113–119.
9. Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S. & Schaffner, W. (1994) *Science* **263**, 808–811.
10. Ma, J. & Ptashne, M. (1988) *Cell* **55**, 443–446.
11. Ho, S. N., Biggar, S. R., Spencer, D. M., Schreiber, S. L. & Crabtree, G. R. (1996) *Nature (London)* **382**, 822–826.
12. Nyanguile, O., Uesugi, M., Austin, D. J. & Verdine, G. L. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 13402–13406.
13. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7586–7590.
14. Wade, W. S., Mrksich, M. & Dervan, P. B. (1992) *J. Am. Chem. Soc.* **114**, 8784–8794.
15. Trauger, J. W., Baird, E. E. & Dervan, P. B. (1996) *Nature (London)* **382**, 559–561.
16. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. (1997) *Nature (London)* **387**, 202–205.
17. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M. & Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12890–12895.
18. Ellenberger, T. E., Brandl, C. J., Struhl, L. & Harrison, S. C. (1992) *Cell* **71**, 1223–1237.
19. Baird, E. E. & Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118**, 6141–6146.
20. Kates, M. J. & Schauble, J. H. (1995) *J. Heterocyclic Chem.* **32**, 971–978.
21. Senear, D. F., Brenowitz, M., Shea, M. A. & Ackers, G. K. (1986) *Biochemistry* **25**, 7344–7354.
22. Gu, W., Malik, S., Ito, M., Yuan, C. X., Fondell, J. D., Zhang, X., Martinez, E., Qin, J. & Roeder, R. G. (1999) *Mol. Cell* **3**, 97–108.
23. Lue, N. F., Flanagan, P. M., Kelleher, P. M. III, Edwards, A. M. & Kornberg, R. D. (1991) *Methods Enzymol.* **194**, 545–551.
24. Lue, N. F., Flanagan, P. M., Sugimoto, K. & Kornberg, R. D. (1989) *Science* **246**, 661–664.
25. Dawson, P. E., Muir, T. E., Clark-Lewis, I. & Kent, S. B. H. (1994) *Science* **266**, 776–779.
26. Kielkopf, C. L., Baird, E. E., Dervan, P. B. & Rees, D. C. (1988) *Nat. Struct. Biol.* **5**, 104–109.
27. Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986) *Methods Enzymol.* **130**, 132–181.
28. Wootner, M. & Jaehning, J. A. (1990) *J. Biol. Chem.* **265**, 8979–8982.
29. Gill, G. & Ptashne, M. (1988) *Nature (London)* **334**, 721–724.
30. Tasset, D., Tora, L., Fromental, C., Scheer, E. & Chambon, P. (1990) *Cell* **62**, 1177–1187.