Park, K. Kaushansky, L. Levitt, *J. Biol. Chem.* **268,** 6299 (1993); P. N. Cockerill, M. F. Shannon, A. G. Bert, G. R. Ryan, M. A. Vadas, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 2466 (1993).
10. M. Grilli, J. J.-S. Chiu, M. J. Lenardo, *Int. Rev. Cytol.* **143,** 1 (1993); P. A. Baeuerle, *Biochim. Biophys. Acta* **1072,** 63 (1991).
11. P. A. Baeuerle and D. Baltimore, *Science* **242,** 540 (1988).
12. S. C. Sun, P. A. Ganchi, D. W. Ballard, W. C. Greene, *ibid.* **259,** 1912 (1993); K. Brown, S. Park, T. Kanno, G. Franzoso, U. Siebenlist, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 2532 (1993); R. de Martin *et al.*, *EMBO J.* **12,** 2773 (1993); T. Henkel *et al.*, *Nature* **365,** 182 (1993).
13. E. B. M. Traenckner, S. Wilk, P. A. Baeuerle, *EMBO J.* **13,** 5433 (1994); J. A. DiDonato, F. Mercurio, M. Karin, *Mol. Cell. Biol.* **15,** 1302 (1995); I. Alkalay *et al.*, *ibid.*, p. 1294.
14. A. A. Beg and A. S. Baldwin Jr., *Genes Dev.* **7,** 2064 (1993).
15. K. S. Ullman, J. P. Northrop, C. L. Verweij, G. R. Crabtree, *Annu. Rev. Immunol.* **8,** 421 (1990).
16. B. Gaugler, C. Langlet, J. M. Martin, A. M. Schmitt-Verhulst, A. Guimezanes, *Eur. J. Immunol.* **21,** 2581 (1991).
17. A. Vacca *et al.*, *Mol. Endocrinol.* **3,** 1659 (1989).
18. J. Northrop, G. R. Crabtree, P. S. Matilla, *J. Exp. Med.* **175,** 1235 (1992).
19. The IL-2–luciferase plasmid described by Northrop *et al.* (*19*) was electroporated (960 μF, 320 V) by means of a Bio-Rad gene pulser into FJ8.1 cells (25 μg of DNA per 10^7 cells).
20. Nuclear extracts of FJ8.1 cells were prepared as described [Y. Devary, C. Rosette, J. A. DiDonato, M. Karin, *Science* **261,** 1442 (1993)] and 5-μg samples were incubated with κB, Oct-1, or AP-1 binding site probes. Protein–DNA binding was examined by EMSAs,as described above [M. Garner and A. Revsin, *Nucleic Acids Res.* **9,** 3047 (1981)]
21. O. Leo, M. Foo, D. H. Saiks, L. E. Samelson, J. A. Bluestone, *Proc. Natl. Acad. Sci. U.S.A.* **84,** 1374 (1987).
22. N. Auphan, unpublished data.
23. J. DiDonato and N. Auphan, unpublished data.
24. P. A. Baeuerle and D. Baltimore, *Cell* **53,** 211 (1988).
25. S. Haskill *et al.*, *ibid.* **65,** 1281 (1991).
26. C. Rosette, unpublished data.
27. J. E. Thompson, R. J. Philips, H. Erdjument-Bromage, P. Tempst, S. Ghosh, *Cell* **80,** 573 (1995).
28. J. DiDonato, unpublished data. Antiserum to IκBβ was obtained from S. Ghosh.
29. R. I. Scheinman, P. C. Cogswell, A. K. Lofquist, A. S. Baldwin Jr., *Science* **270,** 283 (1995).
30. A. Ray and K. E. Prefontaine, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 752 (1994); R. I. Scheinman, A. Gualberto, C. M. Jewell, J. A. Cidlowski, A. S. Baldwin Jr., *Mol. Cell. Biol.* **15,** 943 (1995).
31. U. Zabel, T. Henkel, M. Silva, P. A. Baeuerle, *EMBO J.* **12,** 201 (1993); D. E. Cressman and R. Taub, *Oncogene* **8,** 2567 (1993).
32. U. Zabel and P. A. Baeuerle, *Cell* **61,** 255 (1990).
33. E. Kopp and S. Ghosh, *Science* **265,** 956 (1994).
34. B. Stein *et al.*, *EMBO J.* **12,** 3879 (1993).
35. M.-L. Allegre, *et al.*, *J. Immunol.* **146,** 1184 (1991).
36. J. D. Dignam, R. M. Lebovitz, R. G. Roeder, *Nucleic Acids Res.* **11,** 1475 (1983).
37. F. Mercurio, J. A. DiDonato, C. Rosette, M. Karin, *Genes Dev.* **7,** 705 (1993).
38. We thank R. Raghow for critical reading of the manuscript, A. Baldwin for helpful discussions, C. Van Orshoven and P. Alford for expert secretarial assistance, A. M. Schmitt-Verhulst for providing the T cell hybrodoma and laboratory facilities to N.A. that were required for completion of this work, and R. Evans for the antibody to GR. Supported by NIH. N.A., J.D., and A.H. were supported by postdoctoral fellowships from the Association pour la Recherche Contre le Cancer, INSERM, NIH, and the Austrian Science Foundation, respectively. C.R. was supported by a minority research supplement from NIH.

19 June 1995; accepted 9 August 1995

Altered DNA Recognition and Bending by Insertions in the α2 Tail of the Yeast a1/α2 Homeodomain Heterodimer

Yisheng Jin, Janet Mead, Thomas Li, Cynthia Wolberger, Andrew K. Vershon*

The yeast MATα2 and MAT**a**1 homeodomain proteins bind cooperatively as a heterodimer to sites upstream of haploid-specific genes, repressing their transcription. In the crystal structure of α2 and **a**1 bound to DNA, each homeodomain makes independent base-specific contacts with the DNA and the two proteins contact each other through an extended tail region of α2 that tethers the two homeodomains to one another. Because this extended region may be flexible, the ability of the heterodimer to discriminate among DNA sites with altered spacing between α2 and **a**1 binding sites was examined. Spacing between the half sites was critical for specific DNA binding and transcriptional repression by the complex. However, amino acid insertions in the tail region of α2 suppressed the effect of altering an **a**1/α2 site by increasing the spacing between the half sites. Insertions in the tail also decreased DNA bending by **a**1/α2. Thus tethering the two homeodomains contributes to DNA bending by **a**1/α2, but the precise nature of the resulting bend is not essential for repression.

Homeodomain proteins are found in a wide range of eukaryotic organisms and form a large family of transcription factors that function in many different cellular processes (*1*). Although many homeodomain proteins bind DNA with relatively low sequence specificity in vitro, they often confer very specific regulatory activities in vivo (*2*). One mechanism that homeodomain proteins use to achieve their biological specificity is through interactions with additional protein factors (*3*). Such protein-protein interactions increase DNA binding affinity and specificity to the target sites. For example, in the yeast *Saccharomyces cerevisiae*, the α2 homeodomain protein interacts with two other proteins to regulate cell-type specific gene expression (*4*). In haploid α cells, the α2 protein acts in combination with MCM1, a MADS box protein, to bind DNA as a heterotetramer to repress **a**-specific genes (*asg*) (*5*). In **a**/α diploid cells, the α2 protein also interacts with **a**1, another homeodomain protein, to bind DNA as a heterodimer and repress haploid-specific genes (*hsg*) (*6*). Therefore, MCM1 and **a**1 contribute to the target site selection of α2.

In *asg* repression, the MCM1 protein contributes to the sequence specific binding of α2 in two different ways. First, the cooperative interactions between the two proteins increases the apparent DNA binding affinity of the complex (*5*, *7*). Second, interactions with MCM1 dictate the spacing and orientation of the α2 binding sites, which increases the DNA binding specificity of the complex (*8*).

In *hsg* repression, the **a**1 protein appears to play a similar role in helping α2 bind selectively to the **a**1/α2 target sites. It has been shown that protein-protein interactions between α2 and **a**1 contribute to the DNA binding affinity (*9*). However, we wondered whether these protein-protein interactions also dictate the spacing and orientation of the homeodomain DNA binding sites and therefore contribute to the DNA binding specificity of the complex. The crystal structure of the **a**1/α2/DNA ternary complex shows that although there are no direct contacts between the α2 and **a**1 homeodomains, the two proteins interact through a region immediately following the α2 homeodomain that forms a short α helix and makes a set of mainly hydrophobic contacts with **a**1 (*10*). This helix is tethered to the end of the third helix in the α2 homeodomain by an extended linker region (residues 59 to 63 in the crystal structure, residues 189 to 193 of the intact protein) which does not appear to be making contacts with either protein or DNA. Because both NMR and x-ray studies suggest that the linker is extended, it is possible that the linker is somewhat flexible (*10*, *11*). The **a**1 protein may therefore not de-

Y. Jin, J. Mead, A. K. Vershon, Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08855–0759, USA.
T. Li, Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205–2185, USA.
C. Wolberger, Howard Hughes Medical Institute and Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205–2185, USA.
*To whom correspondence should be addressed.

termine the spacing between the DNA binding sites in the same manner as MCM1. We wanted to test whether the protein-protein interactions between **a**1 and α2 are rigidly fixed and whether this interaction determines the spacing of the homeodomain binding sites.

In order to test repression by **a**1/α2 sites with altered spacing, oligonucleotides with a consensus *hsg* site, similar to the one used in the crystallographic study, as well as sites with a one base insertion or deletion in the center of the site, were cloned between the UAS and TATA sequences of a *CYC1-lacZ* fusion promoter. These constructs were co-transformed into yeast with wild-type α2 and assayed for *lacZ* expression by determining β-galactosidase activity (Fig. 1A) (*12*). The first column in Fig. 1 shows the repression data for wild-type α2 protein with the wild-type *hsg* consensus site and with *hsg* sites having altered spacing between the half sites. An insertion or deletion of a single base pair between the two half sites, at a position that is not contacted by either protein, causes a large reduction in repression. In order to correlate repression in vivo with DNA binding in vitro, we assayed the binding affinities with purified **a**1 and α2 proteins in electrophoretic mobility shift assays (*13*). The binding affinity of the wild-type protein to the altered DNA site with a 1–base pair (bp) insertion is at least 100-fold weaker than to the wild-type site (Fig. 1, B and C). Deletion of base pair 11 in the operator resulted in a similar loss of DNA binding activity (*14*). This result shows that **a**1/α2 repression and DNA binding is dependent on the spacing of the two half sites and that the relative position of the sites is rigidly fixed. The failure of **a**1/α2 sites with altered spacing to repress transcription suggests a reason why all known wild-type operators have 6-bp spacing between the α2 and **a**1 half sites. In contrast, the natural α2/MCM1 binding sites have either a 4- or 5-bp spacing between the α2 and MCM1 half sites, suggesting that the spacing requirements between these proteins is more flexible than for **a**1 and α2.

Because the linker region between the α2 homeodomain and the helix that contacts **a**1 is extended, we reasoned that insertion of residues within this linker might be able to suppress the decrease in DNA binding that occurs when the **a**1 and α2 sites are moved apart. On the basis of the crystal structure of the **a**1/α2–DNA complex (*10*), we inserted either one, two, or three glycines between residues 59 and 60 of the linker region and tested how these insertions affect repression from wild-type and altered spacing sites. All of these insertions in the α2 tail were able to repress transcription from a wild-type site at approximately wild-type levels (Fig. 1A). The affinity of these mutant proteins for the wild-type DNA site is also within threefold of wild-type binding affinity (Fig. 1B). These results show that the position of the short helix in the α2 tail in relation to the α2 homeodomain is not critical for protein-protein interactions, cooperative DNA binding, and repression. It also suggests that the extended region is somewhat flexible, because it can adopt an altered conformation that accommodates the inserted glycine residues, while maintaining proper spacing between the **a**1 and α2 homeodomains binding to the wild-type consensus *hsg* site.

We next tested whether insertions in the α2 tail would suppress the defects of the *hsg* site with a single base insertion (Fig. 1A). We found that increasing the number of residues in the extended region results in greater repression from the *hsg* site with a single base insertion. An α2 protein with three glycines inserted into the extended region, α2+3Gly, restores repression to near wild-type levels. This result shows that insertions in the tail can at least partially suppress the effect of altering the spacing of an *hsq* site. These in vivo results predict that α2 proteins with insertions in the tail should bind to the altered site with higher affinity than the wild-type protein. We found that the α2+3Gly mutant bound to the *hsg*+A site with more than 25-fold higher affinity than the wild-type protein (Fig. 1C). Although this level of binding was still down fivefold from the binding affinity to the wild-type site, it shows that increasing the length of the extended region in the tail can restore binding to the altered site. These results also demonstrate that the conserved spacing observed in the wild-type sites is restricted by the length of the extended region between the α2 homeodomain and the short helix that contacts **a**1.

Biochemical and structural studies have shown that the **a**1/α2 complex produces a

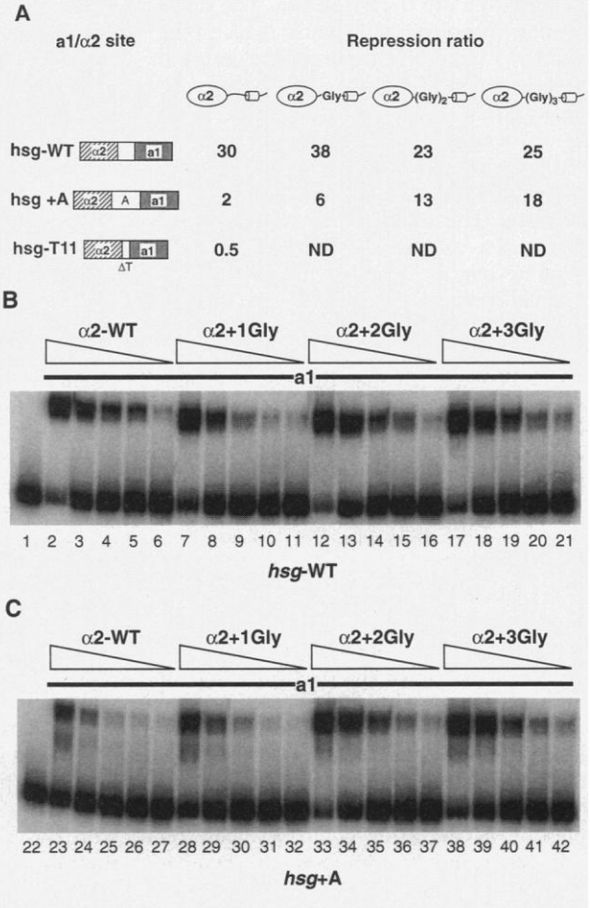

Fig. 1. Glycine insertions in the α2 tail restore the repression and DNA binding of an *hsg* operator with altered spacing. (**A**) A *CYC1-lacZ* promoter reporter vector containing a consensus wild-type (pYJ103) and altered spacing *hsg* sites (pYJ148, pYJ182) were co-transformed into a MAT**a** strain with plasmids that express either no α2 (pAV114), wild-type α2 (pJM130), or mutant α2 proteins with either one, two, or three glycines inserted into the tail (pYJ201, pYJ202, and pYJ203) and assayed for β-galactosidase activity (*12*). Values shown represent the repression ratio calculated by comparing the β-galactosidase activity in the presence and absence of α2. In the absence of α2, a *CYC1-lacZ* construct that contains a wild-type consensus *hsg* site (pYJ103) expresses an average of 335 ± 15 units of β-galactosidase activity. A strain that contains the wild-type α2 protein (pJM130) expresses 11 ± 1 units of β-galactosidase activity from the same *CYC1-lacZ* reporter plasmid (pYJ103) thereby showing ~30-fold repression of the promoter. (**B**) α2 proteins with glycine insertions in the tail bind to the wild-type *hsg* site with wild-type affinity. A labeled fragment containing the wild-type consensus *hsg* site was assayed for binding in the presence of a constant amount of **a**1 and dilutions of wild-type and mutant α2 proteins ranging by fivefold from 3×10^{-10} M (lanes 2, 7, 12, and 17) to 4.8×10^{-13} M (lanes 6, 11, 16, and 21). (**C**) Insertion in the α2 tail restores the binding of α2 to an *hsg* site with altered spacing. A labeled fragment containing the *hsg*+A site was assayed for binding in the presence of a constant amount of **a**1 and dilutions of α2 proteins ranging by fivefold from 1.5×10^{-9} M (lanes 23, 28, 33, and 38) to 2.4×10^{-12} M (lanes 27, 32, 37, and 42). Protein concentrations were normalized by Bradford assay and Coomassie-strained SDS gels.

smooth bend in the DNA that is localized in the **a**1 half site and the center region between the **a**1 and α2 sites (*10*, *15*). This bending may be required for the **a**1 protein to make optimal contacts with the DNA. It may also be the result in part to tethering the two DNA binding domains together at a fixed distance such that the only way both proteins can simultaneously recognize their sites is to bend the DNA. A natural question that arises from our experiments is whether insertions in the tail reduce the DNA bending by **a**1/α2. In performing gel shift experiments of the α2 glycine insertion mutants on the wild-type *hsg* site, we noticed that the **a**1/α2 mutant complexes migrated slightly faster than the complex of the wild-type protein (Fig. 1B). This faster mobility could be the result of a decrease in DNA bending by the **a**1/α2 mutant complex. In order to test this model, we assayed the ability of these proteins to bend a wild-type *hsg* site in circular permutation experiments (Fig. 2) (*16*). As was observed in previous experiments (*15*), the wild-type α2 protein produces a bend of 100° when it binds with **a**1 to the **a**1/α2 site. The single glycine insertion α2 mutant reduces the bend by 13°, and the triple glycine insertion reduces the bend angle by 25°. These results demonstrate that a large component of the DNA bending by **a**1/α2 is the result of tethering the two homeodomains and that increasing the tether length reduces the bending. The **a**1/α2+3Gly complex still produces a bend in the DNA. This DNA bending may be required for **a**1 to make optimal contacts with its DNA site (*10*), or there still may be some distance constraints in tethering the two proteins together. It might be expected that the insertion mutants would bind with higher affinity since they do not require the same energy to bend the DNA as much as wild-type protein. However, we observe that the insertions and wild-type protein have the same apparent affinity for the wild-type DNA site. It is possible that the energy gained in reduced DNA bending by the mutant proteins may be lost through sub-optimal protein-protein or protein-DNA contacts. We conclude however, that since the α2+3Gly mutant binds the *hsg* site with wild-type affinity and represses transcription at nearly wild-type levels, the DNA bending that results from tethering the two proteins together is not absolutely required for DNA binding or repression by the **a**1/α2 complex. These results show that a portion of the bend is produced by protein-protein as well as protein-DNA contacts and that in the case of **a**1/α2, the two contributions can be partially separated.

Our results show that spacing between DNA binding domains contributes to the specificity of a complex of transcriptional regulatory proteins. Specific spacing and orientation of the DNA binding sites is also required for cooperative interactions between α2 and MCM1, as well as in other regulatory systems (*8*, *18*). Spacing and orientation of the sites within a complex of DNA binding proteins may therefore be a general component of target site selection.

Altering the spacing between the DNA binding domain and the protein-protein interaction domain allows the **a**1/α2 complex to recognize different target sites. It is possible that differences in sequence specificity may occur naturally, through altering the spacing between the DNA binding domains of proteins in a complex. It is interesting to note that the *Drosophila* homeodomain protein UBX, which interacts cooperatively with the EXD homeodomain protein, has several naturally occuring isoforms that have different spacing between the DNA binding domain and the YPWM protein-protein interaction motif (*19*). It is possible that in combination with EXD, these UBX isoforms may bind different target sites as a result of different spacing between the homeodomains. As other DNA binding cofactors are identified, it may become more apparent whether differences in spacing between DNA binding domains in a complex is a general mechanism used to alter target site selection.

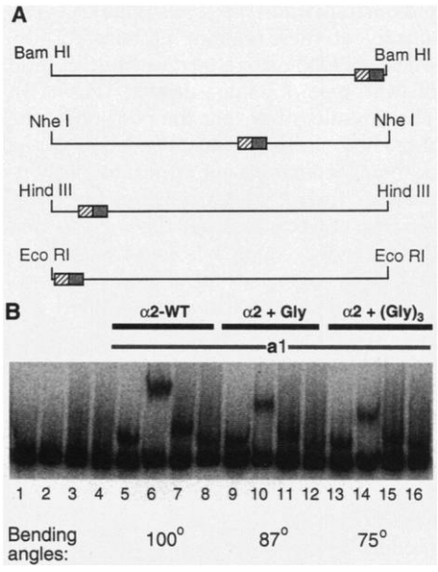

Fig. 2. Insertions in the α2 tail reduce the DNA bending caused by **a**1/α2 DNA binding. (**A**) A diagram indicates the position of the **a**1/α2 site in the 430-bp fragments used as probes in the position permutation assay (*17*). (**B**) The electrophoretic mobility shift assay shows the DNA bending by **a**1 and wild-type α2 (lanes 5 to 8), α2 with one glycine insertion in the tail (lanes 9 to 12), α2 with three glycine insertions in the tail (lanes 13 to 16). Free probes are shown in lanes 1 to 4 and indicate there is no intrinsic bending of the site in the absence of protein. Lanes 1, 5, 9, and 13 contain the Bam HI fragment. Lanes 2, 6, 10, and 14 contain the Nhe I fragment. Lanes 3, 7, 11, and 15 contain the Hind III fragment. Lanes 4, 8, 12, and 16 contain the Eco RI fragment; α2 proteins were added to a final concentration of 6×10^{-11} M.

REFERENCES AND NOTES

1. M. P. Scott, J. W. Tamkun, G. W. Hartzell, *Biochem. Biophys. Acta.* **989**, 25 (1989); W. J. Gehring, M. Affolter, T. Burglin, *Annu. Rev. Biochem.* **63**, 487 (1994).
2. T. Hoey and M. Levine, *Nature* **332**, 858 (1988); C. Desplan, J. Theis, P. H. O'Farrel, *Cell* **54**, 1081 (1988); S. Hayashi and M. Scott, *ibid.* **63**, 883 (1990); P. A. Lawrence and G. Morata, *ibid.* **78**, 181 (1994).
3. S. Stern, M. Tanaka, W. Herr, *Nature* **341**, 624 (1989); D. Xue, Y. Tu, M. Chalfie, *Science* **261**, 1324 (1993); S. Chan, L. Jaffe, M. Capovilla, J. Botas, R. S. Mann, *Cell* **78**, 603 (1994); M. A. van Dijk and C. Murre, *ibid.*, p. 617.
4. I. Herskowitz, *Nature* **342**, 749 (1989); A. D. Johnson, in *Transcriptional Regulation*, K. Yamamoto and S. McKnight, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992), pp. 975–1006.
5. C. A. Keleher, C. Goutte, A. D. Johnson, *Cell* **53**, 927 (1988).
6. C. Goutte and A. D. Johnson, *ibid.* **52**, 875 (1988); A. M. Dranginis, *Nature* **347**, 682 (1990); C. Goutte and A. D. Johnson, *J. Mol. Biol.* **233**, 359 (1993); *EMBO J.* **13**, 1434 (1994).
7. C. A. Keleher, S. Passmore, A. D. Johnson, *Mol. Cell. Biol.* **9**, 5228 (1989).
8. D. L. Smith and A. D. Johnson, *Cell* **68**, 133 (1992).
9. A. Mak and A. D. Johnson, *Genes Dev.* **7**, 1862 (1993).
10. T. Li, M. Stark, A. D. Johnson, C. Wolberger, *Science* **270**, 262 (1995).
11. C. L. Phillips, M. R. Stark, A. D. Johnson, F. W. Dahlquist, *Biochemistry* **33**, 9294 (1994).
12. Oligonucleotides containing glycine codons (GGT) inserted between α2 residues 189 and 190 were synthesized with Bam HI overhang (top strand) or Nhe I overhang (bottom strand). Plasmid pYJ201, pYJ202, and pYJ203, which contain one, two, or three glycines inserted between residues 189 and 190, respectively, were constructed by cloning annealed Bam HI–Nhe I fragments into pJM130, a yeast *CEN, LEU2* plasmid containing an engineered α2 gene with unique Bam HI and Nhe I sites in a 4.3-kb MATα fragment [J. Mead, H. Zhong, T. B. Acton, A. K. Vershon, in preparation]. *CYC1-LacZ* reporter plasmids were constructed by cloning oligonucleotides containing either the consensus wild-type site (pYJ103, hsg-WT: TCGATCATGTAATTAATTACATCA) or altered spacing sites (pYJ148, hsg+A: TCGATCATGTAATTAAATTACATCA; pYJ182, hsg-T11: TCGATCATGTAATAATTACATCA) into the Sal I site of pAV73, a yeast 2μ, *URA3* plasmid containing a *CYC1-lacZ* reporter promoter [A. K. Vershon, N. M. Hollingsworth, A. D. Johnson, *Mol. Cell. Biol.* **12**, 3706 (1992)]. The α2 expression plasmids pAV114, pJM130, pYJ201, pYJ202, and pYJ203 were paired with the appropriate *CYC1-lacZ* reporter vectors and cotransformed into the wild-type *MATa* strain, EG123 (*ura3, trp1, leu2, and his4*) (*5*) and the level of *LacZ* expression in three independent transformants was measured by liquid β-galactosidase assays (*8*).
13. Labeled fragments containing the *hsg* sites were synthesized by polymerase chain reaction using ^{32}P kinased primers that anneal to sequences in the *CYC1* promoter on either side of the *hsg* operator; 0.1 μg of the pYJ103, pYJ148, or pYJ182 was used as the template for 30 rounds of amplification, and the 120-bp labeled operator fragment was then gel-purified, eluted, and precipitated. The α2 protein used in this experiment is a fragment consisting of amino acid residues 123 to 210 (the entire homeodomain and COOH-terminal tail) with six histidine residues fused to the NH_2-terminus. α2 protein was expressed from plasmid pYJ195, a derivative of pET21a(+) (from Novagen) and was purified by chro-

matography on a nickel column according to the manufacturers protocols. Full-length **a**1 protein with six histidine residues fused to the COOH-terminus was expressed from plasmid pYJ173, a derivative of pET21a(+), and purified according to the protocol (Novagen). Electrophoretic mobility shift assays were performed as described (6). Reactions were incubated for 1 hour at RT and then electrophoresed on a nondenaturing 6% polyacrylamide gel with 0.5× TBE. Dried gels were exposed to a phosphor screen and the image was scanned on a Molecular Dynamics model 425 phosphorimager.

14. Y. Jin and A. Vershon, unpublished results.
15. D. L. Smith, A. B. Desai, A. D. Johnson, *Nucleic Acids Res.* **23**, 1239 (1995).
16. H.-M. Wu and D. M. Crothers, *Nature* **308**, 509 (1984).
17. In the position permutation experiment, 430-bp fragments containing the *hsg* site were generated by cutting plasmid pAJ459 (*15*), with appropriate enzymes (Bam HI, Nhe I, Hind III, or Eco RI) and filling in 5′ overhangs with [α-^{32}P]dNTPs. The labeled fragments were purified and electrophoretic mobility shift assays were performed as described above. Reaction mixes were electrophoresed on a nondenaturing 4% polyacrylamide gel with 1X TBE (90 mM trisborate, 2 mM EDTA). Apparent DNA bending angles were calculated based on the Thompson and Landy relationship [J. F. Thompson and A. Landy, *Nucleic Acids Res.* **16**, 9687 (1988)].
18. K. Umesono, K. Murakami, C. Thompson, R. Evans, *Cell* **65**, 1255 (1991); A. M. Naar *et al.*, *ibid.*, p. 1267.
19. K. Kornfeld, R. B. Saint, P. A. Beachy, P. J. Harte, D. A. Peattie, D. S. Hogness, *Genes Dev.* **3**, 243 (1989).
20. We thank C. Abate-Shen, R. Ebright, and D. Norris for comments on the manuscript. Y.J. was supported by a Busch predoctoral fellowship and J.M. was supported by a Busch postdoctoral fellowship. This work was supported by a grant from NSF (MCB-9304526) to C.W. and NIH (GM49265) to A.K.V.

8 September 1995; accepted 22 September 1995

Titins: Giant Proteins in Charge of Muscle Ultrastructure and Elasticity

Siegfried Labeit and Bernhard Kolmerer

In addition to thick and thin filaments, vertebrate striated muscle contains a third filament system formed by the giant protein titin. Single titin molecules extend from Z discs to M lines and are longer than 1 micrometer. The titin filament contributes to muscle assembly and resting tension, but more details are not known because of the large size of the protein. The complete complementary DNA sequence of human cardiac titin was determined. The 82-kilobase complementary DNA predicts a 3-megadalton protein composed of 244 copies of immunoglobulin and fibronectin type III (FN3) domains. The architecture of sequences in the A band region of titin suggests why thick filament structure is conserved among vertebrates. In the I band region, comparison of titin sequences from muscles of different passive tension identifies two elements that correlate with tissue stiffness. This suggests that titin may act as two springs in series. The differential expression of the springs provides a molecular explanation for the diversity of sarcomere length and resting tension in vertebrate striated muscles.

During the past decade, additional filament systems formed by two giant proteins, titin and nebulin, had to be incorporated into the established view of vertebrate striated muscle as a two-filament sliding system [for reviews see (*1*)]. Progress on the molecular characterization of the titin and nebulin filaments initially was limited to histochemical approaches and electron microscopy because of the unusual sizes of these molecules, with molecular weights in the megadalton range. Native molecules of titin (*2*), also referred to as connectin (*3*), are filaments >1 μm long (*4*) that in situ extend from Z discs to M lines (*5*) (Fig. 1). The portion of titin that spans the A band (the region within the thick filament, referred to here as A band titin) is composed of regular arrangements of domains (*5*) that bind to other proteins of the thick filament (*6*, *7*). This suggests that titin is involved in the regulation of the A band ultrastructure. In the I band (the region of titin that spans the thin filament), titin filaments are extensible, as revealed by immunoelectron microscopy (*5*, *8*). This is likely to account for the intrinsic elasticity of vertebrate striated muscle myofibrils (*9*), because degradation of titin by radiation or proteases or its removal by extraction results in a loss of passive tension (*10*). The critical role of the titin filament for muscle structure and function cannot be unraveled without knowledge of its primary structure. Here, we determined the complete complementary DNA (cDNA) sequence of human titin. We suggest that titin specifies a sarcomeric building plan, into which tissue-specific features such as different passive tensions are provided by differential splicing.

We have previously isolated five partial titin cDNA clones from a human heart cDNA library (*6*). The partial titin cDNAs have been extended systematically into both the 5′ and the 3′ direction by anchored polymerase chain reaction (PCR) techniques. A total of 49 extensions were found to link the partial cDNAs into one 82-kb sequence contig. This contig represents the full-length coding sequence of titin expressed in the human heart, as indicated by the presence of untranslated regions at the 5′ and 3′ ends. The 81-kb open reading frame predicts a 26,926-residue protein with a molecular mass of 2993 kD. Ninety percent of the mass is contained in a repetitive structure composed of 244 copies of 100-residue repeats (Fig. 2). These repeats encode 112 immunoglobulin (Ig)-like and 132 FN3-like domains. The pool of 54 partial cardiac titin cDNAs identified 60 λ phage clones in a human genomic DNA library to which the partial cDNAs were assigned in a co-linear order. A subset of 16 phage clones was sufficient to span the 54 cardiac titin cDNAs and covered a contiguous span of 300 kb of genomic DNA. This argues against the presence of cloning artifacts and also suggests that the portion of the titin gene harboring the coding se-

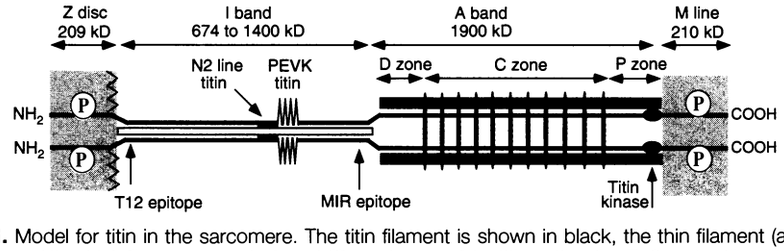

Fig. 1. Model for titin in the sarcomere. The titin filament is shown in black, the thin filament (actin) in yellow, and the thick filament (myosin) in red. The epitopes of the titin antibodies T12 and antibodies to the MIR have been mapped in the sarcomere by immunoelectron microscopy (*5*, *26*); the positions of their epitopes in the titin sequence are known (*27*). Antibodies to the titin kinase domain react with the periphery of the M line (*12*). Therefore, it can be estimated which sections of the titin sequence are in the Z disc, I band, A band, and the M line. For the I band, the range of variation as predicted by the observed splice variants is indicated. The presumed extensible element of the I band, the PEVK element, is located between the N2 line titin and the second tandem Ig block (zig-zag pattern). Within the thick filament in the central C zone (green stripes), titin binds to both the C protein and myosin (*7*, *8*) and is likely to specify the presence of 11 copies of the 430 Å thick filament repeat in vertebrate striated muscles. Phosphorylation of tandemly arranged Ser-Pro repeats in the Z disc and the M line titin (red P) may control integration of the titin filament into Z discs and M lines during myogenesis (*13*).

European Molecular Biology Laboratory, P.O. Box 102209, 69012 Heidelberg, Germany.