albumin) containing 100 μM ATP (except in Fig. 3c) at 27 °C. During the 10 min, the reaction mixture was either kept in darkness (under dim red light as mentioned above) or irradiated continuously with blue light (450 nm, 10 μmol m^{-2} s^{-1}) provided by the Okazaki Large Spectrograph[29]. cAMP was assayed with a BIOTRAK EIA system (Amersham Pharmacia Biotech).

RNAi experiments

Synthesis and electroporation of the dsRNAs were performed in accordance with the methods for *Trypanosoma*[24], with several modifications (for details see Supplementary Information). Because a dark-grown culture of *Euglena* clearly shows the step-up photophobic response, the electroporated cells were maintained in darkness for measurement of this response. In contrast, light-grown cells, which exhibit a clear step-down response, were used for measurement of that response. Both dark-grown and light-grown cells showed a disappearance of the PFB after introduction of the dsRNAs. Measurements of the step-up and the step-down responses were made as described previously[12]. Stimulus light (450 nm) was provided by the Okazaki Large Spectrograph[29].

Received 28 September 2001; accepted 7 January 2002.

1. Senger, H. & Schmidt, W. in *Photomorphogenesis in Plants* 2nd edn (eds Kendrick, R. E. & Kronenberg, G. H. M.) 301–325 (Kluwer Academic, Dordrecht, 1994).
2. Watanabe, M. in *Handbook of Organic Photochemistry and Photobiology* (eds Horspool, W. & Song, P.-S.) 1276–1288 (CRC Press, Boca Raton, 1995).
3. Ahmad, M. & Cashmore, A. R. *HY4* gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor. *Nature* **366,** 162–166 (1993).
4. Cashmore, A. R., Jarillo, J. A., Wu, Y.-J. & Liu, D. Cryptochromes: blue light receptors for plants and animals. *Science* **284,** 760–765 (1999).
5. Huala, E. *et al. Arabidopsis* NPH1: a protein kinase with a putative redox-sensing domain. *Science* **278,** 2120–2123 (1997).
6. Briggs, W. R. *et al.* The phototropin family of photoreceptors. *Plant Cell* **13,** 993–997 (2001).
7. Emery, P. *et al.* CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* **95,** 669–679 (1998).
8. van der Horst, G. T. J. *et al.* Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398,** 627–630 (1999).
9. Lebert, M. in *Comprehensive Series in Photosciences 1: Photomovement* (eds Häder, D.-P. & Jori, G.) 297–341 (Elsevier, Amsterdam, 2001).
10. Buder, J. Zur Kenntnis der phototaktischen Richtungsbewegungen. *Jb. Wiss. Bot.* **58,** 105–220 (1919).
11. Barghigiani, C., Colombetti, G., Franchini, B. & Lenci, F. Photobehavior of *Euglena gracilis*: action spectrum for the step-down photophobic response of individual cells. *Photochem. Photobiol.* **29,** 1015–1019 (1979).
12. Matsunaga, S. *et al.* Discovery of signaling effect of UV-B/C light in the extended UV-A/blue-type action spectra for step-down and step-up photophobic responses in the unicellular flagellate alga *Euglena gracilis*. *Protoplasma* **201,** 45–52 (1998).
13. Benedetti, P. A. & Lenci, F. *In vivo* microspectrofluorometry of photoreceptor pigments in *Euglena gracilis*. *Photochem. Photobiol.* **26,** 315–318 (1977).
14. Brodhun, B. & Häder, D.-P. Photoreceptor proteins and pigments in the paraflagellar body of the flagellate *Euglena gracilis*. *Photochem. Photobiol.* **52,** 865–871 (1990).
15. Bessey, O. A., Lowry, O. H. & Love, R. H. The fluorometric measurement of the nucleotides of riboflavin and their concentration in tissues. *J. Biol. Chem.* **180,** 755–769 (1949).
16. Gomelsky, M. & Kaplan, S. AppA, a redox regulator of photosystem formation in *Rhodobacter sphaeroides* 2.4.1, is a flavoprotein. Identification of a novel FAD binding domain. *J. Biol. Chem.* **52,** 35319–35325 (1998).
17. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1474 (1997).
18. Kaneko, T. *et al.* Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3,** 109–136 (1996).
19. Danchin, A. Phylogeny of adenylyl cyclases. *Adv. Sec. Mess. Phosphoprot. Res.* **27,** 109–162 (1993).
20. Fraser, C. M. *et al.* Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281,** 375–388 (1998).
21. Katayama, M., Wada, Y. & Ohmori, M. Molecular cloning of the cyanobacterial adenylate cyclase gene from the filamentous cyanobacterium *Anabaena cylindrica*. *J. Bacteriol.* **177,** 3873–3878 (1995).
22. Kasahara, M., Yashiro, K., Sakamoto, T. & Ohmori, M. The *Spirulina platensis* adenylate cyclase gene, *cyaC*, encodes a novel signal transduction protein. *Plant Cell Physiol.* **38,** 828–836 (1997).
23. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806–811 (1998).
24. Ngô, H., Tschudi, C., Gull, K. & Ullu, E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA* **95,** 14687–14692 (1998).
25. Darszon, A., Labarca, P., Nishigaki, T. & Espinosa, F. Ion channels in sperm physiology. *Physiol. Rev.* **79,** 481–510 (1999).
26. Hasegawa, E., Hayashi, H., Asakura, S. & Kamiya, R. Stimulation of *in vitro* motility of *Chlamydomonas* axonemes by inhibition of cAMP-dependent phosphorylation. *Cell Motil. Cytoskel.* **8,** 302–311 (1987).
27. Torre, V., Ashmore, J. F., Lamb, T. D. & Menini, A. Transduction and adaptation in sensory receptor cells. *J. Neurosci.* **15,** 7757–7768 (1995).
28. Tessier, L.-H. *et al.* Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*. *EMBO J.* **10,** 2621–2625 (1991).
29. Watanabe, M. *et al.* Design and performance of the Okazaki Large Spectrograph for photobiological research. *Photochem. Photobiol.* **36,** 491–498 (1982).
30. Hurley, J. H. Structure, mechanism, and regulation of mammalian adenylyl cyclase. *J. Biol. Chem.* **274,** 7599–7602 (1999).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We thank J. W. Hastings for critical reading of the manuscript, M. Ohmori and M. Ishiura for discussions, I. Nakashima and Y. Yamakawa for advice on subcellular fractionation, N. Murata for encouragement, Y. Makino for operation of the protein sequencer at NIBB Center for Analytical Instruments, and M. Itoh for technical assistance. Part of this work was done under the NIBB cooperative-research programme for the use of the Okazaki Large Spectrograph. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan, and by Grants for Ground Research for Space Utilization from the Japan Space Forum. M.I. thanks NIBB and BRAIN for a postdoctoral fellowship during part of this study.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.W. (e-mail: machakou@nibb.ac.jp).

Structure and dynamics of KH domains from FBP bound to single-stranded DNA

Demetrios T. Braddock*†, John M. Louis*, James L. Baber*, David Levens† & G. Marius Clore*

** Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases; and † Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA*

Gene regulation can be tightly controlled by recognition of DNA deformations that are induced by stress generated during transcription[1–3]. The KH domains of the FUSE-binding protein (FBP), a regulator of c-*myc* expression[1,4], bind *in vivo* and *in vitro* to the single-stranded far-upstream element (FUSE), 1,500 base pairs upstream from the c-*myc* promoter[4–6]. FBP bound to FUSE acts through TFIIH at the promoter[4]. Here we report the solution structure of a complex between the KH3 and KH4 domains of FBP and a 29-base single-stranded DNA from FUSE. The KH domains recognize two sites, 9–10 bases in length, separated by 5 bases, with KH4 bound to the 5′ site and KH3 to the 3′ site. The central portion of each site comprises a tetrad of sequence 5′d-ATTC for KH4 and 5′d-TTTT for KH3. Dynamics measurements show that the two KH domains bind as articulated modules to single-stranded DNA, providing a flexible framework with which to recognize transient, moving targets.

FBP contains four K homology (KH) repeats[7] separated by linkers of varying lengths[5]. The minimal single-stranded DNA (ssDNA) binding domain, as seen from gel electrophoresis[6], comprises KH3 and KH4 (Fig. 1a). The target for KH3 and KH4 has been localized by sensitivity to the ssDNA selective agent permanganate to a 29-base stretch (M29 in Fig. 1b) in the non-coding strand of the FUSE element (−1525 to −1553 of the c-*myc* gene)[6]. We solved the three-dimensional structure for a complex of relative molecular mass (M_r) of 30,000 comprising the KH3 and KH4 domains of human FBP (residues 278–447 of the full length sequence, referred to as FBP3/4 and numbered 5–174) bound to M29 ssDNA using multidimensional nuclear magnetic resonance[8].

The complex of FBP3/4 with M29 ssDNA as well as with shorter ssDNAs, 25 (M25) and 20 (M20) bases long (Fig. 1b), is long-lived ($K_{diss} \approx 10$ nM) and readily purified as a single complex by gel filtration. Intermolecular nuclear Overhauser experiments (NOEs) showed that KH4 and KH3 bound to bases 4–10 and 16–21, respectively, of M29 ssDNA. However, the spectrum of M29

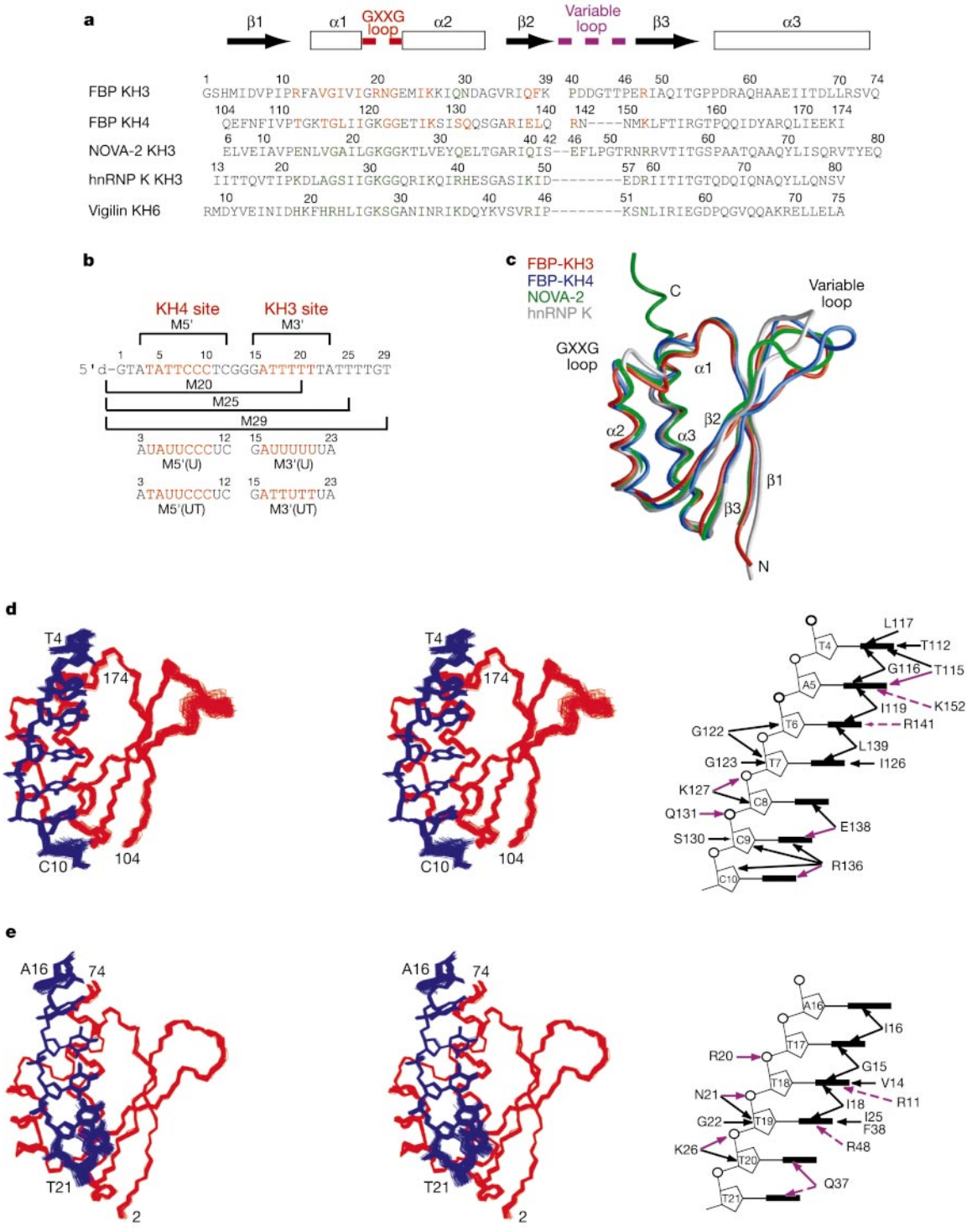

Figure 1 Structural analysis of the FBP3/4-ssDNA complex. **a**, Structure-based sequence alignment of the KH domains of FBP3/4, NOVA-2, hnRNP K and vigilin. Residues of the KH3 and KH4 domains of FBP3/4 that contact ssDNA are indicated in red; the equivalent residues in the other KH domains are coloured green. **b**, ssDNAs used in the current study. NMR analysis was carried out on FBP3/4 complexed to M29 and intermolecular NOE contacts were confirmed using complexes of the isolated KH4 and KH3 domains bound to M5′ and M3′(UT), respectively. **c**, Backbone superposition of the KH domains of FBP3/4 (KH3 and KH4 in red and blue, respectively), NOVA-2 (ref. 9) (green) and hnRNP K[10] (grey). The Cα root mean square (r.m.s.) differences range from 1.1 Å for FBP KH3 and KH4 versus NOVA-2 KH3[9] to 1.6 Å for FBP KH3 versus hnRNP K KH3 (ref. 10). **d**, **e**, Stereoviews showing best-fit superposition of the final 80 simulated annealing structures (protein backbone in red, DNA in blue) and a summary of the observed intermolecular contacts (with H-bonds and salt bridges represented by purple arrows; the dashed arrows indicate potential electrostatic interactions) for the KH4 and KH3 domains, respectively. The coordinate precision for the protein backbone plus DNA heavy atoms is 0.30 and 0.38 Å for the KH3 and KH4 halves of the complex, respectively. (The corresponding values for all heavy atoms are 0.64 and 0.70 Å, respectively.) The experimental NMR restraints for the KH3 and KH4 halves of the complex are as follows: 1,095/949 interproton distances (including 50/68 intermolecular contacts), 244/261 torsion angles, 33/36 $^{3}J_{HN_{\alpha}}$ couplings, 120/121 $^{13}C\alpha/\beta$ shifts, and 61/61 $^{1}D_{NH}$, 47/39 $^{1}D_{NC'}$ and 46/40 $^{2}D_{HNC'}$ dipolar couplings. There are no interproton distance or torsion angle violations >0.5 Å and $>5°$, respectively; the $^{1}D_{NH}$ dipolar coupling *R*-factor[30] is $<10\%$. The percentage residues in the most favourable region of the Ramachandran map is 96% for KH3 and 90% for KH4. A complete table of structural statistics is provided in the Supplementary Information.

ssDNA displayed considerable overlap, particularly at the KH3-binding site. To resolve ambiguities and confirm the assignment of intermolecular NOEs, NMR data for the individual KH domains complexed with shorter ssDNAs (Fig. 1b) were also analysed: KH4 with M5′ (bases 3–12 of M29) and KH3 with M3′(UT) (bases 15–23 of M29 with the T bases at positions 19 and 22 substituted by U to alleviate spectral degeneracy arising from the stretch of six T bases. K_{diss} for the binding of KH4 to M5′ and KH3 to M3′, M3′(U) and M3′(UT) is around 3 μM; K_{diss} for the binding of KH4 to M5′(U) and M5′(UT) is around 12 μM. Thus, substitution of the methyl group of T by a hydrogen in U has no effect for the KH3-binding site and only a marginal effect in energetic terms for the KH4-binding site. The structure of the FBP3/4–M29 complex was determined from 3,153 experimental NMR restraints, including 118 intermolecular interproton distance restraints and 294 protein dipolar couplings. A best-fit superposition of the final 80 simulated annealing structures for the KH4-ssDNA and KH3-ssDNA domains is shown in Fig. 1d, e, respectively.

The KH fold[9–11] comprises three α-helices packed onto a three-stranded antiparallel β-sheet in a β1-α1-α2-β2-β3-α3 topology. The invariant GXXG sequence in the loop connecting helices 1 and 2 is characterized by positive ϕ angles at positions 1, 3 and 4, with the ϕ/ψ angles of residues 3 and 4 in a left-handed helical conformation. Strands β2 and β3 are linked by a loop of variable length (7–10 residues) and conformation. A best-fit superposition of the KH domains of FBP3/4, NOVA-2 (ref. 9) and hnRNP K[10] is shown in Fig. 1c, and the corresponding structure-based sequence alignments are provided in Fig. 1a.

Single-stranded DNA binds to a groove delineated by helices 1 and 2 and the GXXG loop on one side and strand β2 on the other (Fig. 2). The centre of the groove is hydrophobic and the edges are hydrophilic and charged (Fig. 2a, c). The sugar-phosphate backbone is directed towards the left-hand side of the binding site while the bases point towards the centre and right-hand side of the site in the view shown in Fig. 2. The narrow binding site (~10 Å) favours pyrimidines over purines, explaining why the KH3 and KH4 binding sites each contain a single purine, located at the 5′ end. The B-like ssDNA is slightly underwound and extended, positioning the 5′ end at the top of the binding site in close proximity to the GXXG motif (Fig. 2). All the sugar-phosphate backbone torsion angles of the ssDNA lie in a narrow range characteristic of right-handed DNA[12], as evidenced by the restricted dispersion of the ^{31}P resonances (between 4 and 4.5 p.p.m.)[13]. The pattern of intra- and inter-nucleotide NOEs throughout is typical of right-handed DNA[8].

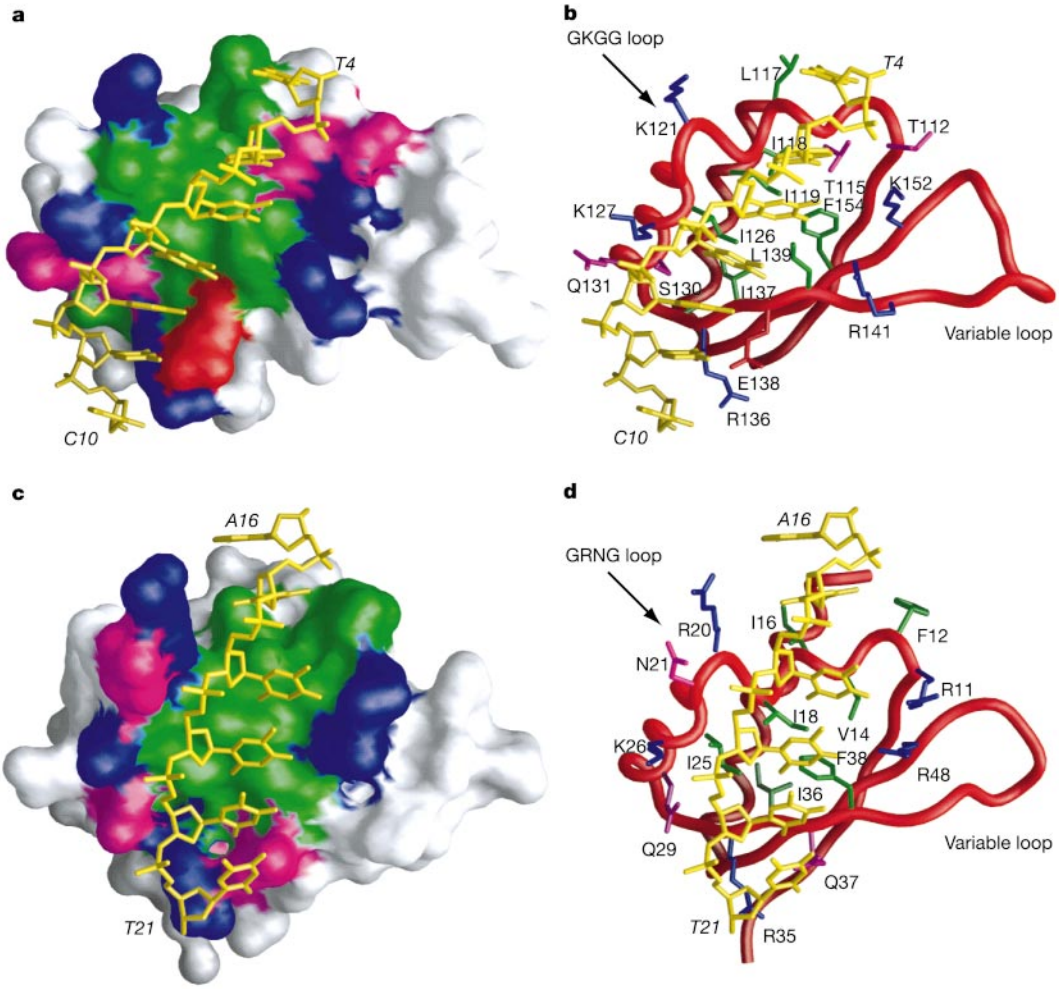

Figure 2 Structure of the FBP3/4–ssDNA complex. **a**, **b**, KH4 domain bound to bases 4–10 of M29. **c**, **d**, KH3 domain bound to bases 16–21 of M29. In **a** and **c** the protein is depicted as a molecular surface with hydrophobic, uncharged hydrophilic, positively charged and negatively charged residues comprising the ssDNA binding sites depicted in green, pink, blue and red, respectively; the ssDNA heavy atoms are shown in yellow. In **b** and **d**, the protein backbone is depicted as a red tube, and the colouring for the protein side chains in the DNA-binding site and ssDNA are the same as in **a** and **c**. The coordinates of the KH3-ssDNA and KH4-ssDNA structures (restrained regularized mean) have been best-fitted to each other, and the same view is shown in all four panels. Nucleotide numbering is in italics.

Hydrophobic contacts to the central T bases of the KH3 (T18, T19) and KH4 (T6, T7) binding sites involve Ile 18, Ile 25 and Phe 38, and Ile 119, Ile 126 and Leu 139, respectively. The methyl groups of these thymines are directed towards solvent, explaining why substitution of T for U has little effect on binding affinity.

Recognition of the KH4 site involves a number of intermolecular H-bonds (Figs 1d and 2a, b): at the 5′ end from the Oγ atom of Thr 115 to the N6 amino group of A5; at the 3′ end from the carboxylate of Glu 138 to the N4 amide group of C9, and from the O2 and O4′ atoms of the base and sugar of C10 to the guanidino group of Arg 136. Other electrostatic interactions involve the Nζ amino group of Lys 152 and the N7 atom of A5, and the guanidino group of Arg 141 and the O4 atom of T6. Contacts to the sugar-phosphate backbone are afforded by Gly 122 and Gly 123 of the GXXG motif, and Lys 127, Ser 130 and Gln 131 of helix 2.

The KH3 domain recognizes a stretch of four T bases (T17–T20), the contacts with which are mainly hydrophobic (Figs 1e and 2c, d). At the 5′ end there are H-bonds between the guanidino groups of Arg 11 and Arg 48 and the O4 atoms of T18 and T19, respectively; at the 3′ end the side-chain amido group of Gln 37 is H-bonded to the O2 atom of T20, and possibly the O2 or N3 atoms of T21. Contacts to the sugar-phosphate backbone are afforded by Arg 20, Asn 21 and Gly 22 of the GXXG motif and Lys 26 of helix 2. Thr 115 in KH4, H-bonded to the base of A5, is replaced by Val 14 in KH3 which favours a T base at this position.

The KH domain presents a scaffold for nucleic acid binding that can be tuned for ssDNA[5,6] or RNA[7,9] specificity. The ssDNA-binding surfaces of the KH domains of FBP3/4 overlap extensively with that of the NOVA-2 KH3 domain that recognizes a stem-loop RNA[9]. There are, however, major differences. Extensive intermolecular contacts are made between ssDNA and residues in helix 2 and the amino-terminal end of strand β2 in the FBP3/4 complex (Fig. 2), which are not seen in the NOVA-2 KH3–RNA complex where equivalent interactions are precluded owing to the presence of the wider double-stranded RNA stem. In contrast, in the NOVA-2 KH3–RNA complex contacts are made to both arms of the RNA hairpin and, as a consequence, the binding surface also includes the flexible loop (connecting strands β2 and β3) and the carboxy-terminal end of helix 3. The latter extends well beyond the protein core of NOVA-2 KH3 (Fig. 1c), presumably as a consequence of protein–RNA contacts and/or crystal packing. The core of the RNA-recognition sequence, 5′r-UCAC[9], is located and oriented similarly to the central portion of the KH4 (5′d-ATTC) and KH3 (5′d-TTTT) ssDNA recognition sequences, but the RNA and ssDNA tetrads have very different conformations. The bases of the ssDNA tetrads are stacked upon one another (Fig. 2). In contrast, the first three bases of the RNA tetrad are not stacked: the base at position 2 is oriented at right angles to the bases at positions 1 and 4, and concomitantly the conformation of the sugar-phosphate backbone departs markedly from that characteristic of right-handed RNA or DNA[12]. The unusual features of the RNA tetrad arise from a network of inter- and intra-molecular H-bonds involving the bases and 2′-OH groups of the ribose sugars.

The KH3 and KH4 domains of FBP (Fig. 1a) do not interact with one another in the FBP3/4-M29 ssDNA complex. They are connected by a glycine-rich, flexible 30-residue linker (Fig. 3a), and the DNA is bridged by 5 bases (bases 11–15) that do not display any contacts to protein. If the FBP3/4–M29 ssDNA complex were rigid, dipolar couplings measured for both halves of the complex in a dilute liquid crystalline medium, such as phage fd[14], would be described by a single alignment tensor[15,16]. However, the magnitude of the alignment tensor[15] for the KH3 domain is half that of the KH4 domain in the complex (Fig. 3b), indicative of significant interdomain motion[16] (see Supplementary Information for further details), which was characterized by model free analysis of ^{15}N relaxation data using the extended Lipari–Szabo model[17,18]. Assuming an axially symmetric diffusion tensor[18], the data yield values of 21.5 and 1.85 ns for the effective overall rotational correlation time and diffusion anisotropy, respectively, of the whole complex. The timescale and magnitude of the interdomain motions are described by an effective slow internal correlation time (τ_s) of 4.1 ns for KH3 and 3.6 ns for KH4, and an order parameter S_s^2 of 0.67 for KH3 and 0.70 for KH4, which correspond to the two KH domains wobbling independently in cones with semi-angles of about 30° (Fig. 3c). Alignment of the principal components of the diffusion tensors of the two domains indicates that the average orientation of the two domains is parallel (average interhelical angle between the third helix of each domain is about 1°, see Fig. 3c).

The interdomain mobility of the FBP3/4–M29 ssDNA complex is

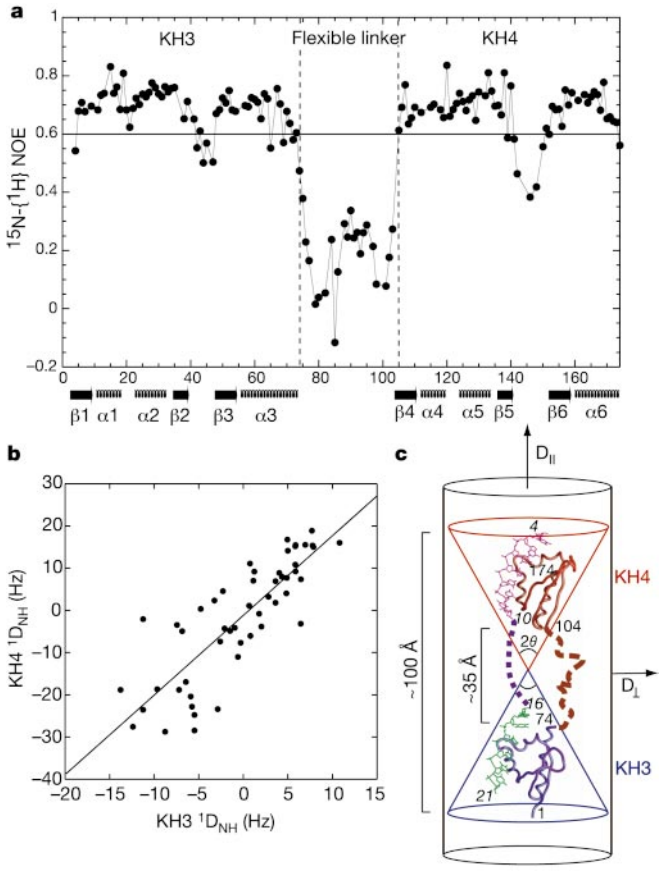

Figure 3 Interdomain motion in the FBP3/4–M29 ssDNA complex, **a**, ^{15}N-{^{1}H} NOE values as a function of residue number. The low NOE values for the 30-residue linker indicate high flexibility for this segment of the polypeptide chain. **b**, Correlation of $^1D_{NH}$ dipolar couplings of structurally equivalent residues measured for the KH4 and KH3 domains in the FBP3/4–M29 complex. The correlation coefficient is 0.83, as expected for very similar structures, but the magnitude of the alignment tensor[15] for the N–H vectors in the KH3 domain (–7.2 Hz) is half that in the KH4 domain (–14.5 Hz), diagnostic of significant interdomain motion[16]. **c**, Depiction of interdomain motion in the FBP3/4–M29 complex. The red and blue cones indicate the slow motion of the KH4 and KH3 halves of the complex, respectively. The KH4 and KH3 domains are shown as red and blue ribbons, respectively, and the ssDNA bound to them is shown in purple and green, respectively. The 5-base linker for the ssDNA and 30-residue linker for the protein are depicted by blue and red dashed lines, respectively. The semi-cone angle θ, derived from the internal slow order parameter S_s^2 is about 30° for both domains. $D_{||}$ nd $D_\perp$ represent the parallel and perpendicular components of an axially symmetric diffusion tensor. The two domains are aligned relative to the long axis of the diffusion tensor. The overall length of the complex and the separation between the two domains is indicated and calculated as described in the text. Nucleotide numbering is in italics.

determined by the intrinsic flexibility of the ssDNA (bases 11–15 of M29) linking the KH3 and KH4 binding sites. The 30-residue protein linker between the KH3 and KH4 domains is highly disordered (Fig. 3a), such that in free FBP3/4 the two domains reorient essentially independently of each other (unpublished data), as in the case of other modular proteins connected by long linkers[16]. Further, the ssDNA linker, although it is flexible, is not disordered because the observed NOEs are typical of right-handed DNA[8].

The distances (L_i) from the centre of KH3 and KH4 domains to the flexible hinge point at the centre of the overall complex can be derived from the relation $L_i = (D_{trans,i}/D_{w,i})^{1/2}$, where $D_{trans,i}$ and $D_{w,i}$ are the translational and rotational diffusion coefficients of domain i (ref. 19). $D_{trans,i}$ is calculated from the hydrodynamic radius (~16.5 Å for each domain) and $D_{w,i}$ from the values of S_S^2 and τ_s (ref. 19 and Supplementary Information). This yields $D_{trans,i} \approx 1.9 \times 10^{-10}\ m^2\ s^{-1}$, $D_{w,i} \approx 1.7 \times 10^{7}\ s^{-2}$, and $L_i \approx 33$ Å. The overall average length of the complex is therefore about 100 Å, corresponding to a cylinder length to diameter ratio of 3 with a predicted diffusion anisotropy of 1.8 (ref. 20), in agreement with the value determined from the model free analysis of the data. These dimensions imply that the average distance between the C terminus of KH3 (residue 74) and the N terminus of KH4 (residue 104) is about 35 Å, which is close to the expected average end-to-end distance of about 40 Å for a random-coil polypeptide of $n = 30$ residues and a 40% Gly content[20]. The intervening stretch of ssDNA comprises six base steps, so the average rise for this segment of ssDNA is around 5.8 Å, about 70% longer than regular B-DNA.

Interdomain motion observed in the FBP3/4–M29 ssDNA complex in which the linkers, both protein and ssDNA, possess intrinsic flexibility, is likely to be an essential component of FBP function. This property can readily accommodate changes in the direction of the DNA. FBP interacts directly with and modulates the helicase activity of TFIIH[4], a generalized transcription factor that has been proposed to act as a torque-generating machine[21] and is required throughout the early stages of transcription from initiation through promoter escape. Hence, interdomain flexibility in the FBP–ssDNA complex allows FBP bound to FUSE to maintain contact with a moving component, TFIIH, of the transcription machinery. As such, the ssDNA-binding domain of FBP presents an unusual architectural transcription factor: the presence of multiple KH domains, separated by linkers, generates a series of single-stranded flexible hinge points in the DNA matrix, facilitating interactions between other, sometimes mobile, components of the transcription machinery that would otherwise be prohibited by stereochemical and topological restraints imposed by the more rigid double-stranded DNA. This predicts that reduction in interdomain flexibility, by reducing the length of the intervening ssDNA sequence between the KH3- and KH4-binding sites, should decrease FUSE activity, whereas increasing the length of this linker segment slightly should have less effect as it could be accommodated by the flexible protein linker between the two KH domains. This is borne out experimentally: deletion of four of the five intervening bases between the KH4 and KH3 ssDNA binding sites (corresponding to bases 11–14 of M29) reduces c-*myc* reporter expression approximately fourfold[22]; conversely, insertion of an additional four bases between the two ssDNA binding sites has minimal effect on FUSE activity (D.L. and M. Avigan, unpublished work).

Underlying the structure of the FBP3/4–ssDNA complex is the concept that DNA transcription can be monitored and controlled by the recognition of ssDNA generated during the course of transcription. The present structure clearly shows that ssDNA located in a *cis*-regulatory element (FUSE) upstream from the c-*myc* promoter can be recognized specifically by FBP3/4, forming a high-affinity complex in solution. These features, combined with previous functional studies on the role of FBP at the c-*myc* locus[1,4–6], provide the experimental basis for this mechanism of transcriptional control. Finally, we note that overexpression of c-*myc* is associated with many human cancers[23] and that loss of FBP function shuts off c-*myc* expression and arrests cellular proliferation[24]. Thus, the FBP3/4–ssDNA complex may represent a potentially useful target for therapeutic intervention. □

Methods

Sample preparation

FBP3/4, comprising residues 278–447 of human FBP[2] and numbered from residues 5–174, as well as the individual KH3 (residues 5–77) and KH4 (residues 101–174) domains were cloned, expressed and purified by affinity chromatography using standard procedures. Samples for NMR contained 1:1 complexes of protein (^{15}N, $^{15}N/^{13}C$ or $^{15}N/^{13}C/^{2}H$/VLI-methyl protonated) and ssDNA in 50 mM sodium phosphate, 20 mM EDTA and 0.02% NaN_3, pH 6.8. Further details are provided in Supplementary Information.

NMR spectroscopy

All NMR experiments were carried out at 35 °C on Bruker 600, 750 and 800 MHz spectrometers. ^{1}H, ^{15}H and ^{13}C backbone and side-chain resonances of the protein were assigned by three-dimensional (3D) double- and triple-resonance NMR experiments[8]; ^{1}H resonances for the ssDNA by two-dimensional (2D) ^{12}C-filtered experiments[8]; and ^{31}P resonances of the ssDNA from a 2D ^{1}H-^{31}P correlation spectrum. The observation of strong $^{31}P(i)$–H4′(i) correlations and the absence of $^{31}P(i)$–H5′/H5″(i) correlations indicates that the β and γ sugar-phosphate torsion angles are in the *t* and g^+ conformations, respectively[13]. Interproton distance restraints (classified into ranges[8]) within the protein were derived from 3D and four-dimensional (4D) ^{15}N- and ^{13}C-separated nuclear Overhauser enhancement (NOE) experiments; within the DNA from 2D ^{12}C-filtered NOE experiments; and between the protein and DNA from 3D ^{13}C-separated/^{12}C-filtered and ^{15}N-separated/^{12}C-filtered NOE experiments[8]. Backbone ϕ/ψ torsion angle restraints were derived from backbone chemical shifts using the program TALOS[25]. Heteronuclear ^{3}J couplings were measured by quantitative J-correlation spectroscopy[8]. Dipolar couplings were measured in a dilute liquid crystalline medium of 25 mg ml^{-1} phase fd[15].

^{15}N relaxation measurements

^{15}N T_1, T_2 and NOE measurements were carried out at 600 and 750 MHz and analysed essentially as described[18]. All the relaxation data were fitted simultaneously on the basis of the atomic coordinates optimizing the value of the effective overall rotational correlation time and anisotropy for the whole complex; a single-order parameter S_f^2 used to describe all fast internal motions; the order parameters (S_s^2) and correlation times (τ_s) for slow interdomain motions of the two domains; and the two Euler angles describing the orientation of the long axis of the rotational diffusion tensor relative to the coordinates of each domain. The correlation time for fast internal motions, τ_e, was held fixed at 15 ps, but the results are not affected for values of less than 30 ps (ref. 17). The optimized value of S_f^2 is 0.85, typical of fast librational motions[17–19]. The value of the error function *E*/*N* is 3.3.

Structure calculations

Structures were calculated from the experimental restraints by simulated annealing in torsion angle space[26] using XPLOR_NIH (http://nmr.cit.nih.gov/xplor_nih). The non-bonded contacts in the target function are represented by a quartic van der Waals repulsion term supplemented by a torsion angle database potential of mean force[27]. Structure figures were generated with the programs VMD-XPLOR[28] and GRASP[29].

Received 27 July; accepted 11 December 2001.

1. Liu, J. *et al.* Defective interplay of activators and repressors with TFIIH in xeroderma pigmentosum. *Cell* **104,** 353–363 (2001).
2. Rothman-Denes, L. B., Dai, X., Davydova, E., Carter, R. & Kazimierczak, K. Transcriptional regulation by DNA structural transitions and single-stranded DNA binding proteins. *Cold Spring Harbor Symp. Quant. Biol.* **63,** 63–73 (1999).
3. Michelotti, E. F., Sanford, S. & Levens, D. Marking of active genes on mitotic chromosomes. *Nature* **388,** 895–899 (1997).
4. Liu, J. *et al.* The FBP interacting repressor targets TFIIH to inhibit activated transcription. *Mol. Cell* **5,** 331–341 (2000).
5. Duncan, R. C. *et al.* A sequence-specific, single strand binding protein activates the far-upstream of c-*myc* and defines a new DNA binding motif. *Genes Dev.* **8,** 465–480 (1994).
6. Michelotti, G. A. *et al.* Multiple single-stranded *cis* elements are associated with activated chromatin of the human c-*myc* gene *in vivo*. *Mol. Cell. Biol.* **16,** 2656–2669 (1996).
7. Burd, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA binding proteins. *Science* **265,** 615–621 (1994).
8. Clore, G. M. & Gronenborn, A. M. Determining structures of larger proteins and protein complexes by NMR. *Trends Biotechnol.* **16,** 22–34 (1998).
9. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100,** 323–332 (2000).
10. Baber, J. L., Libutti, D., Levens, D. & Tjandra, N. High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a *c-myc* transcription factor. *J. Mol. Biol.* **289,** 949–962 (1999).
11. Musco, G. *et al.* Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* **85,** 237–245 (1996).
12. Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).

13. Gorenstein, D. G. Conformation and dynamics of DNA and protein-DNA complexes by ^{31}P NMR. *Chem. Rev.* **94,** 1315–1138 (1994)..
14. Clore, G. M., Starich, M. R. & Gronenborn, A. M. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc.* **120,** 10571–10572 (1998).
15. Clore, G. M., Gronenborn, A. M. & Bax, A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **131,** 159–162 (1998)
16. Braddock, D. T., Cai, M., Baber, J. L., Huang, Y. & Clore, G. M. Rapid identification of medium to large scale interdomain motion in modular proteins using dipolar couplings. *J. Am. Chem. Soc.* **123,** 8634–8635 (2001).
17. Clore, G. M. *et al.* Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **112,** 4989–4990 (1990).
18. Baber, J. L., Szabo, A. & Tjandra, N. Analysis of slow interdomain motion of macromolecules using NMR relaxation data. *J. Am. Chem. Soc.* **123,** 3953–3959 (2001).
19. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear-magnetic-relaxation in macromolecules. 1 Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559 (1980).
20. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry* Part II, 564 and Part III, 1008–1010 (Freeman, San Francisco, 1980).
21. Kim, T.-K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288,** 1418–1421 (2000).
22. Avigan, M. I., Stober, B. & Levens, D. A far upstream element stimulates c-myc expression in undifferentiated leukemia cells. *J. Biol. Chem.* **265,** 18538–18545 (1990).
23. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* **16,** 653–699 (2000).
24. He, L. *et al.* Loss of FBP function arrests cellular proliferation and extinguishes *c-myc* expression. *EMBO J.* **19,** 1034–1044 (2000).
25. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database of protein chemical shift and sequence homology. *J. Biomol. NMR* **31,** 289–302 (1999).
26. Schwieters, C. D. & Clore, G. M. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J. Magn. Reson.* **152,** 288–302 (2001).
27. Clore, G. M. & Gronenborn, A. M. New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl Acad. Sci. USA* **95,** 5891–5898 (1998).
28. Schwieters, C. D. & Clore, G. M. The VMD-XPLOR visualization package for NMR structure refinement. *J. Magn. Reson.* **149,** 239–244 (2001).
29. Nicholls, A., Sharp, K. A. & Honig, G. Protein folding and association: insights into interfacial and thermodynamic properties of hydrocarbons. *Proteins* **17,** 297–309 (1991).
30. Clore, G. M. & Garrett, D. S. R-factor, free R and complete cross-validation for dipolar coupling refinement of NMR structures. *J. Am. Chem. Soc.* **121,** 9008–9012 (1999).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We thank D. Garrett and F. Delaglio for software support; C. A. Bewley, M. Caffrey, W. A. Eaton, J. Kuszewski, L. Murphy, C. Schwieters, A. Szabo and N. Tjandra for discussions; and C. A. Bewley for critically reading the manuscript. This work was supported in part by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.M.C. (e-mail: clore@speck.niddk.nih.gov). Coordinates have been deposited in the Protein Data Bank (RCSB accession code 1J4W).

correction

Transmission potential of smallpox in contemporary populations

Raymond Gani & Steve Leach

Nature **414,** 748–751 (2001).

There was an error in equations (1) in this Letter. The θI terms in dI/dt and dQ/dt should have been omitted. Thus, the set of differential equations should have been:

$$\begin{aligned}
\frac{dS}{dt} &= \chi_1(1-\epsilon_1)C_i - \beta(\varphi+\rho-\varphi\rho)SI & \frac{dI}{dt} &= \alpha(1-\theta)E_n - \gamma I \\
\frac{dE_n}{dt} &= \beta\varphi(1-\rho)SI - \alpha E_n & \frac{dQ}{dt} &= \alpha(1-\epsilon_2)E_i + \alpha\theta E_n - \chi_2 Q \\
\frac{dE_i}{dt} &= \beta\varphi\rho SI - (\chi_1\epsilon_2 + \alpha(1-\epsilon_2))E_i & \frac{dU}{dt} &= \gamma I + \chi_2 Q \\
\frac{dC_i}{dt} &= \beta\rho(1-\varphi)SI - \chi_1 C_i & \frac{dV}{dt} &= \chi_1(\epsilon_2 E_i + \epsilon_1 C_i)
\end{aligned} \qquad (1)$$

and the definition of θ in Table 2 should have read 'the proportion of infectious individuals from E_n that are quarantined'. This change only affects the analysis of the data for Kosovo. The value of R_0 that gives the best fit to the data for Kosovo is now 10 (5 discounting hospital-associated cases), rather than 10.8 (5.4 discounting hospital-?associated cases), as stated in the original paper in Table 1. This small error does not alter our conclusions. □