V-ATPase fragment was amplified from the genomic DNA of Osome plants by PCR using the primers of its conserved region[26]. Washing and detection were carried out as described for DNA gel blot analysis.

Detection of transgene by PCR

S-genotypes of self-pollinated and crossed progeny were determined as follows. Genomic DNA was prepared from young leaves. The SLG^{28} transgene was amplified by PCR with SLG^{28} specific primers, PS18 and PS15 (ref. 27). The SRK^{28} transgene was amplified by PCR with SRK^{28} specific primers, PK28 (5′-CCTCTTATATTTTTCTGCCTCTGG-3′) and PK4 (ref. 28). PK28 was designed on the basis of the nucleotide sequence of the transmembrane domain of SRK^{28}, and PK4 was designed on that of exon 4 of SRK^{28}. The expected PCR product of the SRK^{28} gene was 648 bp. To discriminate between the SRK^{28} transgene and the endogenous SRK^{52} gene, PCR-RFLP was conducted by using the primers, PS18 and B (ref. 29) to obtain PCR products, which were then digested with *Mbo*I and electrophoresed on 5% polyacrylamide gels; the fragments were visualized by silver-staining. The endogenous SLG^{52} and SLG^{60} genes were amplified by PCR with PS5 and PS15, and PS3 and PS21, respectively[13,27].

Immunoblot analysis

Total protein was extracted from five stigmas for the S^{28} homozygote and ten stigmas for all the other plants in 50 mM Tris-HCl pH 7.5. The extract was subjected to thin-layer polyacrylamide gel IEF (Ampholine PAG Plate, pI 3.5–9.5; LKB Pharmacia) and transferred to PVDF membranes (Millipore) by electroblotting. SLG proteins were detected with the anti-SLG[43] (a class I SLG) monoclonal antibody which crossreacts with most of class I SLGs (ref. 30).

Received 21 October; accepted 14 December 1999.

1. Bateman, A. J. Self-incompatibility systems in angiosperms. III. Cruciferae. *Heredity* **9,** 52–68 (1955).
2. Stein, J. C. *et al.* Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proc. Natl Acad. Sci. USA* **88,** 8816–8820 (1991).
3. Takayama, S. *et al.* Sequences of S-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. *Nature* **326,** 102–105 (1987).
4. Nasrallah, J. B., Kao, T.-h., Chen, C.-h., Goldberg, M. L. & Nasrallah, M. E. Amino-acid sequence of glycoproteins encoded by three alleles of the *S*-locus of *Brassica oleracea*. *Nature* **326,** 617–619 (1987).
5. Goring, D. R., Glavin, T. L., Schafer, U. & Rothstein, S. J. An *S* receptor kinase gene in self-compatible *Brassica napus* has a 1-bp deletion. *Plant Cell* **5,** 531–539 (1993).
6. Nasrallah, J. B., Rundle, S. J. & Nasrallah, M. E. Genetic evidence for the requirement of the *Brassica S*-locus receptor kinase gene in the self-incompatibility response. *Plant J.* **5,** 373–384 (1994).
7. Nasrallah, M. E., Kandasamy, M. K. & Nasrallah, J. B. A genetically defined *trans*-acting locus regulates *S*-locus function in *Brassica*. *Plant J.* **2,** 497–506 (1992).
8. Goring, D. R. & Rothstein, S. J. The *S*-locus receptor kinase gene in a self-incompatible *Brassica napus* line encodes a functional serine/threonine kinase. *Plant Cell* **4,** 1273–1281 (1992).
9. Nasrallah, J. B. & Nasrallah, M. E. Pollen–stigma signaling in the sporophytic self-incompatibility response. *Plant Cell* **5,** 1325–1335 (1993).
10. Hinata, K., Watanabe, M., Toriyama, K. & Isogai, A. A review of recent studies on homomorphic self-incompatibility. *Int. Rev. Cytol.* **143,** 257–296 (1993).
11. Suzuki, G. *et al.* Genomic organization of the *S* locus: Identification and characterization of genes in *SLG/SRK* region of S^9 haplotype of *Brassica campestris* (syn. *rapa*). *Genetics* **153**, 391–400 (1999).
12. Schopfer, C. R., Nasrallah, M. E. & Nasrallah, J. B. The male determinant of self-incompatibility in *Brassica*. *Science* **286,** 1697–1700 (1999).
13. Takasaki, T. *et al.* Introduction of *SLG* (*S*-locus glycoprotein) alters the phenotype of endogenous *S*-haplotype, but confers no new *S*-haplotype specificity in *Brassica rapa* L. *Plant Mol. Biol.* **40,** 659–668 (1999).
14. Conner, J. A. *et al.* Transgene-induced silencing of *S*-locus genes and related genes in *Brassica*. *Plant J.* **11,** 809–823 (1997).
15. Stahl, R. J., Arnoldo, M., Glavin, T. L., Goring, D. R. & Rothstein, S. J. The self-incompatibility phenotype is altered by the transformation of a mutant *S* locus receptor kinase. *Plant Cell* **10,** 209–218 (1998).
16. Hatakeyama, K., Takasaki, T., Watanabe, M. & Hinata, K. Molecular characterization of *S* locus genes, *SLG* and *SRK*, in a pollen-recessive self-incompatibility haplotype of *Brassica rapa* L. *Genetics* **149,** 1587–1597 (1998).
17. Hatakeyama, K., Watanabe, M. Takasaki, T., Ojima, K. & Hinata, K. Dominance relationships between *S* alleles in self-incompatible *Brassica campestris* L. *Heredity* **80,** 241–247 (1998).
18. Watanabe, M. *et al.* A high degree of homology exists between the protein encoded by *SLG* and the *S* receptor domain encoded by *SRK* in self-incompatible *Brassica campestris* L. *Plant Cell Physiol.* **35,** 1221–1229 (1994).
19. Suzuki, G., Watanabe, M., Toriyama, K., Isogai, A. & Hinata, K. Expression of SLG^9 and SRK^9 genes in transgenic tobacco. *Plant Cell Physiol.* **37**, 866–869 (1996).
20. Kish-Nishizawa, N. *et al.* Ultrastructure of papillar cells in *Brassica campestris* revealed by liquid helium rapid-freezing and substitute-fixation method. *Plant Cell Physiol.* **31,** 1207–1219 (1990).
21. Stone, S. L., Arnoldo, M. & Goring, D. R. A breakdown of *Brassica* self-incompatibility in ARC1 antisense transgenic plants. *Science* **286,** 1729–1731 (1999).
22. Suzuki, G., Watanabe, M., Toriyama, K., Isogai, A. & Hinata, K. Molecular cloning of members of the *S*-multigene family in self-incompatible *Brassica campestris*. *Plant Cell Physiol.* **36,** 1273–1280 (1995).
23. Jones, J. G. *et al.* Effective vectors for transformation, expression of heterologous, and assaying transposon excision in transgenic plant. *Transgenic Res.* **1,** 285–297 (1992).
24. Takasaki, T. *et al.* Factors influencing *Agrobacterium*-mediated transformation of *Brassica rapa* L. *Breed. Sci.* **47,** 127–134 (1997).
25. Kho, Y. O. & Bear, J. Observing pollen tubes by means of fluorescence. *Euphytica* **17,** 298–302 (1968).
26. Manolson, M. F., Ouellette, B. F., Filion, M. & Poole, R. J. cDNA sequence and homologies of the '57-Kda' nucleotide-binding subunit of the vacuolar ATPase from *Arabidopsis*. *J. Biol. Chem.* **263,** 17987–17994 (1988).
27. Nishio, T., Kusaba, M., Watanabe, M. & Hinata, K. Registration of *S* alleles in *Brassica campestris* L. by the restriction fragment sizes of *SLGs*. *Theor. Appl. Genet.* **92,** 388–394 (1996).
28. Nishio, T., Kusaba, M., Sakamoto, K. & Ockendon, D. Polymorphism of the kinase domain of the *S*-locus receptor kinase gene (*SRK*) in *Brassica oleracea* L. *Theor. Appl. Genet.* **95,** 335–342 (1997).
29. Brace, J., Ockendon, D. J. & King, G. J. Development of a method for identification of *S* alleles in *Brassica oleracea* based on digestion of PCR-amplified DNA with restriction endonucleases. *Sex. Plant Reprod.* **6,** 133–138 (1993).
30. Shiba, H., Hinata, K., Suzuki, A. & Isogai, A. Breakdown of self-incompatibility in *Brassica* by the antisense RNA of the *SLG* gene. *Proc. Jpn. Acad. B* **71,** 81–83 (1995).

Acknowledgements

We thank T.-h. Kao for his critical review and editing of the manuscript. This work was supported in part by Grants-in-Aid for Special research on Priority Areas from the Ministry of Education, Science, Culture and Sports, Japan.

Correspondence and requests for materials should be addressed to T.T. (e-mail: taka@kobe-u.ac.jp).

Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding

Ketan S. Gajiwala*, Hua Chen*†, Fabrice Cornille‡, Bernard P. Roques‡, Walter Reith§, Bernard Mach§ & Stephen K. Burley*†

**Laboratories of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, † Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA*
‡ Département de Pharmacologie Moléculaire et Structurale, INSERM U266, CNRS URA D1500, UFR des Sciences Pharmaceutiques et Bioligiques, 4 Avenue de l'Observatoire, 75720 Paris Cedex 06, France
§ Département de Génétique et Microbiologie, Centre Médical Universitaire (CMU), 1, rue Michel-Servet, CH 1211 Genéve, 4 Switzerland

Regulatory factor X (RFX) proteins are transcriptional activators that recognize X-boxes (DNA of the sequence 5′-GTNRCC(0–3N)RGYAAC-3′, where N is any nucleotide, R is a purine and Y is a pyrimidine) using a highly conserved 76-residue DNA-binding domain (DBD). DNA-binding defects in the protein RFX5 cause bare lymphocyte syndrome or major histocompatibility antigen class II deficiency[1]. RFX1, -2 and -3 regulate expression of other medically important gene products (for example, interleukin-5 receptor α chain, IL-5Rα)[2]. Fusions of the ligand-binding domain of the oestrogen receptor with the DBD of RFX4 occur in some human breast tumours[3]. Here we present a 1.5 Å-resolution structure of two copies of the DBD of human RFX1 (hRFX1) binding cooperatively to a symmetrical X-box[4,5]. hRFX1 is an unusual member of the winged-helix subfamily of helix–turn–helix proteins[6] because it uses a β-hairpin (or wing) to recognize DNA instead of the recognition helix typical of helix–turn–helix proteins. A new model for interactions between linker histones and DNA is proposed.

We used multiwavelength anomalous dispersion (MAD) to determine the structure of the hRFX1 DBD (Fig. 1) recognizing a symmetrical X-box (see Methods). Unexpectedly, the hRFX1 DBD proved to be a member of the winged-helix hepatocyte nuclear factor (HNF)-3/*forkhead*-related subfamily of HTH transcription factors (Fig. 2, reviewed in ref. 7). The DBD consists of three α-helices (H), three β-strands (S) and three connecting loops (L), arranged in the order H1-S1-H2-L1-H3-L2-S2-W1-S3. The third loop, connecting β-strands S2 and S3, forms wing W1 of the winged-helix motif. Residues forming the hydrophobic core of the DBD originate from the secondary structural elements H1, S1, H2,

H3, L2 and S3 (Fig. 1). Fifteen out of nineteen of these amino acids are either identical or conserved among RFX proteins, indicating that all five distinct DBDs are likely to share the same structure. In contrast to previously described winged-helix DBDs (typically 110 residues long with two wings, W1 and W2)[6], the shorter RFX DBD has only one wing (Fig. 2). A structure-based sequence alignment of the hRFX1 DBD with HNF-3γ revealed 8% identity (Fig. 2b), which explains why we could not discern their similarity. As deduced from

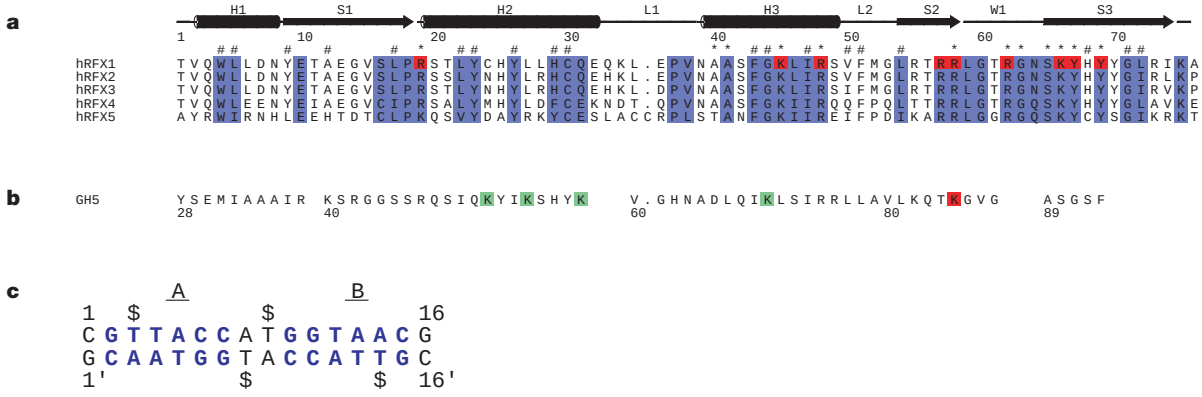

Figure 1 Protein and oligonucleotide sequences. **a**, DBDs of five RFX proteins with the hRFX1 DBD secondary structure; pairwise identities range between 36 and 90%. Violet denotes conserved amino acids; red identifies hRFX1 residues making direct or water-mediated DNA contacts. Residues marked with hash signs comprise the hydrophobic core, and side chains buried on DNA binding are indicated with asterisks. Residue numbering is based on the hRFX1 DBD with Thr 1 corresponding to Thr 438 in full-length hRFX1. **b**, Structure-based sequence alignment of GH5 with the hRFX1 DBD. Green coded GH5 residues are protected against chemical modification in chromatin, with Lys 85 (red) being maximally protected. **c**, Oligonucleotide used for crystallization with numbering scheme, X-box half-site labelling and 5-bromouracil positions (dollar signs). Purple lettering denotes half-sites.

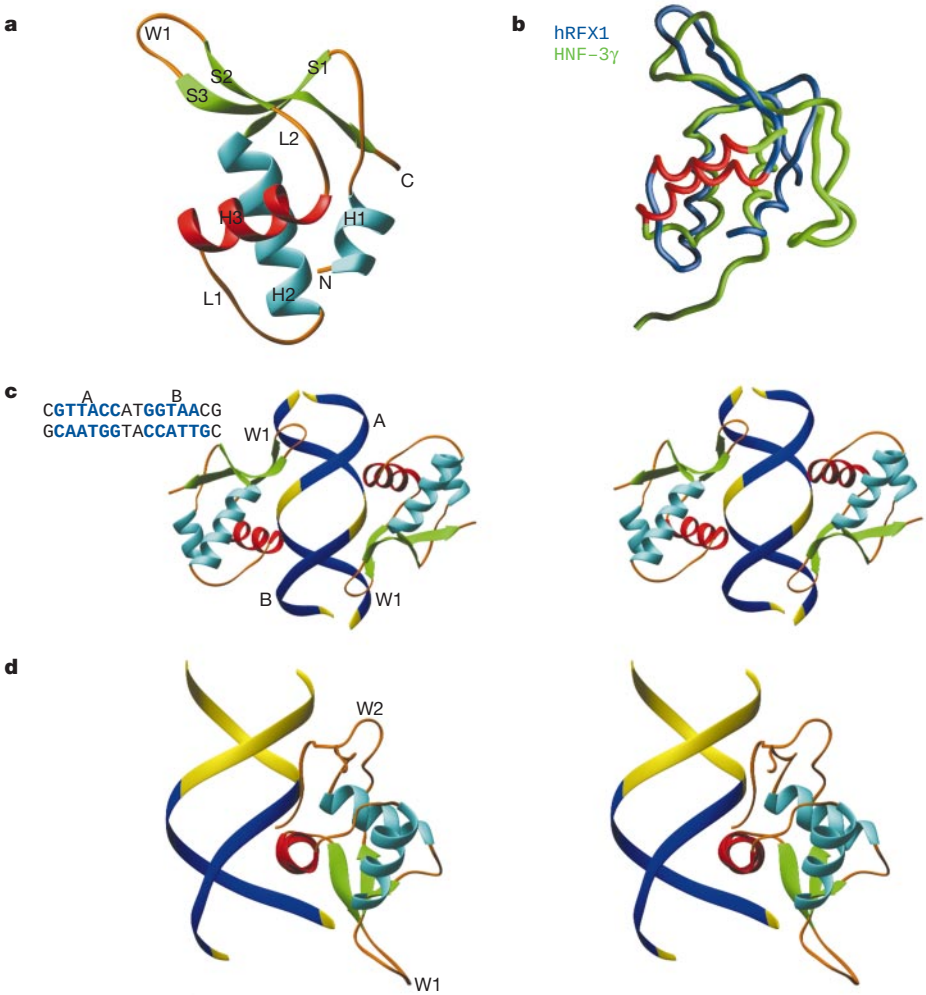

Figure 2 hRFX1 and HNF-3γ DBDs. **a**, hRFX1 DBD with labelled N and C termini and secondary structural elements (HTH recognition helix H3, red). **b**, Superposition of hRFX1 and HNF-3γ DBDs (r.m.s. deviation for α-carbon atoms, 3.5 Å), shown in the same orientation as **a**. **c**, Stereo diagram of the hRFX1–DNA 2:1 complex, viewed along the 2-fold crystallographic symmetry axis relating each half of the asymmetric unit. Inset, crystallization oligonucleotide, with blue characters denoting X-box half-sites, labelled A and B. **d**, Stereo diagram of HNF-3γ bound to its cognate DNA[6].

previous biochemical studies[5], two RFX1 DBDs bind a single X-box. The resulting 2:1 protein–DNA complex is perfectly symmetrical, with a crystallographic two-fold axis passing through the centre of the DNA (Fig. 2c).

The hRFX1 consensus binding site is an imperfect inverted repeat with variable spacing between two half-sites (5′-NNGTNRCC/n(0–3N)RGYAACNN-3′; underlining denotes X-box half-site)[8]. The palindromic 16-base pair (bp) duplex oligonucleotide used here (5′-CGTTACCATGGTAACG-3′) has a 2-bp spacer (5′-AT-3′). This X-box closely resembles the physiological hRFX1 binding site (5′-GTTGCCCGGCAAC-3′) in the hepatitis B virus (HBV) enhancer[9]. Each half-site of the symmetric X-box interacts with both DBDs. Protein–DNA interactions are shown in Figs 2 and 3. Among previously reported structures of *bona fide* winged-helix proteins (that is, those sharing the HNF-3γ fold), four have been determined with cognate B-DNAs (HNF-3γ [6], genesis[10], E2F4 and DP2 (ref. 11)). In the HNF-3γ co-crystal structure, H3 (that is, the recognition helix; Fig. 2d, red) makes specific contacts with the major groove[6]. W1 of HNF-3γ overlies the minor groove of a neighbouring oligonucleotide in the crystal lattice, and the carboxy-terminal wing (W2) makes one minor-groove contact (Fig. 2d). This mode of DNA binding is shared by the transcription factors genesis, E2F4 and DP2, and is common to all fully characterized HTH proteins except hRFX1.

The hRFX1 DBD uses an alternative mechanism of DNA recognition. β-strands S2 and S3 and their connecting loop, wing W1, make extensive contacts with the major groove of one half-site of the symmetric X-box (Figs 2c and 3). A single side chain from α-helix H3 (red in Figs 2c and 3) interacts with the minor groove of the other half-site. X-box binding by hRFX1 is mediated primarily by polar side chains, through direct contacts or via interfacial water molecules (Fig. 3). This unprecedented protein–DNA complex is not a crystal packing artefact. The total solvent-accessible surface area buried per DBD is 1,800 Å^2, which far exceeds values characteristic of adventitious lattice contacts and is compatible with a specific protein–DNA complex (reviewed in ref. 12). The observed side-chain–base contacts (Fig. 3) are entirely consistent with hRFX1 binding *in vivo* and *in vitro*.

Binding of hRFX1 to the Py site in the polyoma virus genome is abolished by changing G11 to any other base[8]. This finding can be explained by our structure (Fig. 3a), which contains two hydrogen bonds between Arg 58 and G11 (NH1–O6 = 2.7 Å, NH2–N7 = 3.1 Å). Replacement of G11 with A would eliminate the NH1–O6 hydrogen bond, and both hydrogen bonds would probably be impossible following pyrimidine substitution. Our explanation is supported by the results of electrophoretic gel mobility shift assays (EMSA). When G11 and G10 are changed to T simultaneously, eliminating three hydrogen bonds (G10 interacts with Tyr 67, OH–N7 = 2.7 Å, not depicted), DNA binding is abolished (data not shown). A similar effect is observed when G15′ is substituted with T, presumably because hydrogen bonds with Arg 62 (Fig. 3b, NE–N7 = 3.1 Å, NH2–O6 = 2.7 Å) can no longer form (data not shown). Additional, albeit indirect, support for the physiological relevance of our co-crystal structure comes from studies of RFX proteins

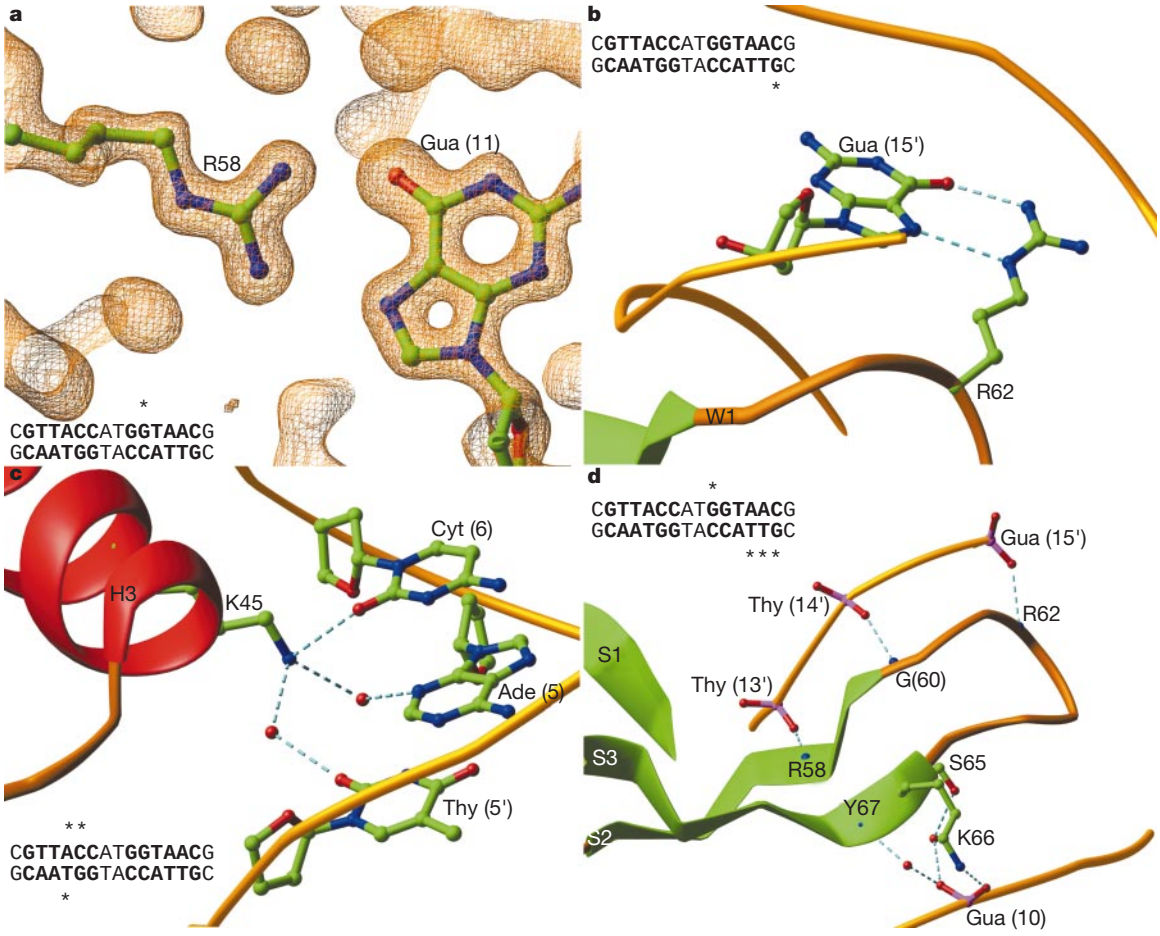

Figure 3 Protein–DNA interactions. **a**, Representative 1.5 Å resolution electron-density map contoured at 2σ, showing Arg 82 bound to G11 in the major groove of the B half-site. The crystallization oligonucleotide is inset, with bold denoting each half-site, and an asterisk identifying the depicted nucleotide(s). **b**, Arg 62 bound to G15′ in the major groove of the B halfsite. **c**, Lys 45 from H3 (red) interacts with the minor groove face of the A half-site. Red spheres denote water molecules. **d**, Direct and water-mediated protein–DNA contacts made by W1 with the phospho-deoxyribose backbones, flanking the major groove of the B half-site.

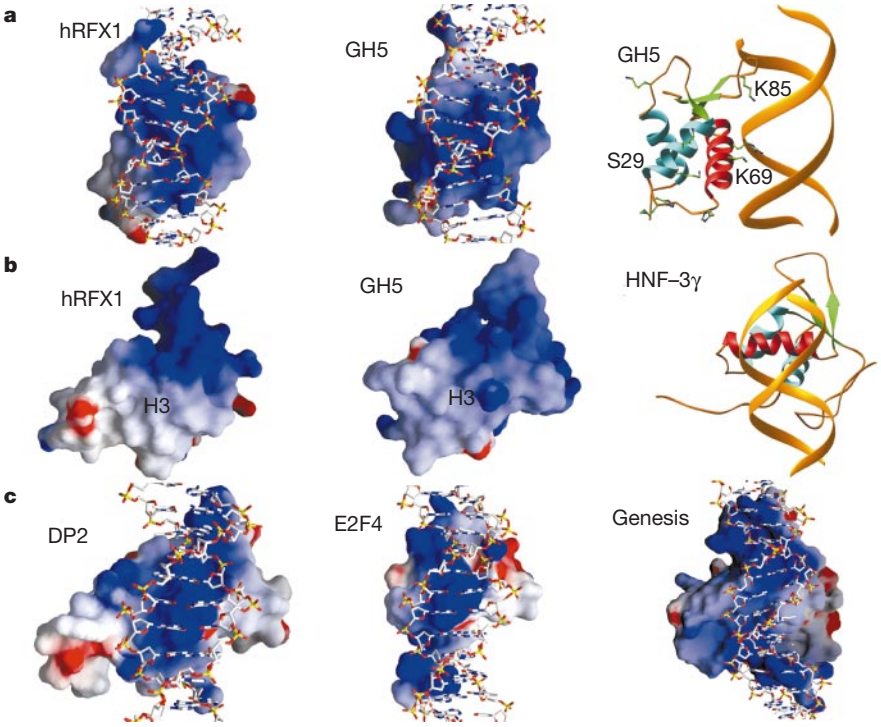

Figure 4 Model for linker histone–DNA interactions. GRASP[25] representations of the chemical properties of the solvent-accessible DBD surface, calculated using a water probe radius of 1.4 Å. The surface electrostatic potential is coloured red and blue, representing electrostatic potentials <–20 to >+20k_BT, where k_B is the Boltzmann constant and T is the temperature. The calculations were performed with an ionic strength of 0 and dielectric constants of 80 and 2 for solvent and protein, respectively. DNA is denoted as an atomic stick figure or a pair of yellow ribbons denoting phospho-deoxyribose backbones. **a**, hRFX1 DBD–DNA complex (left), and our alternative model of GH5–DNA complex (middle, surface electrostatic potential; right, ribbon diagram). Depicted GH5 side chains (right) are thought to be important for DNA packaging. Label denoting Ser 29 shows its approximate, partially obscured position. **b**, H3 faces colour coded for electrostatic potential (left: hRFX1 DBD; middle: GH5). The HNF-3γ–DNA ribbon drawing (right) shows the horizontal orientation of H3 common to all views in **b** and **c**. **c**, DP2, E2F4 and genesis bound to their cognate DNAs, with H3 faces colour-coded for electrostatic potential.

binding to X-boxes containing methylated 5′-CpG-3′ dinucleotides (consensus sequence for RFX binding to methylated sites is 5′-N RTMRYYAMRGMRAYN-3′, where R is a purine, Y is a pyrimidine and M is 5-methyl-Cyt or T[13]). Methyl groups of T4 and T6, respectively, make van der Waals contacts with Arg 58 (Me–CB = 4 Å, not shown) and Tyr 67 (Me–CZ = 3.6 Å, not shown).

Finally, studies of the IL-5Rα promoter show that our findings with RFX1 can be generalized to other members of the RFX family. Expression of IL-5Rα is regulated by RFX1, -2 and -3 through homomeric and heteromeric interactions with a unique enhancer-like *cis* element, which resembles the consensus RFX binding site (5′-NNGTNRCC/n(0-3N)RGYAACNN-3′)[2]. Mutations of three guanines within the IL-5Rα promoter completely abolish binding of RFX1, -2, and -3 (ref. 2). These deleterious substitutions involve nucleotides making major groove hydrogen bonds with Arg 58, Arg 62 and Tyr 67 in our RFX1 co-crystal structure. It is, therefore, likely that RFX2 and -3 bind to the IL-5Rα X-box using the same mechanism as RFX1, because all nine of the residues involved in DNA contacts are identical among these RFX family members.

Our structure of the hRFX1–DNA complex reveals no protein–protein interactions between the two DBDs comprising the 2:1 protein–DNA complex (Fig. 2c), which readily explains how RFX1, -2 and -3 can form heterocomplexes with the IL-5Rα X-box. EMSA has revealed cooperative hRFX1 DBD binding to an X-box[5]. These observations indicate that DNA-binding cooperativity may arise from a through-DNA effect. Geometric analysis of X-box DNA conformational parameters[14] reveals that the geometry of each X-box half-site is distorted when bound to the hRFX1 DBD. The

Table 1 Statistics of the Crystallographic Analysis

Data set	Resolution (Å)	Reflections, measured/unique	Completeness (%), overall/outer shell	R_{sym} (%), overall/outer shell
λ_1 (0.9201 Å)	20.0–1.85	112,611/12,998	99.3/93.7	4.0/16.4
λ_2 (0.9197 Å)	20.0–1.85	112,848/13,014	99.3/93.7	4.2/18.0
λ_3 (0.9062 Å)	20.0–1.85	114,270/12,998	99.8/98.3	4.2/18.8
Overall MAD figure of merit, 0.61				
Refinement statistics	Resolution (Å)	R_{sym}/completeness (%)	*R* factor, overall/outer shell	Free *R* factor
Data	10.0–1.5	5.2/96.3	0.192/0.216	0.227
r.m.s. deviations	bond lengths, 0.016 Å	bond angles, 1.8°	thermal parameters, 1.3 Å^2	

$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is observed intensity and $\langle I\rangle$ is average intensity obtained from multiple observations of symmetry related reflections.
r.m.s. bond lengths and r.m.s. bond angles are the respective root-mean-square deviations from ideal values. r.m.s. Thermal parameter is the root-mean-square deviation between the *B* values of covalently bonded atomic pairs. Free *R* factor was calculated with 10% of the data omitted from the structure refinement.

minor groove is widened (by >3 Å) where H3 inserts Lys 45 to make a hydrogen bond with C6 (Fig. 3c). On the opposite face of the double helix, where the other DBD presents W1 to the major groove, narrowing is observed (by 1 Å). Hence major groove recognition of an X-box half-site by one DBD yields a narrowed major groove conformation, which in turn could stabilize binding of the second DBD because minor groove widening would facilitate insertion of Lys 45 (or vice versa).

During cooperative binding of two RFX DBDs to an X-box, reciprocal interactions would occur between both DBDs and the A and B half-sites (Figs 1c and 2c). An alternative explanation of DNA-binding cooperativity is provided by the presence of overlapping protein–DNA contacts. In the A half-site, Ser 42 of one DBD makes a water-mediated, major groove contact with G6′ (not shown), whereas Lys 45 of the other DBD in the 2:1 complex makes a direct minor groove hydrogen bond with C6 on the opposite strand (Fig. 3c). Corresponding contacts are seen in the B half-site with G11 and C11′. A comparable overlap has been used to explain cooperativity in the Oct1 POU domain–DNA complex[15]. It is not possible to distinguish the relative contributions of these two effects in the absence of an exhaustive analysis of hRFX1 DNA-binding cooperativity.

An X-ray structure of the globular portion of histone H5 (GH5) was determined in the absence of DNA[16]. Recognizing that GH5 was a HTH protein, biochemical data[17,18] allowed the authors[16] to suggest that H3 of GH5 functions as a recognition helix, which would interact with the major groove of DNA in chromatin. Our hRFX1 co-crystal structure reveals a second mechanism by which the linker histones H1 and H5 may bind DNA. Using a structure-based sequence alignment[19], GH5 was superimposed on the hRFX1 DBD–DNA complex (Fig. 4a, left), yielding the structural model for linker histones binding to DNA depicted in Fig. 4a (middle and right). This alternative W1/major-groove binding model is consistent with available biochemical data. Lys 85 in GH5 (Arg 58 in hRFX1 DBD, Fig. 1b) is essential for DNA binding and is protected from reductive methylation in chromatin[17]. Mutation of Lys 85 to either Glu or Gln renders GH5 unable to protect nucleosomal DNA from enzymatic digestion[18]. In the model shown in Fig. 4a (right), GH5 of Lys 85 can hydrogen bond to G11, which represents an important contact site for the hRFX1 DBD (Fig. 3a). Another basic amino acid in GH5 (Lys 69, Fig. 4a) has been implicated in DNA binding[17]. The corresponding residue in the hRFX1 DBD (Lys 45, Fig. 3c) extends from H3 into the minor groove of the X-box and interacts with C6. Finally, photoaffinity-labelling studies place His 62 of GH5 close to DNA[20], which is also compatible with the model shown in Fig. 4a. These indirect biochemical observations do not allow us to choose between the H3/major groove and W1/major groove models for linker histone–DNA interactions, because H3 and W1 interact with DNA in both the HNF-3γ and the hRFX1 co-crystal structures.

The electrostatic properties of the winged-helix proteins provide a means of resolving the issue. Examination of solvent-accessible surfaces of the hRFX1 DBD and GH5 reveals that the W1 face of GH5 (that is, the DNA-binding surface of hRFX1) is significantly more basic than its H3 face (compare Fig. 4a and 4b, middle). The same is true for the hRFX1 DBD (compare Fig. 4a and 4b, left). (The basic patch immediately adjacent to the hRFX1 label in Fig. 4b is located in the background and does not sit in the same plane as the hydrophobic face of H3.) In contrast, the H3 faces of HNF-3γ, DP2, E2F4 and genesis (Fig. 4c) are significantly more basic than their respective W1 faces (not depicted). Winged-helix proteins can, therefore, be divided into two functional classes on the basis of their electrostatic properties. HNF-3γ, DP2, E2F4 and genesis bind DNA with basic H3 faces, placing their recognition helices in the major groove. hRFX1 uses a basic W1 face for DNA binding. We believe that the striking similarity of the electrostatic properties of the hRFX1 DBD and GH5 provides compelling support for the model of DNA binding by linker histones shown in Fig. 4a (middle, right). This model is not incompatible with the fact that a murine HNF-3 can functionally substitute for the linker histone H1 in the mouse prealbumin gene enhancer[21], because we do not know how HNF-3α interacts with DNA in the presence of histones H2a, H2b, H3 and H4, and other transcription factors. Moreover, the model does not preclude another portion of GH5 acting as a secondary DNA-binding site, as characterized in ref. 20. Ser 29 of GH5 (Fig. 4a, right) participates in DNA binding[22], and may well represent part of this secondary nucleic acid binding surface.

In conclusion, our 1.5 Å resolution co-crystal structure of the hRFX1 DBD reveals an unprecedented mode of DNA recognition for an HTH protein. This result provides insight into DNA binding by linker histones, and demonstrates the functional versatility of the winged-helix DNA-binding motif. Finally, this structure provides a framework for further studies of the bare lymphocyte syndrome and other medically important roles of the RFX proteins. □

Methods

X-ray crystallography

Synthetic hRFX1 DBD[5] bearing two substitutions, Cys 24 Asn and Cys 30 Ser, was crystallized with a synthetic brominated 16-bp oligonucleotide (Fig. 1c). Co-crystals (space group *C*2: a = 72.8 Å, b = 42.8 Å, c = 52.3 Å, β = 109.2°) with one DBD plus one DNA strand per asymmetric unit, were grown by hanging drop vapour diffusion at 4 °C against 20% PEG-400, 80 mM magnesium acetate, 50 mM cacodylate, pH 6.5. MAD data were collected under cryogenic conditions using CHESS beamline F2, and the refinement data were collected using Brookhaven NSLS beamline X9B (Table 1).

Structure determination and refinement

Both bromines were located using anomalous difference Patterson syntheses. MAD phases at 1.85 Å resolution yielded an interpretable experimental electron-density map. After model building, the structure was refined to convergence at 1.5 Å resolution using CNS[23]. The final refinement model includes hRFX1 DBD residues 1–76, the crystallization DNA, 124 water molecules, 2 molecules of ethylene glycol and 1 molecule of diethylene glycol (Table 1). PROCHECK[24] revealed no unfavourable (ϕ,ψ) combinations in the hRFX1 DBD, with main-chain and side-chain structural parameters consistently better than average (overall G value 0.1).

Received 6 September 1999; accepted 5 January 2000.

1. Mach, B., Steimle, V., Martinez-Soria, E. & Reith, W. Regulation of MHC class II genes: lessons from a disease. *Annu. Rev. Immunol.* **14,** 301–331 (1996).
2. Iwama, A. *et al.* Dimeric RFX proteins contribute to the activity and lineage specificity of the interleukin-5 receptor alpha promoter through activation and repression domains. *Mol. Cell Biol.* **19,** 3940–3950 (1999).
3. Dotzlaw, H., Alkhalaf, M. & Murphy, L. C. Characterization of estrogen receptor variant mRNAs from human breast cancers. *Mol. Endocrinol.* **6,** 773–785 (1992).
4. Reith, W. *et al.* MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. *Genes Dev.* **4,** 1528–1540 (1990).
5. Cornille, F. *et al.* DNA binding properties of a chemically synthesized DNA binding domain of hRFX1. *Nucleic Acids Res.* **26,** 2143–2149 (1998).
6. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364,** 412–420 (1993).
7. Lai, E., Clark, K. L., Burley, S. & James E. Darnell, J. Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: A family of transcription factors of diverse biological function. *Proc. Natl Acad. Sci. USA* **90,** 10421–10423 (1993).
8. Emery, P. *et al.* A consensus motif in the RFX DNA binding domain and binding domain mutants with altered specificity. *Mol. Cell Biol.* **16,** 4486–4494 (1996).
9. Ostapchuk, P., Scheirle, G. & Hearing, P. Binding of nuclear factor EF-C to a functional domain of the hepatitis B virus enhancer region. *Mol. Cell Biol.* **9,** 2787–2797 (1989).
10. Jin, C., Marsden, I., Chen, X. & Liao, X. Dynamic DNA contacts observed in the NMR structure of winged helix protein–DNA complex. *J. Mol. Biol.* **289,** 683–690 (1999).
11. Zheng, N., Fraenkel, E., Pabo, C. O. & Pavletich, N. P. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.* **13,** 666–674 (1999).
12. Janin, J. Principles of protein–protein recognition from structure to thermodynamics. *Biochimie* **77,** 497–505 (1995).
13. Reith, W. *et al.* RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. *Mol. Cell Biol.* **14,** 1230–1244 (1994).
14. Lavery, R. & Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynamics* **6,** 63–91 (1988).
15. Klemm, J. D. & Pabo, C. O. Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev.* **10,** 27–36 (1996).
16. Ramakrishnan, V., Finch, J., Graziano, V. & Sweet, R. Crystal structure of the globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362,** 219–223 (1993).
17. Thomas, J. O. & Wilson, C. M. Selective radiolabelling and identification of a strong nucleosome binding site on the globular domain of histone H5. *EMBO J.* **5,** 3531–3537 (1986).
18. Buckle, R. S., Maman, J. D. & Allan, J. Site-directed mutagenesis studies on the binding of the globular domain of linker histone H5 to the nucleosome. *J. Mol. Biol.* **223,** 651–659 (1992).

19. Holm, L. & Sander, C. Families of structurally similar proteins, version 1. 0. *J. Mol. Biol.* **233,** 123–138 (1993).
20. Goytisolo, F. A. *et al.* Identification of two DNA-binding sites on the globular domain of histone H5. *EMBO J.* **15,** 3421–3429 (1996).
21. Cirillo, L. A. *et al.* Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* **17,** 244–254 (1998).
22. Zhou, Y. B., Gerchman, S. E., Ramakrishnan, V., Travers, A. & Muyldermans, S. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* **395,** 402–405 (1998).
23. Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921 (1998).
24. Laskowski, R. J., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Cryst.* **26,** 283–290 (1993).
25. Nicholls, A., Sharp, K. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11,** 281–296 (1991).

Acknowledgements

We thank D. Thiel and the MacCHESS staff, and Z. Dauter and K.R. Rajashankar and the NSLS X9B staff for help with X-ray data collection. We thank C. Lenoir for assistance in peptide synthesis, and J. Bonanno, K. L. Clark, J. E. Darnell, R. C. Deo, D. Jeruzalmi, J. Kuriyan, E. Lai, H. A. Lewis, S. K. Nair and G. A. Petsko for many useful discussions. S.K.B. is an Investigator in the Howard Hughes Medical Institute. This work was supported by the National Institute of General Medical Sciences.

Correspondence and requests for materials should be addressed to S.K.B. (e-mail: burley@rockvax.rockefeller.edu). Atomic coordinates have been deposited to the Protein Data Bank under accession number 1dp7.

Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin

Kelvin Chu*†, Jaroslav Vojtchovský‡, Benjamin H. McMahon§, Robert M. Sweet‖, Joel Berendzen† & Ilme Schlichting‡

**P-21 Biophysics Group, MS-D454, Los Alamos National Laboratory, Los Alamos, New Mexico, 87545, USA*
† Department of Physics, Cook Building, University of Vermont, Burlington, Vermont 05405, USA.
‡ Max Planck Institute for Molecular Physiology, Department of Biophysical Chemistry, Otto-Hahn-Str 11, 44227 Dortmund, Germany
§ Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, 87545, USA
‖ Department of Biology, Brookhaven National Laboratory, Upton, New York 11973-5000, USA

Small molecules such as NO, O_2, CO or H_2 are important biological ligands that bind to metalloproteins to function crucially in processes such as signal transduction, respiration and catalysis. A key issue for understanding the regulation of reaction mechanisms in these systems is whether ligands gain access to the binding sites through specific channels and docking sites, or by random diffusion through the protein matrix. A model system for studying this issue is myoglobin, a simple haem protein. Myoglobin has been studied extensively by spectroscopy, crystallography, computation and theory[1–11]. It serves as an aid to oxygen diffusion but also binds carbon monoxide, a byproduct of endogenous haem catabolism. Molecular dynamics simulations[3–5], random mutagenesis[6] and flash photolysis studies[7–10] indicate that ligand migration occurs through a limited number of pathways involving docking sites. Here we report the 1.4 Å resolution crystal structure of a ligand-binding intermediate in carbonmonoxy myoglobin that may have far-reaching implications for understanding the dynamics of ligand binding and catalysis.

The ligand-binding reaction in myoglobin (Mb) has been studied by a variety of techniques[1–14]. Time-resolved and kinetic studies over wide ranges in time and temperature indicate that ligand binding is not a simple process, but consists of a series of sequential reactions as the ligand moves from the surrounding solvent through the protein to the binding site at the haem iron.

A		**B**		**D**		**S**
MbCO (bound)	⇔	Mb*CO (photolysed)	⇔	Mb**CO (photorelaxed)	⇔	Mb + CO (deoxy)

Initially the system has CO bound to the haem iron to form carbonmonoxy myoglobin (**A**, MbCO). After either thermal- or photodissociation, the CO migrates inside the haem pocket to a primary 'docking site' (**B**, Mb*CO). At room temperature, mid-infrared bands corresponding to state-**B** formation are generated within picoseconds after photolysis of MbCO[11]; as time evolves, the bands disappear either by geminate rebinding (**B** →**A**)[10] or by relaxation to a new state. This second intermediate, which has been observed by for sperm whale myoglobin (swMb)[11], is also geminate and builds up transiently at room temperature. The secondary docking site can also be populated at low temperature by illumination with intense laser light and is denoted photorelaxed state (**D**, Mb**CO)[10]. The two primary ligand-binding intermediates, states **B** and **D**, have median binding enthalpies of 10 and 30 kJ mol^{-1}, respectively[10]. At high temperatures and after long times, the ligand escapes the protein matrix to the solvent state (**S**, Mb + CO, deoxy).

Whether the second, spectroscopically distinct intermediate state is caused by a new docking site or by conformational changes induced in the protein has been the subject of much discussion[7–14]. A detailed analysis of recombination kinetics reveals that the ligand must be located elsewhere than the distal pocket in state **D** and suggests that proximal cavities are implicated in the rebinding pathway[9]. To address this issue, we determined the crystal structure of Mb**CO and compared it with that of Mb*CO and MbCO.

We chose horse heart MbCO (hhMbCO) for our study because there is a single ligand-bound conformational substate at neutral pH (ref. 10) whereas sperm whale MbCO (swMbCO) has three taxonomic substates (A_0, A_1, A_3)[2,11,15] that potentially complicate the analysis. Because of differences in kinetics between the two species, we also determined the structures of hhMbCO and hhMb*CO (Table 1; Fig. 1). In hhMbCO, the ligand is bound almost perpendicular to the haem plane and the haem iron is in plane. In hhMb*CO, the photodissociated CO is located parallel to the haem plane on top of the carbon pyrole at a distance of 3.6 Å from the haem iron. The iron has moved by ~0.25 Å out of the mean haem plane, as described previously for swMb*CO (refs 16, 17). A CO orientation that has the carbon atom closer to the iron was preferred in the refinement. Neither the distal nor the proximal histidine move significantly upon photolysis, as described previously[17,18]. Spectroscopic measurements of the photolysed state[11,12] reveal two bands that correspond to CO in different orientations in the distal pocket, one of which converts into the other with a low enthalpic barrier.

In hhMb**CO, the iron is ~0.23 Å out of the mean haem plane (Fig. 2). There is no electron density at the positions of the bound or photolysed CO in the data set collected under illumination (some electron density for bound CO was observed in the data set collected without illumination during data acquisition; otherwise the two complexes are identical), but a new roundish peak appears below the haem plane (see Fig. 1b, c). It was modelled with a CO molecule having 30% occupancy, on the basis of results from equivalent spectroscopic studies carried out in solution. The round electron density assigned to the CO position is in agreement with the observation that the infrared band of state **D** is broader than that of the state **B** (refs [10,11,19]) and the ligand is probably not oriented within the binding site. The ligand density is sandwiched between the side chains of the proximal His 93 and Phe 138, and surrounded