# Design and Characterisation of an Artificial DNA-Binding Cytochrome

D. Dafydd Jones*[a, b] and Paul D. Barker[a]

*We aim to design novel proteins that link specific biochemical binding events, such as DNA recognition, with electron transfer functionality. We want these proteins to form the basis of new molecules that can be used for templated assembly of conducting cofactors or for thermodynamically linking DNA binding with cofactor chemistry for nanodevice applications. The first examples of our new proteins recruit the DNA-binding basic helix region of the leucine zipper protein GCN4. This basic helix region was attached to the N and C termini of cytochrome $b_{562}$ (cyt $b_{562}$) to produce new, monomeric, multifunctional polypeptides. We have fully characterised the DNA and haem-binding properties of these proteins, which is a prerequisite for future application of the new molecules. Attachment of a single basic helix of GCN4 to either the N or C terminus of the cytochrome does not result in specific DNA binding but the presence of DNA-binding domains at both termini converts the cytochrome into a specific DNA-binding protein. Upon binding haem, this chimeric protein retains the spectral characteristics of wild-type cyt $b_{562}$. The three forms of the protein, apo, oxidised holo and reduced holo, all bind the designed (ATGAcgATGA) target DNA sequence with a dissociation constant, $K_D$, of approximately 90 nM. The protein has a lower affinity ($K_D$ ca. 370 nM) for the wild-type GCN4 recognition sequence (ATGAcTCAT). The presence of only half the consensus DNA sequence (ATGAcgGGCC) shifts the $K_D$ value to more than 2500 nM and the chimera does not bind specifically to DNA sequences with no target recognition sites. Ultracentrifugation revealed that the holoprotein–DNA complex is formed with a 1:1 stoichiometry, which indicates that a higher-order protein aggregate is not responsible for DNA binding. Mutagenesis of a loop linking helices 2 and 3 of the cytochrome results in a chimera with a haem-dependent DNA binding affinity. This is the first demonstration that binding of a haem group to a designed monomeric protein can allosterically modulate the DNA binding affinity.*

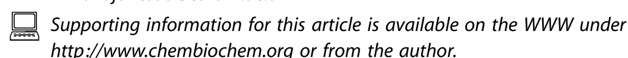
## Introduction

Current interest in self-assembling nanodevices suggests that it would be useful to link biological activity to an electron transfer reaction within a protein device. The resulting proteins could be designed to self-assemble at specific locations upon well-defined macromolecular scaffolds that would form the basis of current-carrying molecular devices and ultimately act as transistors in which the gating voltage is replaced by a biochemical process. How proteins may ultimately be used in electronic devices is a matter for discussion, but the ability to design and construct proteins with new functions is now becoming a reality[1–3] and it is useful to direct some protein engineering towards these goals. Although it might seem simplest to use the biological function of natural redox proteins out of context in electronic devices, natural proteins have evolved in very different environments to those ultimately required for electronic device construction. We therefore believe that proteins must be radically redesigned or new scaffolds created for use in electronics, although natural proteins will clearly provide inspiration and guidance for construction of molecules with novel electronic functions. Ligand binding has been exploited as a method for controlling protein conformation and hence function.[1,4–6] Herein we describe the design and construction of a chimeric protein that marries two disparate functions, electron transfer and DNA binding, by combining two separate proteins in a single unit capable of carrying out both the original functions.

Cytochrome $b_{562}$ (cyt $b_{562}$) is a haem-binding four-helix-bundle protein (Figure 1) found in the periplasm of *Escherichia coli* and has been the subject of extensive structural, folding and protein engineering studies. The haem group is noncovalently attached to the protein through two ligands, a methionine residue (Met7) towards the N terminus and a histidine residue (His102) towards the C terminus, both of which coordinate the haem iron moiety.[7,8] Upon cofactor binding, major structural and dynamic changes occur in the polypeptide,[9,10] particularly in the fourth helix, which changes from a dynamic structure into a unique fully-formed helix locked in its conformation by the coordination of His102 to the haem iron moiety.

[a] Dr. D. D. Jones, Dr. P. D. Barker
University Chemical Laboratories and MRC Centre for Protein Engineering
University of Cambridge
Lensfield Road, Cambridge, CB2 1EW (UK)
E-mail: pdb30@cam.ac.uk

[b] Dr. D. D. Jones
Current address:
School of Biosciences
Biomedical Sciences Building, Cardiff University
Museum Avenue, Cardiff, CF10 3US (UK)
Fax: (+44) 29-20874048
E-mail: jonesdd@cardiff.ac.uk


Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim DOI: 10.1002/cbic.200300569 ChemBioChem 2004, 5, 964 – 971

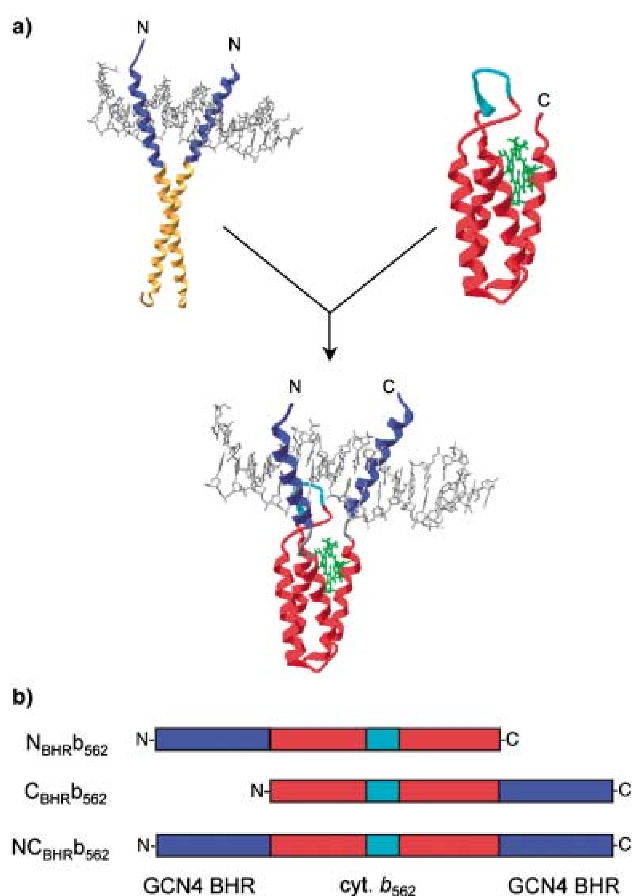



**Figure 1.** *a) Structures of GCN4 (Protein Databank (PDB) file 1YSA) and oxidised holocytochrome* $b_{562}$ *(PDB file 1QPU), together with a modelled, schematic structure of the target chimeric protein,* $NC_{BHR}b_{562}$*. The BHRs are coloured blue, the leucine zipper is orange, the cytochrome is red and the haem group is shown in green. The loop region, replaced by KA, AKA and AKRA in some of the designed proteins, is coloured cyan in the cytochrome structure. The N and C termini are indicated as N and C, respectively. b) A schematic outline of the chimeric proteins constructed. Colour coding as in (a). BHR, basic helix region.*

One of the major mechanisms that proteins use to bind DNA with high affinity is dimerisation, which allows the protein to recognise and bind palindromic DNA sequences.[11] An example of a protein that uses such a mechanism is GCN4 (Figure 1), a member of the leucine zipper (bZip) family of DNA-binding proteins.[11–13] GCN4 consists of a leucine zipper sequence, responsible for dimerisation, and a basic helix region (BHR) that binds one half site of the DNA recognition sequence. The leucine zipper region of GCN4 is largely structured but the helical nature of the BHR is induced upon binding to DNA. Both the BHR and the leucine zipper region of GCN4 have been the subject of many protein engineering experiments. Several groups have successfully removed the requirement for the leucine zipper for DNA binding by covalently linking two BHR peptides together with a disulfide bridge.[14–16] Ceunoud and Schepartz[17,18] have mimicked the covalent linkage by attaching metal chelators to the C-terminal BHR region so that dimerisation is mediated by metal binding.

We have taken this approach in a new direction by linking BHR peptides to a functionally unrelated metalloprotein in a way that integrates the two disparate functions in a monomeric protein unit. We describe herein the construction of this novel protein, which unites the inherent electron transfer activity of cyt $b_{562}$ with the DNA-binding properties of the BHR of GCN4. Our initial objective was to determine the basic features of this new scaffold so as to assess whether or not two such distinct and unrelated proteins can be successfully married. In particular, we have assessed its affinity, sequence specificity and integrity, which are properties we need to understand for future use of these proteins in nanodevices.

## Results and Discussion

### Design

Dimerisation of GCN4 through the leucine zipper region is essential for high-affinity, specific binding to DNA,[19–21] so mimicking this dimerising property is an intrinsic requirement of our monomeric protein design. One of the initial criteria used to select cyt $b_{562}$ was that the N and C termini of the four-helix bundle of the cytochrome are close enough to each other in space to allow the correct placement of GCN4 BHRs relative to each other when they are attached to the termini. The inherent flexibility of the termini of cyt $b_{562}$ allows the BHRs to sample the required conformational space to determine DNA binding. When BHRs are present at the termini, DNA binding should occur; cyt $b_{562}$ replaces the dimerisation function of the leucine zipper region of GCN4.

Combining two DNA-binding helices in one polypeptide in this manner has one obvious and important structural consequence; the BHRs will have the opposite orientation (antiparallel) to that (parallel) of the BHRs in GCN4 (Figure 1). There is no absolute requirement for the DNA-binding helix to be at the N terminus since transcription factor Skn 1 has one BHR-like recognition helix at its C terminus and recognises a DNA sequence equivalent to the half-site of AP-1.[22] Furthermore, studies with synthetic peptides of the Jun bZIP BHR dimerised through N- or C-terminal cysteine residues have shown that the nature of the recognition of the DNA sequence half-sites depends upon the orientation of the basic regions with respect to each other (*parallel versus antiparallel*).[23]

As mentioned above, the structural changes accompanying haem binding to cyt $b_{562}$ are relatively well understood. The large amount of free energy (around 15 $kJ\,mol^{-1}$[24a]) that is made available upon cofactor binding could also be linked to processes other than the folding of the cytochrome. This energy is oxidation-state dependent, so there is a thermodynamic link to electron transfer activity. Such a high-energy coupled folding process could be utilised to modulate DNA binding activity. Another important reason for selecting cyt $b_{562}$ as the scaffold is the position of the loop linking helices 2 and 3, close in space to the proposed point of attachment of the BHRs and hence the DNA binding site (Figure 1). Recent studies have shown that the size and nature of this loop can influence the properties of the cytochrome as a whole (see below). This loop could therefore be used to influence the properties

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

of the chimera either by direct binding to the DNA or by allosteric modulation of the stability of the protein.

Three chimeric cyt $b_{562}$ proteins were constructed, two with the GCN4 BHR at either the N or C terminus ($N_{BHR}b_{562}$ and $C_{BHR}b_{562}$, respectively) and the third with a BHR at both the N and the C terminus ($NC_{BHR}b_{562}$; Figure 1). The properties of the former two proteins illustrate the requirement for two BHRs acting cooperatively to acquire tight, site-specific DNA binding. No additional linking sequence was added between the two functional components. We also studied three examples of mutants of $NC_{BHR}b_{562}$ in which changes were made to the loop joining helices 2 and 3. Specifically, the residues from Glu49 to Glu57 were replaced by KA, AKA or AKRA.

The potential DNA recognition site was chosen rationally according to information relating to GCN4 and Skn 1. As the topology of one of the BHR helices is inverted in our chimeric proteins, the predicted recognition sequence (dsNC1, ATGAcgATGA) for the $NC_{BHR}b_{562}$ construct has one half-site inverted. The half-sites are separated by CG base pairs. The natural DNA recognition sequence for wild-type GCN4 is dsGCN4 (ATGAcTCAT), but we did not assume that this sequence is optimal for $NC_{BHR}b_{562}$. Previous work has shown that GCN4 in fact only requires one half-site of the palindromic sequence to bind to DNA.[25] Therefore, $NC_{BHR}b_{562}$ may show some limited binding to dsGCN4 but the affinity should be lower than that for binding to the designed optimal recognition sequence, dsNC1. To rule out the possibility of one BHR playing the predominant DNA-sequence-specifying role with the other BHR making stabilising yet nonspecific phosphate interactions, binding to DNA with only one half of the recognition sequence (dsNC$^1/_2$, ATGAcgGGCC) was assessed. Oct2a is a double-stranded oligonucleotide with no potential binding sites and was used to monitor nonspecific DNA interactions.

### Ability of chimeric proteins to bind haem

The most easily characterised property of the new chimeric proteins is their ability to bind haem. The electronic spectra provide evidence that the presence of the BHR at either one or both of the termini does not cause the spectral properties of the protein to change in comparison to those of wild-type cyt $b_{562}$ (Figure S1 in the Supporting Information). Like wild-type cyt $b_{562}$, $N_{BHR}b_{562}$, $C_{BHR}b_{562}$ and $NC_{BHR}b_{562}$ have $\lambda_{max}$ values of 426.5 nm, 531 nm and 561.5 nm in their reduced form and the signal at 426.5 nm undergoes the characteristic blue shift to 418 nm upon oxidation of the proteins (Figure S1). This retention of the cytochrome spectral properties by the new chimeric proteins indicates that the four-helix bundle and, more importantly, the first and fourth helices (containing the ligand-binding residues) are structurally intact. The haem environment is thus unaltered. The affinity of each of these constructs for ferric haem was measured by titration with free haem at pH 8.0. We found that the dissociation constants of these proteins (20–50 nM) were slightly higher than that of the wild-type protein (5–10 nM).[24] The affinity of wild-type cyt $b_{562}$ for reduced haem has been calculated as much higher than these values ($K_D \approx 10$ pM) and so far has not been measured directly by any method. We conclude that attachment of the DNA-binding helices to the wild-type cytochrome does not significantly alter the intrinsic haem affinity.

The affinity of the protein $NC_{BHR}b_{562}$ for oxidised haem was also measured in the presence and absence of the target double-stranded oligonucleotide, dsNC1. Experiments were performed by addition of haem to solutions of protein or preformed protein/DNA mixtures, or vice versa. The dissociation constant for the haem–$NC_{BHR}b_{562}$ complex is 30–50 nM and is independent of the presence of DNA. Addition of DNA to the preformed protein–haem complex did not result in any changes in the haem spectrum.

### DNA binding and specificity of $NC_{BHR}b_{562}$

Traditional gel mobility shift assays (GMSA) were used to establish whether the chimeric proteins have the ability to recognise and bind DNA. When the $NC_{BHR}b_{562}$ chimera was in the apo form, tight and specific DNA binding was observed (Figure 2 a)

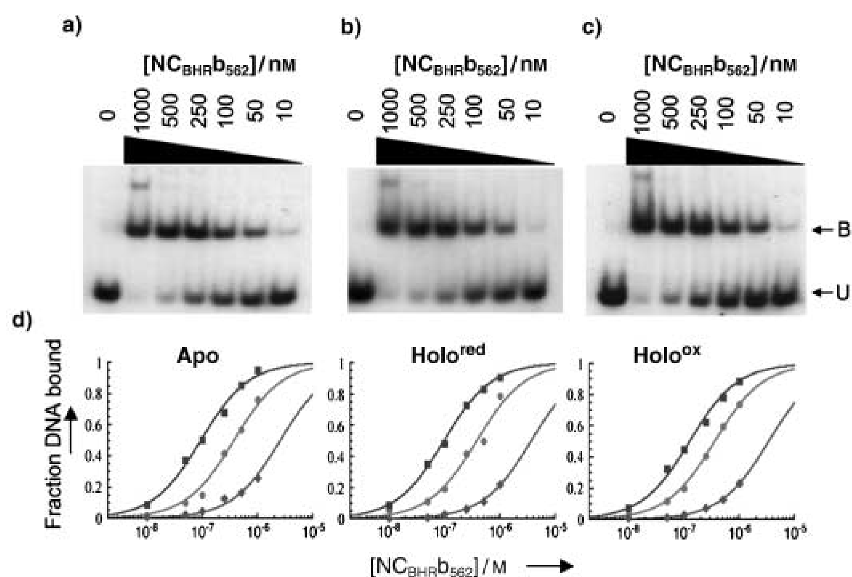



*Figure 2. a, b and c) Binding of $NC_{BHR}b_{562}$ to dsNC1 DNA. Gels of the three forms of $NC_{BHR}b_{562}$, (a) apo, (b) holo$^{red}$ and (c) holo$^{ox}$, are shown. U signifies the position of unbound DNA and B refers to the protein–DNA complex. In the case of the apo and reduced holo forms, 1,4-dithiothreitol (DTT) was present in the binding buffer. For the oxidised holo form, DTT was replaced by potassium ferricyanide. d) Graphical representation of the extent of binding of the three forms of $NC_{BHR}b_{562}$ to dsNC1 (■), dsGCN4 (●) and dsNC$^1/_2$ (◆) DNA. The fraction of DNA bound to protein was determined by densitometry and each solid line represents a least-squares fit to the data.*

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

in the presence of 5 nM digoxigenin (DIG)-labelled dsNC1 and competitor DNA (poly-I-C). The equilibrium dissociation constant was 92±6 nM (Table 1) under these conditions and dropped to below 10 nM in the absence of competitor DNA (data not shown). This result compares favourably with the $K_D$ value (90 nM in the presence of competitor DNA) most recently determined by GMSA for wild-type GCN4 (Table 1 and ref. [25]). We have generated a protein that recognises a DNA sequence different from that recognised by the parent GCN4 and maintains a high binding affinity despite alteration of the orientation of the DNA-binding helices.

The $K_D$ value of the holo$^{red}$ form of the chimera is 96±3 nM, which is similar to that of the apo form. The $K_D$ value of the oxidised form is only slightly higher at 125±8 nM (Figure 2, Table 1). These results show that each state of the protein maintains the same binding affinity for the DNA target sequence. Although binding occurs under equilibrium conditions, the electrophoresis procedure removes the system from equilibrium. Experiments with the wild-type cytochrome alone indicate that the haem group remains associated with the protein under nondenaturing electrophoresis conditions. Absolute proof of this association is provided by the results of analytical ultracentrifugation experiments (see below).

**Table 1.** *Equilibrium dissociation constants of the apo, holo$^{red}$ and holo$^{ox}$ forms of the chimera proteins for binding with dsNC1 and dsGCN4 DNA (for sequences, see the Experimental Section).[a]*

| | $K_D$ for binding dsNC1 DNA [nM] | | | $K_D$ for binding dsGCN4 DNA [nM] | | |
|---|---|---|---|---|---|---|
| Protein | Apo | Oxidised | Reduced | Apo | Oxidised | Reduced |
| $NC_{BHR}b_{562}$ | 92±6 | 96±3 | 125±8 | 372±25 | 358±37 | 382±88 |
| $N_{BHR}b_{562}$ | NB[b] | NB | NB | NB | NB | NB |
| $C_{BHR}b_{562}$ | NB | NB | NB | NB | NB | NB |
| AKRA-$NC_{BHR}b_{562}$ | 135±17 | nd[c] | 70±6 | 590±25 | nd | 210±16 |
| AKA $NC_{BHR}b_{562}$ | 535±71 | nd | 649±52 | >μM | nd | >μM |
| KA $NC_{BHR}b_{562}$ | 113±9 | nd | 162±13 | nd | nd | nd |
| GCN4 | nd | na[d] | na | 90[e] | na | na |

[a] The $K_D$ values were calculated by determining the fraction of DNA bound to protein at a fixed DNA concentration by using densitometry and curve fitting.[25] [b] NB, no binding observed in GMSA experiments. [c] nd, not determined. [d] na, not applicable. [e] See ref. [25].

When only one BHR was present at either the N or C terminus, as in $N_{BHR}b_{562}$ and $C_{BHR}b_{562}$, respectively, no specific DNA binding to dsNC1 was observed in the GMSA assays (Figure 3). We conclude that the two BHR regions of the $NC_{BHR}b_{562}$ chimera act cooperatively to achieve tight and specific DNA binding.

Two oligonucleotides, dsGCN4 and dsNC$^1/_2$, were used to investigate the specificity of the $NC_{BHR}b_{562}$ chimera with respect to half-site recognition. All three forms of the chimera (apo, holo$^{red}$ and holo$^{ox}$) were tested against dsGCN4 and the $K_D$ values (372±25 nM, 382±37 nM and 358±88 nM, respectively) were found to be a factor of four higher than those measured for the proposed optimal recognition sequence, dsNC1 (see Figure 2d and Figure S2b in the Supporting Information). The binding of the protein to dsGCN4 may result from residual contacts between one BHR region and part of the nonoptimal half-site. Structural studies have shown that each BHR is separated into two units, one that recognises the inner two bases and another that recognises the outer two bases of the half-site. If the N-terminal BHR of $NC_{BHR}b_{562}$ binds correctly to the consensus half-site, then the C-terminal BHR region may partially bind to the other half of the sequence. It is possible that such interaction occurs with dsGCN4 and leads to stabilisation of the complex. The second half of dsGCN4 could be split into two regions, TC and AT, recognised by the two units of the BHR. If the same principle is applied to dsNC1, the two DNA regions would be AT and GA; the AT region is common to both dsNC1 and dsGCN4. It may well be that the BHR at the C terminus of $NC_{BHR}b_{562}$ has sufficient conformational freedom to allow recognition of the AT unit of dsGCN4 by its corresponding BHR unit, but the GA unit is not recognised. This partial recognition will account thermodynamically for the difference in $K_D$ values and also suggest that the chimera still provides a high degree of specificity. If this is the case, removal of the AT unit (or potentially the GA unit) from the second half-site should reduce the ability of $NC_{BHR}b_{562}$ to bind DNA. The binding of the chimera to dsNC$^1/_2$ demonstrates that the hypothesis is correct since $K_D$ values of more than 2500 nM were observed for binding of all forms of the chimera to this sequence (see Figure S2a in the Supporting Information; Table 1), a 30-fold increase in dissociation constant compared to the values determined for dsNC1. These results also demonstrate that the chimera has a higher degree

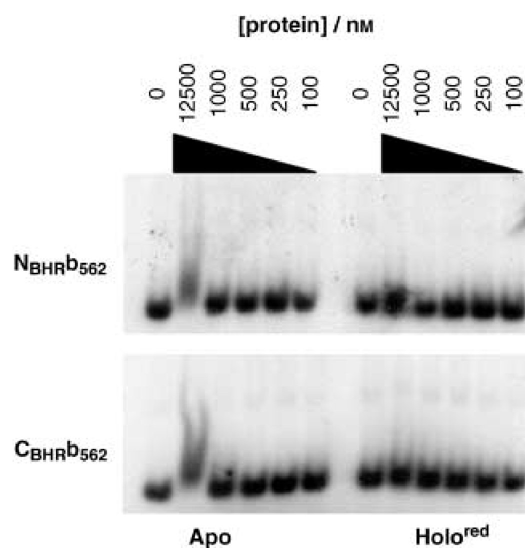



**Figure 3.** *Effect of attachment of a single BHR to either the N ($N_{BHR}b_{562}$) or C terminus ($C_{BHR}b_{562}$) on the ability of cyt $b_{562}$ to bind dsNC1 DNA. Conditions as described in the legend of Figure 2 and the Experimental Section.*

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

of specificity than the wild-type GCN4 protein, which has a $K_D$ value of 800 nM for a similar DNA sequence with only one half-site.[25] The weak binding that occurs between the chimera and dsNC$^1/_2$ could be the result of the N-terminal BHR of $NC_{BHR}b_{562}$ retaining the ability to recognise the consensus half-site and the C-terminal BHR making weak, peripheral phosphate backbone contacts. This hypothesis also explains why no binding is observed when one BHR is attached to either the N or C terminus (Figure 3).[11]

A third sequence, Oct2a DNA, that bears no relationship to dsNC1 was investigated. Neither the apo nor the holo$^{red}$ form of the chimera bound Oct2a DNA specifically below 1 mM (see Figure S3 in the Supporting Information). These data indicate that binding of the chimera to dsNC1 is specific to that DNA sequence and the protein can discriminate between closely related sequences.

### DNA–protein binding stoichiometry

The complex formed by $NC_{BHR}b_{562}$ and dsNC1 DNA was analysed by equilibrium ultracentrifugation to determine the molecular weight of the complex and hence its stoichiometry. Experiments performed with a chimera containing an oxidised haem moiety allowed the sedimentation behaviour of the haem-containing material to be measured independently (at 420 or 562 nm) of that of the protein or nucleic acid. The apo and oxidised holoproteins were sedimented alone and in the presence of dsNC1 DNA. The apparent molecular weights of apo and holo $NC_{BHR}b_{562}$ were measured as 17300 Da and 17500 Da, respectively. Analysis of 1:1 mixtures of the protein and nucleic acid at three different concentrations well above the $K_D$ value indicated that a single species with an apparent molecular mass of about 39000 Da was present. This mass is very close to the expected molecular weight of the 1:1 protein–DNA complex (40844 Da). Spectra of samples containing haem were recorded from 700 to 400 nm at several radial positions in the centrifugation cell and the results showed that the haem remained fully complexed with the protein. There was no evidence of the existence of higher-order aggregates.

### Alteration of the loop linking helices 2 and 3 of cytochrome $b_{562}$ affects DNA binding

One part of the four-helix bundle that may have a direct steric and/or electrostatic influence on DNA binding is the loop between helices 2 and 3 in cytochrome $b_{562}$ (Figure 1). This 15 amino acid loop is dynamic and not specifically structured in the apoprotein but is much more ordered in the holoprotein. In the context of the chimeric protein, our models predict that this loop is the part of the four-helix bundle closest to the DNA within the DNA–protein complex. As part of a different project aimed at examining the role of this long loop in the stability and haem affinity of cytochrome $b_{562}$, we replaced this 15 amino acid sequence with a variety of shorter sequences. A detailed thermodynamic analysis of the resulting proteins[26] revealed that, in the cytochrome alone, removal of all 15 residues is tolerated and the protein retains significant stability and haem affinity. We incorporated some of these specific loop deletions into our chimeric $NC_{BHR}b_{562}$ proteins. For example, replacement of residues Glu49–Glu57 with the sequence AKRA results in a protein with interesting properties. This loop-shortened chimera binds oxidised haem with a slightly lower affinity than the wild-type protein ($K_D$ = 100 nM). However, the affinity of this loop-shortened protein for DNA is dependent on the presence or absence of haem (Figure 4). The dissociation con-

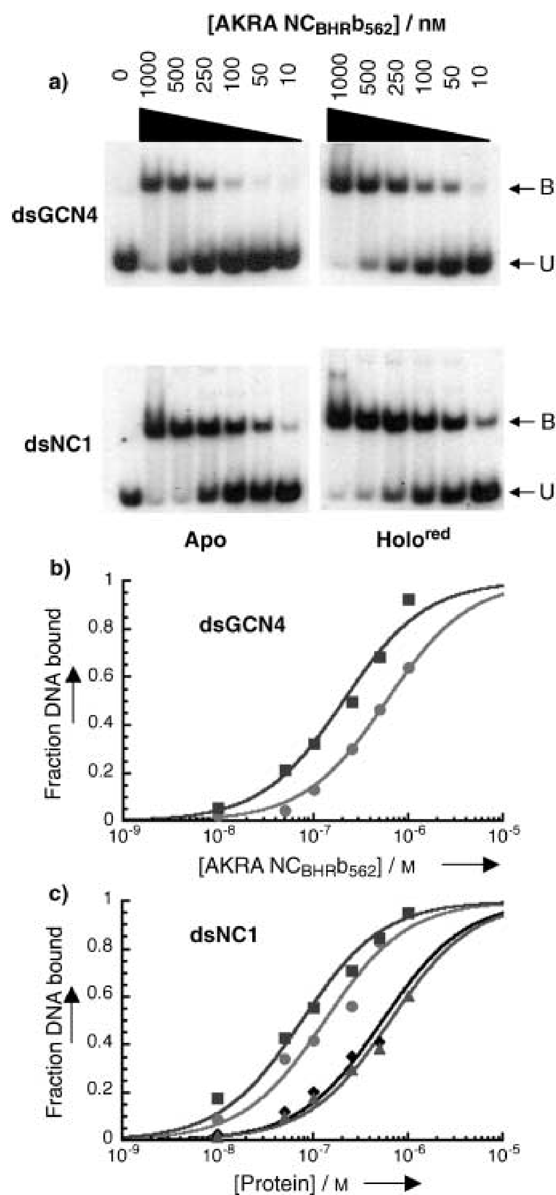



*Figure 4. DNA binding selectivities of AKRA $NC_{BHR}b_{562}$ and AKA $NC_{BHR}b_{562}$. a) The affinities of apo and holo$^{red}$ $NC_{BHR}b_{562}$ AKRA for dsNC1 and dsGCN4. U signifies unbound DNA and B refers to the protein–DNA complex. A graphical representation of the extent to which the two forms (holo$^{red}$, ■; apo, ●) of AKRA-$NC_{BHR}b_{562}$ bind to (b) dsGCN4 and (c) dsNC1 DNA is shown. The binding of the apo (◆) and holo$^{red}$ (▲) forms of AKA $NC_{BHR}b_{562}$ to dsNC1 is also shown in (c). The fraction of DNA bound to protein was determined by densitometry and each solid line represents a least-squares fit to the data. Data for the oxidised holoproteins are left out for clarity. The $K_D$ values for DNA binding of this form of each protein were, within experimental error, the same as those of the equivalent, reduced holoprotein.*

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

stants of the apoprotein and reduced holoprotein for dsNC1 DNA are 135 nm and 70 nm, respectively, while the values for binding to dsGCN4 DNA are 590 nm and 210 nm, respectively. The difference between the binding affinities of the two forms is greater for binding to the dsGCN4 DNA than for the higher-affinity binding to dsNC1 DNA. In both cases, the $K_D$ values for the two forms of the protein straddle those observed for the full-length, parent construct. The complex subsite recognition discussed above may be relevant in this context. We observed a threefold difference in DNA binding affinity between the apo and holo forms of the protein and we believe that this is the first evidence that haem binding can be coupled to DNA binding in this chimeric construct. Other loop replacements provide further evidence that the loop sequence can influence function in the context of the chimera. In some cases, there is a general decrease in the affinity of the chimera for DNA without loss of specificity (Figure 4). For example, the $K_D$ values of the AKA $NC_{BHR}b_{562}$ protein for binding to dsNC1 ($535\pm71$ nm and $649\pm52$ nm for the apo and reduced holo forms, respectively) are lower than those of $NC_{BHR}b_{562}$, while the $K_D$ values for binding to the nonoptimal dsGCN4 sequence are above the micromolar range (Table 1). The binding affinity of the KA $NC_{BHR}b_{562}$ protein is little altered for either DNA sequence compared to the values measured for the $NC_{BHR}b_{562}$ protein with the full-length loop (Table 1).

We believe that the origin of the link between haem and DNA binding in these proteins lies in a coupled folding process and that the stability of the four-helix bundle is lowered and its flexibility constrained by loop shortening. DNA binding therefore increases the stability of the protein and preorders the helical bundle for haem binding, or vice versa. Our data show that the cytochrome is not just a passive scaffold for displaying DNA-binding helices and that further engineering of the cytochrome is a valid route to stronger coupling.

## Conclusion

We have successfully integrated the very dissimilar functions of two proteins into one monomeric protein to create a novel, DNA-binding cytochrome. The new protein retains the spectral characteristics of wild-type cyt $b_{562}$ but has gained the ability to bind DNA. The absence of any high-affinity, specific DNA binding by chimera proteins with only one BHR region shows that two BHRs are required to act cooperatively to bind DNA. Binding by the chimera $NC_{BHR}b_{562}$ to its specific DNA target occurs with an affinity comparable to that of GCN4. The complex formed is specific, with a 1:1 ratio of protein to double-stranded DNA. No higher-order aggregates are formed. This work represents the first characterisation of such a system that illustrates the potential of these molecules for precise assembly in defined locations on DNA and introduction of cofactor chemistry into DNA nanostructures.

The chimera shows a high degree of selectivity for the DNA sequence it recognises. The tightest binding was observed with dsNC1 DNA, which has a recognition sequence designed to take into account the topological reversal of the BHR region attached to the C terminus of the cytochrome. The $K_D$ value for dsGCN4 DNA is a factor of four higher than that for dsNC1 DNA, which confirms the discriminatory nature of $NC_{BHR}b_{562}$. Such binding to dsGCN4 DNA could be attributed to the modular ("2+2") nature of the binding of the BHR to the DNA recognition element; two base pairs in the half-site may be bound rather than the full four. Such an event could explain why the $K_D$ value is not as high as expected since limited specific binding provides a degree of stability within the complex. Binding to the full palindromic sequence is required since the removal of one half-site from the DNA (dsNC$^1/_2$) results in a dramatic shift of the $K_D$ value to much higher values (> μm). It appears that $NC_{BHR}b_{562}$ can discriminate between DNA sequences to a greater extent than is achieved by GCN4. Importantly, cytochromes with only one BHR region exhibit no specific DNA binding activity at all. The selectivity of $NC_{BHR}b_{562}$ is also demonstrated by the lack of any binding to Oct2a DNA.

The construct based upon the wild-type cyt $b_{562}$ protein sequence binds DNA with an affinity that is independent of cofactor binding or cofactor oxidation state. Such a feature is desirable if this protein is to be used to assemble porphyrin binding sites into an ordered conducting array on DNA. If conformational coupling is required then engineering of the long loop in the four-helix bundle represents one route to a suitable protein and our initial results with constructs designed for this purpose are described above. Replacement of part of the long loop with the sequence AKRA results in a protein whose affinity for the dsGCN4 oligonucleotide is linked to its affinity for haem. When the equivalent loop residues are replaced by AKA, the affinity for DNA decreases without apparent loss in specificity. The influence of the sequence of this loop upon the DNA affinity may reflect some additional direct interactions between this loop and the bound DNA (reminiscent of Skn-1). However, we believe that the principle mechanism linking changes in the cytochrome sequence to changes in DNA affinity is the coupled folding energy that links the cytochrome stability with the BHR stability. Engineering such indirect cooperative interactions may therefore be key to exaggerating the coupling of the functions. The GCN4 and cyt $b_{562}$ proteins alone show allosteric properties in the sense that their fold is dependent on the presence or absence of DNA and haem, respectively. We have shown herein that joining the two proteins together allows transmission of this conformation-dependent ligand-binding behaviour between the different proteins upon further engineering of the basic scaffold.

Recent work has revealed that nature has used conformational coupling of haem chemistry with DNA chemistry to generate a carbon monoxide sensor. The CO sensor from the bacterium *Rhodospirillum rubrum*, CooA, is a transcription factor that activates the expression of certain genes in response to CO.[27] CooA is a homodimer containing one haem group per subunit, but the axial ligands of each haem are provided by both subunits in a domain-swapping manner.[28] As with other haem proteins, CooA is unable to bind CO when the haem Fe moiety is oxidised. Upon reduction, there is an unusual switch of axial ligands in the six-coordinate haem group and one of the axial ligands is displaced by CO. This binding stabilises a conformation of the dimeric protein that allows high-affinity

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

sequence-specific DNA binding, which leads to transcription of the *coo* genes. It is thus clear that haem chemistry and DNA binding can be coupled within a single polypeptide unit. CooA has evolved for a very specific role in vivo and our project attempts to create a different basic scaffold that can be engineered or evolved in vitro to give a protein more suited to device construction. One goal of this project is to generate a protein in which DNA binding and the electronic state of the haem group are functionally linked. Another key target is the alignment of metalloproteins into a conducting chain. As they stand, the proteins described herein will clearly generate a specific metalloporphyrin binding site at a desired location on a double-stranded DNA molecule. Careful design of DNA sequences containing adjacent recognition sites for the proteins could allow us to align a chain of metalloporphyrins on a DNA template by using our DNA-binding cytochromes. This idea is currently under investigation by single molecule methods. DNA is being exploited as an excellent material for creating self-assembling nanostructures and our novel proteins provide an integrated route to incorporation of specific redox chemistry into these structures.

## Experimental Section

**DNA manipulation**: Wild-type cyt $b_{562}$ was subcloned from the pPB10 vector[29] into pBluescript II KS (Stratagene) between the *Xba*I and *Eco*RV sites. Prior to subcloning, new restriction sites corresponding to *Stu*I and *Nru*I were introduced into the cyt $b_{562}$ gene by splice-overlap mutagenesis.[30,31] The addition of the basic region of GCN4 to cyt $b_{562}$ was carried out by cassette mutagenesis.[32] To construct the N-terminal GCN4–$b_{562}$ chimera, the cyt $b_{562}$ gene was cut with *Nru*I and *Bgl*II and the oligonucleotide cassette encoding the GCN4 basic region was inserted between those sites. To synthesise the C-terminal GCN4–$b_{562}$ chimeric protein, the oligonucleotide cassette encoding the reversed GCN4 BHR was inserted between *Stu*I and *Xba*I sites. To construct the N and C-terminal GCN4–$b_{562}$ subgene, both cassette mutagenesis procedures were performed. The leader sequence used to direct cyt $b_{562}$ to the periplasm was removed by PCR amplification of the genes coding for the mature proteins alone. All the GCN4–$b_{562}$ genes were inserted between the *Nco*I and *Bam*HI sites of a pET11d plasmid (Novagen). Genes coding for three proteins in which the 15 amino acid loop between helices 2 and 3 of cytochrome $b_{562}$ is significantly shortened were generated by amplification of two halves of the gene by PCR, followed by splicing of these separate PCR products. In the resulting mature protein sequences, the nine-residue stretch EDKSPDSPE is replaced by the sequence AKRA, AKA or KA. The entire sequence of each of the genes generated was confirmed by sequencing (Biochemistry Department Sequencing Service, University of Cambridge) prior to expression.

**Protein purification**: The above constructs were expressed in *E. coli* BL21 Gold (Novagen) cells. A single colony was used to inoculate a one-litre culture of 2xTY medium containing ampicillin (100 μg mL$^{-1}$). At an absorbance at 600 nm ($A_{600}$) of 1.0, the cells were induced by the addition of isopropyl-β-D-thiogalactopyranoside (1 mM) and left at 30 °C overnight. The cells were harvested and resuspended in tris(hydroxymethyl)aminomethane (Tris)-HCl (15 mL per litre of medium, pH 7.4, 100 mM) and ethylenediaminetetraacetate (EDTA; 5 mM). The cells were lysed with a homogeniser (Emulsiflex-C5, Avenstin) and spun at 39000 g (18000 rpm in an SS34 rotor, Sorvall RC-5B) for 30 min. The supernatant was applied to a Sepharose S column in Buffer A (20 mM Tris-HCl, pH 7.4, 1 mM DTT) and developed with an NaCl gradient. The pooled fractions were concentrated and applied to a Sepharose 75 gel filtration column equilibrated in Buffer A containing NaCl (300 mM). The apo and holo forms of the proteins were separated on a 1 mL mono S column developed with Buffer A by using a NaCl gradient under reducing conditions. The charge difference between the reduced holo and apo forms is +2. The holoprotein could be produced by adding haem (1.5 equiv) to the apoprotein. Excess haem and apoprotein were removed by using a mono S column under reducing conditions, as described above. The apoprotein was judged to be homogeneous based on the results of sodium dodecylsulfate gel electrophoresis and high-resolution ion exchange chromatography. Protein concentrations were determined spectrophotometrically by using the following extinction coefficients: 3000 M$^{-1}$cm$^{-1}$ at 277 nm for the apoprotein; 117000 M$^{-1}$cm$^{-1}$ at 417 nm and 181000 M$^{-1}$cm$^{-1}$ at 427.5 nm for the oxidised and reduced holoproteins, respectively. Full haem occupancy of the chimeric proteins was confirmed by titrating the proteins with haem and monitoring any further haem binding. The extinction coefficients were estimated by monitoring the absorbance of the proteins during titration with haem and are virtually identical to those of the parent proteins. Haem insertion or removal was performed as previously described.[29] Reduction and oxidation were achieved by the addition of dithiothreitol or potassium ferricynide, respectively, followed by removal of the reductant/oxidant by ultrafiltration if necessary.

The proteins were checked for correct expression and proteolytic fragmentation by measuring the mass of each protein by ESI mass spectrometry (Quattro LC, Micromass) under acidic denaturing conditions. In all cases, the measured masses were within 2 Da of the expected values.

**Gel mobility shift assay using digoxigenin-labelled oligonucleotides**: The gel mobility shift assay was performed according to the instructions accompanying the DIG gel shift kit supplied by Roche Biochemicals. 5′-DIG-labelled oligonucleotides were synthesised at TAG Copenhagen and annealed by heating equimolar mixtures of complementary oligonucleotides to 95 °C for 5 min then cooling them to 20 °C at 0.1 °C s$^{-1}$. The sequences of the labelled double-stranded oligonucleotides used in the gel mobility shift assays are (only one strand shown, with predicted target sequence capitalised):

dsNC1, 5′-gcctaggttaaacgATGAcgATGAcggtataggtcggg-3′;

dsGCN4, 5′-gcctaggttaaacgATGAcTCATcggtataggtcggg-3′;

dsNC$^1/_2$, 5′-gcctaggttaaacgATGAcgGGCCcggtataggtcggg-3′;

Oct2a, 5′-gtacggagtatccagctccgtagcatgcaaatcctctgg-3′

Varying amounts of protein were added to binding buffer (10 % *v/v* glycerol, 5 mM $MgCl_2$, 5 mM KCl, 50 μg mL$^{-1}$ bovine serum albumin, 20 mM Tris-HCl, pH 7.4), together with poly[d(I-C)] (5 ng mL$^{-1}$), DIG-labelled double-stranded oligonucleotide (5 nM) and either DTT (2 mM; reductant) or potassium ferricyanide (2 mM; oxidant). The samples were left at room temperature for 20 min and then placed on ice. Gel electrophoresis was carried out on Novex precast 6 % nondenaturing polyacrylamide DNA retardation gels in ×0.25 Tris-borate-EDTA buffer at 4 °C. Blotting and chemiluminescent detection were carried out according to the instructions provided in the DIG gel shift kit supplied by Roche Biochemicals. The chemiluminescent signals were recorded by exposure of x-ray film to the emission.

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 

**Other methods**: Spectrophotometric analysis of proteins was carried out at 20 °C on either a Hewlett Packard 8453 or a Cary 500 UV/Vis spectrophotometer. The spectral characteristics of each protein (4 μM) were recorded in water in the presence of DTT (10 mM; reduced form) or potassium ferricyanide (0.5 mM; oxidised form). Equilibrium ultracentrifugation was carried out in a Beckman Xli centrifuge. Samples of protein (200 μM and 5 μM for the apo and holoprotein, respectively) in Tris-HCl buffer (0.1 M, pH 7.4) were centrifuged at speeds of 10000, 12500 and 18000 rpm until equilibrium was reached. The absorbance was monitored at 277 nm for the apoproteins and 420 or 562 nm for the holoproteins. The DNA complexes were formed by adding DNA (1.0 equiv) to the proteins. Data were analysed with ULTRASPIN software.[33]

## Acknowledgements

*This work was funded by the Biotechnology and Biological Sciences Research Council through a grant (no. SBD07575) and an Advanced Research Fellowship awarded to P.D.B. We would like to thank A. Fersht for support through the Cambridge Centre for Protein Engineering. We also thank H. Bosshard and U. Zurich for help and encouragement early in this project. We thank C. Squire and J. Clarke for critical reading of the manuscript.*

**Keywords:** chimeric cytochromes · DNA recognition · haem proteins · nanotechnology · protein design

[1] D. E. Benson, M. S. Wisz, H. W. Hellinga, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6292.
[2] H. W. Hellinga, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10015.
[3] H. W. Hellinga, *Nature Struct. Biol.* **1998**, *5*, 525.
[4] J. S. Marvin, H. W. Hellinga, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4955.
[5] L. S. Mizoue, W. J. Chazin, *Curr. Opin. Struct. Biol.* **2002**, *12*, 459.
[6] D. E. Benson, D. W. Conrad, R. M. de Lorimier, S. A. Trammell, H. W. Hellinga, *Science* **2001**, *293*, 1641.
[7] F. Arnesano, L. Banci, I. Bertini, J. Faronoe-Mennella, A. Rosato, P. D. Barker, A. R. Fersht, *Biochemistry* **1999**, *38*, 8657.
[8] K. Hamada, P. H. Bethge, F. S. Mathews, *J. Mol. Biol.* **1995**, *247*, 947.
[9] Y. Feng, S. G. Sligar, A. J. Wand, *Nature Struct. Biol.* **1994**, *1*, 30.
[10] N. D'Amelio, A. M. Bonvin, M. Czisch, P. Barker, R. Kaptein, *Biochemistry* **2002**, *41*, 5505.
[11] C. W. Garvie, C. Wolberger, *Mol. Cell* **2001**, *8*, 937.
[12] T. E. Ellenberger, *Curr. Opin. Struct. Biol.* **1994**, *4*, 12.
[13] T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **1992**, *71*, 1223.
[14] R. V. Talanian, C. J. McKnight, P. S. Kim, *Science* **1990**, *249*, 769.
[15] C. Park, J. L. Campbell, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9094.
[16] C. Park, J. L. Campbell, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4892.
[17] B. Cuenoud, A. Schepartz, *Science* **1993**, *259*, 510.
[18] B. Cuenoud, A. Schepartz, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1154.
[19] T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **1992**, *71*, 1223.
[20] P. Konig, T. J. Richmond, *J. Mol. Biol.* **1993**, *233*, 139.
[21] T. K. Kerppola, T. Curran, *Curr. Opin. Struct. Biol.* **1991**, *1*, 71.
[22] P. B. Rupert, G. W. Daughdrill, B. Bowerman, B. W. Matthews, *Nature Struct. Biol.* **1998**, *5*, 484.
[23] C. Park, J. L. Campbell, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4892.
[24] a) C. R. Robinson, Y. F. Liu, J. A. Thomson, J. M. Sturtevant, S. G. Sligar, *Biochemistry* **1997**, *36*, 16141; b) T. De Lumley, D. D. Jones, C. Henning-Nagel, P. D. Barker, unpublished results.
[25] J. J. Hollenbeck, M. G. Oakley, *Biochemistry* **2000**, *39*, 6380.
[26] C. Henning-Nagel, D. D. Jones, P. D. Barker, unpublished results.
[27] D. Shelver, R. L. Kerby, Y. He, G. P. Roberts, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11216.
[28] W. N. Lanzilotta, D. J. Schuller, M. V. Thorsteinsson, R. L. Kerby, G. P. Roberts, T. L. Poulos, *Nature Struct. Biol.* **2000**, *7*, 876.
[29] P. D. Barker, E. P. Nerou, S. M. V. Freund, I. M. Fearnley, *Biochemistry* **1995**, *34*, 15191.
[30] S. N. Ho, H. D. Hunt, R. M. Horton, J. K. Pullen, L. R. Pease, *Gene* **1989**, *77*, 51.
[31] R. M. Horton, H. D. Hunt, S. N. Ho, J. K. Pullen, L. R. Pease, *Gene* **1989**, *77*, 61.
[32] J. A. Wells, M. Vasser, D. B. Powers, *Gene* **1985**, *34*, 315.
[33] http://www.mrc-cpe.cam.ac.uk/ultraspin_intro.php?menu=1805.

Received: February 3, 2003
Revised: April 28, 2004

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim 