Repertoire Selection of Variant Single-Chain Cro: Toward Directed DNA-Binding Specificity of Helix−Turn−Helix Proteins†

Mikael T. I. Nilsson and Mikael Widersten*

Department of Biochemistry, Uppsala University, Biomedical Center, Box 576, SE-751 23 Uppsala, Sweden

Received April 30, 2004; Revised Manuscript Received July 27, 2004

ABSTRACT: A single-chain derivative of the lambda Cro repressor (scCro) has been randomly mutated in amino acid residues critical for specific DNA recognition to create libraries of protein variants. Utilizing phage display-afforded affinity selection, scCro variants have been isolated for binding to synthetic DNA ligands. Isolated scCro variants were analyzed functionally, both in fusion with phage particles and after expression of the corresponding free proteins. The binding properties with regard to specificity and affinity in binding to different DNA ligands were investigated by inhibition studies and determination of equilibrium dissociation constants for formed complexes. Variant proteins with altered DNA-sequence specificity were identified, which favored binding of targeted synthetic DNA sequences over a consensus operator sequence, bound with high affinity by wild-type Cro. The specificities were relatively modest (2−3-fold, as calculated from K_D values), which can be attributed to the inherent properties in the design of the selection system; one half-site of the synthetic DNA sequences maintains the consensus operator sequence, and one "subunit" of the variant single-chain Cro dimers was conserved as wild-type sequence. The anticipated interaction between the wild-type subunit and the consensus DNA half-site of target DNA ligands is, hence, expected to contribute to the overlap in sequence discrimination. The binding affinity for the synthetic DNA sequences, however, was improved 10−30-fold in selected variant proteins as compared to "wild-type" scCro.

DNA-binding proteins that utilize the helix−turn−helix (HTH)[1] motif or variants thereof for DNA sequence recognition are found in organisms from all kingdoms as well as in bacteriophages and other viruses (*1*−*4*). Although there is substantial knowledge regarding structure/function relationships in DNA binding of several HTH proteins, the available data describe evolutionary snapshots of DNA-binding proteins and are examples of structural solutions to existing biological problems, for example, regulation of gene expression. Because maximum affinity and recognition specificity in DNA binding is not necessarily benficial to an organism, the evolutionary pressure toward such properties is expected to be weak. It is, hence, not possible to directly derive general structural criteria applicable to predictions of binding properties of a given HTH protein by extrapolating results gathered in studies of naturally occurring protein/DNA complexes.

To address such general aspects of HTH protein−DNA recognition, we have pursued an approach of repertoire selection for novel HTH proteins with new and predefined DNA target specificities. These proteins are derivatives of a single-chain form of the lambda Cro repressor (scCro) (*5*). Large protein libraries of scCro variants of diverse DNA-binding properties have been constructed by randomized mutagenesis of amino acid residues critical for specific DNA recognition (Figure 1a) (*6*). Affinity-based selection afforded by phage display (*7*, *8*) toward predesigned synthetic DNA targets (Figure 1b) has enabled the isolation of scCro variants of distinct functional properties, which can be linked to corresponding primary structures. Previous studies on the DNA binding of zinc-finger proteins have demonstrated the feasibility of the approach (*9*) with regard to addressing fundamental issues of zinc-finger−DNA recognition as well as in the construction of novel zinc-finger proteins of predefined binding properties (*10*).

The DNA molecules applied as target ligands in the present study have been designed to probe different aspects of scCro protein−DNA binding interactions: (i) To explore amino acid sequences favoring selection of a given variant, the nucleotide sequences of the ligands have been drastically altered from a known Cro high-affinity ligand, denoted ORC (*11*), while maintaining the overall nucleotide composition (Figure 1b). (ii) To investigate the importance of bendability of the target DNA during protein binding, two different ligands were synthesized (Figure 1b), which differ in their nucleotide compositions at positions (−4) and (−5) in the respective ORas sequences (Figure 1b). In ORas11, nucleotides (−4) and (−5) are guanosines. In ORas12 the corresponding nucleotides are substituted by inosines lacking the C-2 amino group (Figure 1b). Removal of the C-2 amino groups is expected to increase the tendency of conformational freedom due to a lower degree of interchain base pairing and steric hindrance between neighboring groups in the minor groove (*12*). These C-2 amino groups in ORas11 are not expected to interact with Cro directly because they are

† The work was financially supported by The Carl Trygger Foundation.

* Corresponding author [fax +46 (18) 55 8431; telephone +46 (18) 471 4992; e-mail mikael.widersten@biokem.uu.se].

[1] Abbreviations: EMSA, electrophoretic mobility shift assay; HTH, helix−turn−helix; scCro, single-chain Cro.

10.1021/bi049122k CCC: $27.50
Published on Web 08/28/2004

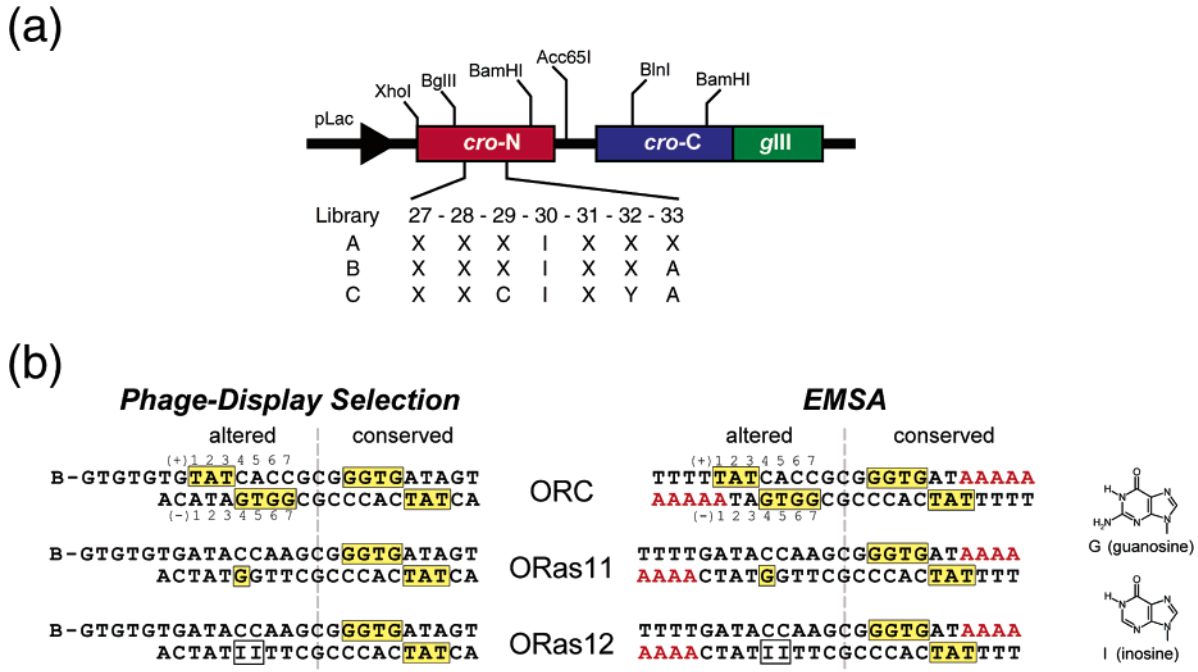

FIGURE 1: Schematic representation of scCro gene libraries and DNA ligands used during phage display-afforded affinity selection or in electrophoretic mobility shift assays (EMSA). (a) scCro gene library constructs. Random mutagenesis of codons for residues making base-specific contacts in the wild-type Cro protein was performed in the 5′-proximal *cro* gene (*cro*-N). Three different libraries, denoted A, B, or C containing different numbers of randomly (NNS) substituted codons, were constructed in which the *cro*-N genes (red) were linked with 24 nucleotides to the *cro*-C gene (blue) encoding the wild-type Cro protein. The *cro*-C gene was further linked to geneIII of phage M13. The proteins expressed were hence covalently linked via short peptide linkers. (b) Nucleotide sequences of dsDNA ligands used during selection by phage display ("B" denotes biotin) or in EMSA. Nucleotides on yellow background indicate positions making specific contacts with wild-type Cro (*15*). Boxed nucleotides in ORas12 indicate the positions of guanosine (in ORas11) to inosine replacements in ORas12. Nucleotides in red indicate positions for labeling with [α-^{32}P]-dATP by Klenow fill-in.

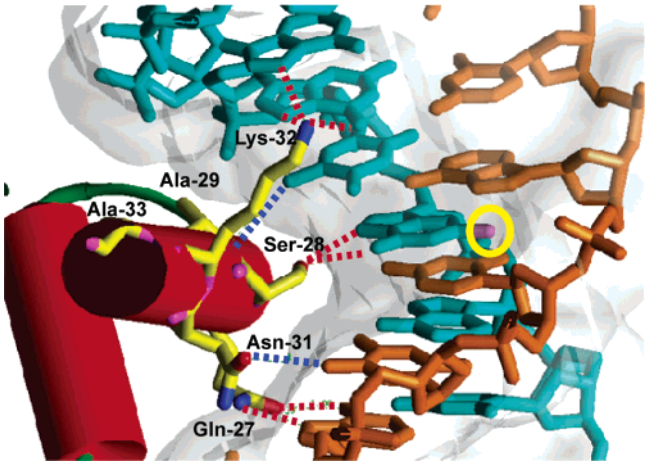

FIGURE 2: Image of the Cro/ORC complex. Specific protein−DNA base contacts are indicated by dotted lines (red, hydrogen bonds; blue, van der Waals interactions). Ser-28 interacts via its side-chain hydroxyl with O-6 and N-7 of G (-4) in ORC. The amino group at C-2 (purple within yellow circle) points toward the minor groove, away from the bound protein. Key: Protein α-helices are represented as red cylinders, loops in green, and side chains subjected to random mutagenesis in, carbon, yellow; oxygen, red; and nitrogen, blue. The strands of cocrystallized ORC DNA are shown in gold and cyan. Image produced in InsightII (Accelrys, San Diego, CA) from the atomic coordinates in 6CRO (*15*).

positioned within the minor groove, hence pointing away from the protein in the binding complex but presenting the same groups for hydrogen bonding and van der Waals contacts toward the major groove (Figure 2). Hence, if DNA bending is a component in sequence discrimination, different scCro variants are expected to bind the two ORas ligands with different relative affinities. Earlier results have shown that it is indeed possible to isolate phage clones displaying scCro variant proteins of distinct primary structures with binding properties favoring recognition for the different ORas ligands (*6*).

The present work describes the outcome of affinity selections of DNA-binding proteins from two new libraries of scCro variants. One of these new libraries was designed to test the significance of a protein sequence motif found in a majority of clones isolated after a previous selection toward ORas12 (*6*) ("B"-clones in Table 1). Selected protein variants have been analyzed for primary structure and characterized for their DNA-binding properties, targeting the structural determinants for their DNA sequence specificity.

EXPERIMENTAL PROCEDURES

Bacteria and Phages. Escherichia coli strains XL1-Blue (Stratagene, La Jolla, CA) and JM109(DE3) (Promega Corp., Madison, WI) were used as host cells during phage-display selection or in overexpression of scCro proteins, respectively. Phage M13K07 (Kan^R) was used as helper phage during phage-display expression.

Annealing, Labeling, and Purification of Double-Stranded DNA Ligand. Oligonucleotide concentrations in water were determined by UV absorption at 260 nm using calculated extinction coefficients. For annealing of double-stranded ligands, oligonucleotides were diluted to 100 μM in water followed by mixing at 1:1 molar ratios in 10 mM Tris-HCl and 50 mM NaCl to final concentrations of 10 μM. Mixtures were heated to 96 °C for 10 min followed by cooling to room temperature over a period of 140 min. Double-stranded DNA was stored at −20 °C until used.

A. Ligands Used in Phage Display Selection. Biotinylated and unbiotinylated double-stranded ORC (ORC+, TGT ATC ACC GCG GGT GAT AGT with ORC−, ACT ATC ACC CGC GGT GAT ACA; ORC+B, biotin-GTG TGT GTA TCA CCG CGG GTG ATA GT and ORC−), ORas11 (ORas1+B, biotin-GTG TGT GAT ACC AAG CGG GTG ATA GT with ORas1−1, ACT ATC ACC CGC TTG GTA TCA), and ORas12 (ORas1+B with ORas1−2, ACT ATC ACC CGC TTI ITA TCA) DNA ligands were obtained by annealing as described above.

B. Ligands for EMSA. Double-stranded DNA ligands (ORC, ORas11, ORas12) were prepared by mixing complementary oligonucleotides N−ORC+EMS, TTT TTA TCA CCG CGG GTG AT, with N−ORC−EMS, TTT TTA TCA CCC GCG GTG AT; 2-ORas1+EMS, TTT TGA TAC CAA GCG GGT GAT, with 2-ORas11−EMS, TTT TAT CAC CCG CTT GGT ATC; and 2-ORas1+EMS with 2-ORas12−EMS, TTT TAT CAC CCG CTT IIT ATC, as described above. Protruding oligo-dT ends were filled in with [α-^{32}P] dATP using Klenow fragment (Cloned, Exo-free, USB). Change of buffer and removal of unincorporated dATP were achieved by passing the labeled oligos through G-25 MicroSpin columns (Amersham Bioscience) equilibrated with 10 mM Tris-HCl, pH 7.4, 100 mM KCl, and 1 mM EDTA.

Phage Display Library Construction. Library A: The tandem *cro* gene construct encoding scCro8 (*5*) was inserted into pC3ΔNX (*13*). The 5′ cro gene was randomly mutated by PCR in codons 27−29 and 31−33 using primers cro8mut, TAA AGA TCT CGG CGT GTA TNN SNN SNN SAT CNN SNN SNN SAT CCA TGC CGG CCG AAA, and Cro8Lrev, GCC GCC AGA GCC ACC, as described previously (*6*). *Library C:* A gene library was constructed, as above, introducing random mutations in codons 27, 28, and 31. Codons 29 (TGC) and 32 (TAC) resulted in Ala-29 to Cys-29 and Lys-32 to Tyr-32 point mutations, respectively. Mutagenesis was performed by PCR as above with primers cro10mut, TAA AGA TCT CGG CGT GTA TNN SNN STG CAT CNN STA CGC CAT CCA TGC CGG CC, and Cro8Lrev. The phage display vector used in the construction of the C library was a derivative of pC3ΔNX, in which the unique *Spe*I site flanking the scCro gene at the 3′-end had been replaced by an *Aat*II site by silent mutagenesis (data not shown). Amplification and harvest of constructed phage display libraries of the different scCro constructs were performed as described earlier (*6*). For evaluation of library quality concerning sequence diversity, plasmid DNA from 24 transformants was isolated and the scCro gene regions were sequenced.

Affinity Selection by Phage-Display Expression. Library A: Phages displaying scCro randomly mutated in residues 27−29 and 31−33 were challenged for binding to biotinylated ORas11 or ORas12 DNA under binding conditions as described previously (*6*) for six selection rounds. All incubations and washing steps during the affinity selection cycles were performed on ice. *Library C:* Two parallel selections of phage-displayed scCro C-variants under binding conditions, as above, were conducted at either 22 °C or on ice for four rounds, with the exception that counterselection by addition of unbiotinylated ORC was not included. Enrichment during selection was assessed from the ratio of phages eluted after a completed selection cycle divided by the number of phages allowed to bind to the biotinylated ORas12 ligand. The scCro variant genes of picked isolates from rounds 2−6 (library A) or rounds 3 and 4 (library C) were analyzed by DNA sequencing.

DNA Binding Selectivity of Phage-Displayed ScCro Variants. Phage clones A12:415, C12:316, C12:401, C12:403, C12:405, C12:408, C12:414, and C12:404* were purified and amplified. Pure variant scCro8 phages (∼10^9 cfu) were subjected to a single round of affinity selection toward three different biotinylated DNA ligands, ORC, ORas11, and ORas12, as described previously (*6*). The relationship of the survival ratios obtained was used as a measure of binding specificity for the different DNA ligands tested.

ScCro Variant Protein Expression and Purification. Genes encoding selected scCro variants were subcloned into vector pETΔB, a *Bgl*II site-deficient derivative of pET21a(+) (Novagen). Confirmation of successful subcloning was verified by DNA sequencing. The resulting plasmid constructs were denoted pETscCro<library:clone#>. The expression plasmids were transformed into *E. coli* JM109 (DE3) for protein synthesis by electroporation.

A. Analytical-Scale Expression. Selected clones of *E. coli* JM109 (DE3) harboring expression plasmids with scCro variants were inoculated in 2TY containing 100 μg of ampicillin/mL and grown overnight at 30 °C. The overnight cultures were diluted 1:50 into 25 mL of 2TY containing 50 μg of ampicillin/mL. Cultures were grown at 30 °C until mid-log phase was reached (OD_{600} of 0.7−1) when scCro overexpression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside. Incubation was continued for 3 h when cells were harvested by centrifugation. Bacteria were resuspended in 0.5 mL of 20 mM sodium phosphate, pH 7.4, containing protease inhibitor cocktail (Roche) and subsequently lysed by ultrasonication. Insoluble debris was removed by centrifugation at 16000*g* for 15 min. Finally, glycerol and 2-mercaptoethanol were added to final concentrations of 50% (v/v) and 12.5 mM, respectively. Total protein content of lysates was determined according to the method of Bradford using a commercial reagent (Bio-Rad, Richmond, CA). The relative concentration of scCro protein variants in the lysates was estimated by Western blotting of lysates using polyclonal rabbit anti-scCro antibodies (AgriSera, Vännäs, Sweden). Immunocomplexes were detected by enhanced chemiluminiscense. Final lysate preparations were stored at −20 °C until used in EMSA.

B. Preparative-Scale Expression and Purification. Growth and expression were performed as described for small-scale cultures with the exception that culture volumes were 1.5 L per expressed variant. The crude lysate was desalted by gel filtration through Sephadex G-25, equilibrated with 50 mM sodium−phosphate, pH 7.5. The G-25 eluate was subsequently cleared either by ultracentrifugation at 100000*g* for 1 h or by filtration through an 0.8/0.2 μm Acrodisc syringe filter (Pall, Ann Arbor, MI). The desalted lysate was loaded at 1 mL/min onto a 5 mL column prepacked with SP-HiTrap cation exchanger resin (Amersham Biosciences). The column, pre-equilibrated in 50 mM sodium−phosphate and 50 mM NaCl, pH 7.5, was washed in the same buffer at 1 mL/min until baseline absorbance was reached. Bound protein was eluted with a linear gradient, 4 mM/mL, of NaCl from 50 mM to 0.7 M. Fractions containing scCro protein were pooled. Buffer was changed to 10 mM Tris-HCl, pH 7.4, and 100 mM KCl (EMSA binding buffer, "BB") by passing

through a PD10 column (Amersham Biosciences). Protein stocks for EMSA studies were subsequently diluted in BB containing 50% (w/v) glycerol and 2 mM dithiothreitol. Proteins were stored at −20 °C until analyzed. Protein concentration was determined by UV absorbance of the solutions using calculated molar extinction coefficients at 276 nm: scCro, 7800 M^{-1} cm^{-1}; A12:407, C12:408, C12:404*, 10150 M^{-1} cm^{-1}.

Electrophoretic Mobility Shift Assay (EMSA). A. Binding Competition Experiments. Equal volumes of diluted scCro protein from diluted bacterial lysates and ^{32}P-labeled ORC, ORas11, or ORas12 were mixed to 100 nM (protein) and 20 pM (DNA) final concentrations in the presence of various concentrations (0.5−450 μg/mL) of unspecific competitor DNA (sonicated salmon sperm DNA) and incubated in 10% (w/v) glycerol, 0.06% (w/v) Nonidet-P40, 1 mM dithiothreitol, 10 μg/mL BSA, and 0.5 mM EDTA in BB, on ice for 60 min. Samples of 20 μL were loaded on a 10% acrylamide gel containing 0.5 × TBE and 50 mg/mL of glycerol running at 16 V/cm, at 4 °C. The gel was prerun for 20 min prior to sample loading. After loading, the gel was run for an additional 20−40 min at 32 V/cm. Following electrophoresis the gel was dried, and separated radioactivity was analyzed by phosphorimaging (Fuji BAS-2500). The fraction of formed scCro/DNA ligand complex was expressed as percent complex of total amount of ligand DNA and plotted versus the competitor DNA concentration. Measures of the inhibition were expressed as competitor concentrations resulting in 50% (IC_{50}) or 90% (IC_{90}) inhibition of complex formation.

B. Determination of Dissociation Constants. Serial dilutions of purified protein were performed prior to mixing with 1:1 volumes of labeled DNA ligands resulting in the same concentrations of additives as described above and 10 μg/mL sonicated salmon sperm DNA. Equal volumes of diluted protein and diluted DNA operator were mixed and incubated on ice for 60 min. Electrophoresis and analysis were performed as described above. The fractional binding of scCro/DNA complex was determined after phosphorimager analysis. Dissociation constants were determined after the fitting of a 1:1 binding isotherm by nonlinear regerssion using program SIMFIT (url:http://www.simfit.man.ac.uk) to the experimental data.

Structure Modeling. The tertiary structures of scCro and three different variants, A12:407, B12:407, and C12:405, were modeled as follows. *(1) Modeling of ScCro:* The AGTGGSGG peptide linker connecting the two subunits in scCro was introduced using the program Biopolymer (Accelrys, San Diego, CA) and the atomic coordinates for the crystal structure of wild-type Cro in 5CRO (*14*). The crude scCro model was soaked in a 5 Å layer of water and subsequently relaxed by energy minimizations of, initially, the solvent only, followed by protein hydrogens, side chains, and finally the whole molecular system using the steepest descent algorithm in Discover (Accelrys). After this relaxation, a conjugate gradient minimization was conducted for 1000 iterations. Following energy minimization the molecular dynamics of the soaked scCro was simulated until the total energy of the system reached equilibrium (∼10 ps) at 283 K. An average equilibrium structure was calculated and finally energy-minimized by conjugate gradient until convergence was reached. *(2) Modeling of Variant ScCro Proteins:* Side-chain replacements in positions 27, 28, 29, 31, and 32 were introduced into the relaxed crude model of scCro8 using Biopolymer. The subsequent calculations were performed as described for scCro. Comparison of the modeled structures was performed by superimposition of the polypeptide backbone heavy atoms using InsightII (Accelrys).

RESULTS

Library Construction and Phage-Display Selection. Library A: The first constructed library was designed for phage-display selection of scCro variants with random substitutions in residues 27−29 and 31−33 (Figure 1a). In wild-type Cro Gln-27, Ser-28, Ala-29, Asn-31, and Lys-32 make specific contacts with nucleotides in the target ORC DNA (*15*) (Figure 2). Ala-33 does not interact with bound DNA directly but was included in the mutagenesis scheme due to its vicinity to the DNA-binding motif. The number of individual scCro-phage clones in the final library was titrated to be 10^7, corresponding to a representation of 1% of all possible gene combinations after random mutagenesis using NNS codons (N = G, A, T, or C and S = G or C) ($32^6 \approx 10^9$). The fraction of wild-type scCro genes, a consequence of vector self-ligation, was <1% as calculated from titration of vector-only transformed host cells. During selection, phages displaying wild-type scCro were identified after rounds 2 (1 of 32 sequenced clones) and 3 (1/16) of the ORas12 selection, but was absent in isolates after rounds 4−6 from selection toward either ligand (data not shown). An unexpected feature of the deduced amino acid sequences found in clones analyzed from the naive library was an over-representation of Trp codons. This was a consequence of suboptimal variation in the codon usage caused by a slight over-representation of T and G nucleotides in the originally synthesized oligonucleotides used for random mutagenesis and library construction. TGG was found in 4 (at position 29) and 2 (at position 32) of 29 sequenced scCro genes from the unselected library. The enrichment of phages during selection toward ORas11 or ORas12 (Figure 1b) binding was monitored by calculating the survival ratios of phages retrieved after the different selection cycles. In both cases, a steady increase in the number of phages surviving selection was obtained up to round 4, when a moderate decline was observed. After two additional cycles, selection was halted and enriched clones from rounds 2−6 were analyzed for their gene sequences in the randomized regions.

The sequence data from isolates selected for binding to ORas11 revealed that two dominating phage clones had been enriched after the sixth selection round. Variant A11:402 ($S_{27}R_{28}T_{29}W_{31}Y_{32}A_{33}$) was first identified as making up 50% (8 of 16 sequenced clones) of the population after four rounds of selection. After the sixth round, however, variant A11:601 ($T_{27}S_{28}Q_{29}W_{31}A_{32}M_{33}$) appeared and dominated this population, being present in six of eight analyzed clones, whereas the remaining two clones were identified as A11:402. After selection toward ORas12, one clone (A12:415) carrying a scCro variant with the sequence $R_{27}S_{28}C_{29}Y_{31}N_{32}W_{33}$ totally dominated the enriched population being found in 26 of 27 sequenced isolates after six selection rounds (Table 1). This variant was first detected in isolates after the fourth cycle as 1 of 16 analyzed clones but had already after the subsequent cycle been amplified to dominate the population

Table 1: Deduced Amino Acid Sequences of ScCro8 Variant Clones Isolated after Phage-Display Selection toward Biotinylated ORas12 DNA[a]

scCro protein	occurrence (ratio)	residue position					
		27	28	29	31	32	33
wild-type		*Q*	*S*	*A*	*N*	*K*	*A*
library A							
A12:407	1/74	R	Q	C	R	Y	A
A12:415	36/74	R	S	C	Y	N	W
library B							
B12:401	3/45	T	Q	C	D	Y	*A*
B12:405	2/45	A	Q	C	Q	Y	*A*
B12:406	1/45	R	E	C	A	Y	*A*
B12:407	4/45	D	Q	C	T	Y	*A*
B12:416	3/45	R	E	C	K	Y	*A*
B12:421	1/45	S	G	C	R	Y	*A*
B12:501	1/45	M	P	C	T	Y	*A*
B12:508	1/45	E	P	C	S	Y	*A*
library C (0 °C)							
C12:302	1/29	I	F	*C*	C	*Y*	*A*
C12:303	1/29	K	C	*C*	R	*Y*	*A*
C12:304	3/29	G	I	*C*	F	*Y*	*A*
C12:305	1/29	T	G	*C*	L	*Y*	*A*
C12:309	1/29	G	P	*C*	Q	*Y*	*A*
C12:313	2/29	R	I	*C*	C	*Y*	*A*
C12:316	1/29	S	A	*C*	R	*Y*	*A*
C12:401	7/29	V	A	*C*	R	*Y*	*A*
C12:402	1/29	Q	L	*C*	A	*Y*	*A*
C12:403	3/29	F	K	*C*	C	*Y*	*A*
C12:405	5/29	P	S	*C*	L	*Y*	*A*
C12:408	1/29	D	R	*C*	L	*Y*	*A*
C12:412	1/29	L	L	*C*	A	*Y*	*A*
C12:414	1/29	W	Q	*C*	S	*Y*	*A*
library C (22 °C)							
C12:301*	2/28	R	I	*C*	C	*Y*	*A*
C12:302*	1/28	W	L	*C*	F	*Y*	*A*
C12:303*	2/28	F	C	*C*	L	*Y*	*A*
C12:305*	1/28	S	A	*C*	R	*Y*	*A*
C12:307*	4/28	P	S	*C*	L	*Y*	*A*
C12:309*	3/28	W	Q	*C*	S	*Y*	*A*
C12:312*	7/28	G	I	*C*	F	*Y*	*A*
C12:401*	3/28	V	A	*C*	R	*Y*	*A*
C12:404*	2/28	Q	R	*C*	L	*Y*	*A*
C12:406*	1/28	D	R	*C*	L	*Y*	*A*
C12:415*	1/28	I	F	*C*	C	*Y*	*A*

[a] Italicized letters indicate residue positions not subjected to mutagenesis. *, phage display selection performed at 22 °C. Sequences from library B are from ref *6*.

(9 of 16 analyzed isolates after round 5). Apart from these strongly enriched variants, no sequence consensus could be deduced after either the ORas11 or ORas12 selections. The over-representation of Trp codons observed in the unselected library remained in several of the isolates after completed selection. Because Trp is a rarely observed residue in naturally occurring DNA-binding protein motifs (*16*, *17*), this over-representation was judged as a possible artifact caused by a skewed library codon usage. It has, however, been demonstrated that a Trp residue at position 33 (i.e., an Ala33Trp mutation) stabilizes the monomeric structure of Cro (*18*), which suggests that Trp-33 in the $R_{27}S_{28}C_{29}Y_{31}N_{32}W_{33}$ variant may have been selected as a consequence of promoting improved stability of the folded state as compared to competing variants.

Library B: To improve both the codon variability and the coverage of possible variants in the library, a new construct (library B) was made (Figure 1a). To limit the number of possible gene combinations in this library, codon 33 was omitted from mutagenesis. An analysis and affinity selection of scCro variants from library B have been reported earlier (*6*). A strong enrichment of clones carrying an $X_{27}X_{28}C_{29}X_{31}Y_{32}$ motif was observed after selection for ORas12 binding. Deduced amino acid sequences of enriched clones are given in Table 1 for comparison.

Library C: To probe the role of the enriched $X_{27}X_{28}C_{29}$-$X_{31}Y_{32}$ motif, library C was constructed (Figure 1a). The rationale behind the design was to preserve the coselected Cys-29 and Tyr-32 codons while allowing for random substitutions at positions 27, 28, and 31. The final library, after subtraction of vector background (12%), was titered to contain 7.9×10^4 individual member clones, covering all of the possible genetic diversity generated by three randomly mutated codons using the NNS strategy ($32^3 \approx 3.3 \times 10^4$).

To test the possibility to directly select for scCro-phage variants with improved specificity rather than overall affinity, two parallel selections for binding toward biotinylated ORas12 at either 0 or 22 °C were performed. According to calorimetric studies of Cro−DNA association, specific recognition of a ligand DNA is favored by a decrease in enthalpy, whereas nonspecific interactions are primarily entropy-driven (*19*). This enthalpic contribution to the change in free energy in specific interactions is increased with elevated temperature as compared to the entropic term. This would increase the demand for specific interactions at higher temperatures to maintain high affinity in binding.

The outcome of the phage enrichment during selection showed an initial decline in the survival ratio caused by more extensive washing introduced at this point, followed by an increase after round three with a plateau being reached after round four. Similar behavior was observed for the two parallel selections. Because no further clonal enrichment was observed after the third selection cycle, and considering the limited number of variants in the initial library, the selection was halted after cycle four.

An indication of successful affinity selection was the finding that only a limited number of different phage clones had been enriched (Table 1). Although the parallel selections had been performed at different temperatures, the amplified scCro-phage populations were overall very similar. Phage clones expressing variant scCro proteins with residue replacements that had the potential to interact with bound DNA through specific hydrogen bonds were chosen for further functional characterization.

DNA-Binding Selectivity of Phage-Displayed ScCro Variants. Variant A12:415 ($R_{27}S_{28}C_{29}Y_{31}N_{32}W_{33}$) together with a selected number of different scCro−phage C clones was challenged for binding to three different biotinylated DNA ligands, ORC, ORas11, and ORas12. The survival ratios obtained for each clone with different DNA ligands after one selection round were treated as a measure of binding affinity and the ratios of survival ratios as a measure of specificity for the different DNAs tested (Figure 3). All of the phage clones tested, except C12:404* ($Q_{27}R_{28}C_{29}L_{31}Y_{32}$), showed a 2−5-fold higher preference for ORas12 over the closely related ORas11 ligand. When the apparent specificity was compared for ORas12 versus ORC, however, only three of the tested variants, C12:404*, C12:408 ($D_{27}R_{28}C_{29}L_{31}Y_{32}$), and C12:414 ($W_{27}Q_{28}C_{29}S_{31}Y_{32}$), showed preference for ORas12 over ORC. These findings are in agreement with earlier results from the analysis of an ORas12 selection of clones isolated from the B library (*6*) (Figure 3), where again

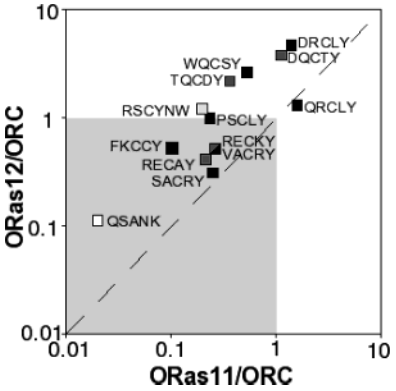

FIGURE 3: DNA binding specificity of variant scCro-phages. Phages presenting different scCro variants were challenged for binding to either ORas11, ORas12, or ORC during one selection cycle. The values of the obtained survival ratios were divided to generate a ratio-of-ratios, a measure for preferred interaction between different scCro-phages and a given DNA. The diagonal line indicates the borderline for discrimination between ORas11 and Oras12. Values for phage clones isolated from library B (*6*) have been included for comparison. White indicates wild-type scCro; light gray, library A; dark gray, library B; black, library C. See Table 1 for clone numbering. Clones in the shaded area displayed a preference in recognition for ORC.

the majority of scCro phage clones discriminated between ORas12 and ORas11 but only clones B12:401 ($T_{27}Q_{28}C_{29}D_{31}Y_{32}$) and B12:407 ($D_{27}Q_{28}C_{29}T_{31}Y_{32}$) favored binding to ORas12 over ORC.

Variant A12:407. The identification of the $X_{27}X_{28}C_{29}X_{31}Y_{32}$ motif as apparently favoring binding to ORas12 motivated a new sequence analysis of the scCro phage clones selected toward ORas12 from the A library. This new search identified variant A12:407 ($R_{27}Q_{28}C_{29}R_{31}Y_{32}A_{33}$). This phage clone had survived four rounds of affinity selection but was outcompeted during later rounds, primarily by the dominating A12:415 clone. The randomly mutated codon 33 in this variant had retained a codon for the wild-type Ala residue, resulting in a selected amino acid sequence fitting the $X_{27}X_{28}C_{29}X_{31}Y_{32}$ consensus. The A12:407 protein was hence selected for further studies on its binding properties.

Binding Properties of ScCro Variants. To analyze the binding affinity of the proteins directly, proteins were expressed as free in solution, not linked to phage, and directly challenged for binding to unbiotinylated radiolabeled DNA ligands. Formed protein/DNA complexes were detected and quantified by electrophoretic mobility shift assays (*20*) (EMSA) in combination with phosphorimager analysis.

A. Competition Studies. ScCro variants were overexpressed in *E. coli*. The individual levels of expression was determined by Western blotting (Figure 4a), and the lysates were diluted to equal concentrations of scCro proteins. All proteins tested behaved as readily soluble and retained DNA-binding activity for weeks when stored at −20 °C in the presence of 50% glycerol. Variant C12:403 ($F_{27}K_{28}C_{29}C_{31}Y_{32}$), however, demonstrated a tendency to form aggregates insoluble also under the denaturing conditions of SDS−PAGE (Figure 4a).

Binding competition by increasing amounts of unrelated DNA was detected as a decrease in complex formation between the various scCro proteins and the different ORas ligands (Figure 4b). The degree of sensitivity to competitor DNA, expressed as IC_{50} or IC_{90} values, was interpreted as an indirect measure of relative binding strength between the proteins and target DNA in the detected complexes (Table 2).

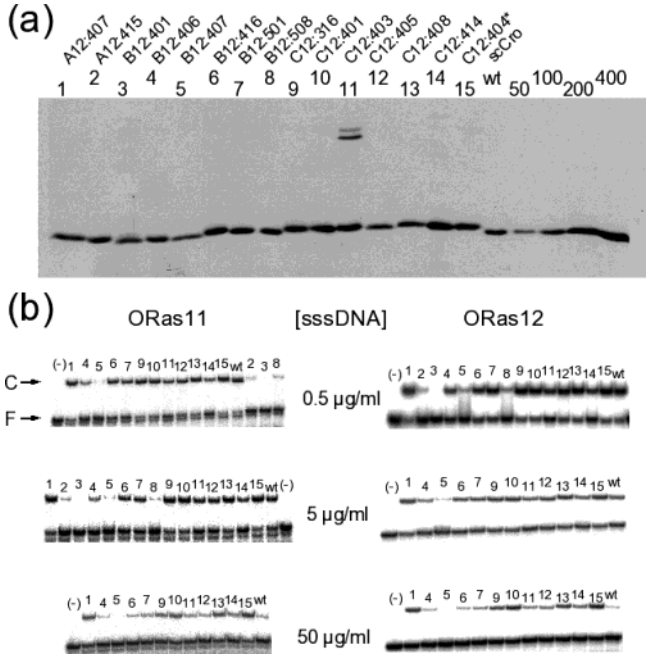

FIGURE 4: Competition by unrelated DNA to scCro protein/ORas complex formation. (a) Western blot analysis of bacterial lysates expressing scCro variants. Lanes are labeled according to clone number. The four right-most lanes were loaded with 50, 100, 200, and 400 ng of purified scCro (wild-type), respectively, for calibration of scCro protein content in lysates. (b) EMSA in the presence of 100 nM scCro variants and <20 pM radiolabeled ORas11 (left) or ORas12 (right). Competition by addition of increasing amounts of unrelated (sonicated salmon sperm) DNA (sssDNA) was estimated from detection of the fraction of formed scCro protein/ORas complex (band C, "complex") as compared to the total amount of added ORas ligand (sum of bands C + F, "free"). The concentration of unrelated competitor DNA is indicated.

Table 2: Inhibition of ScCro/DNA Ligand Complex Formation by Unrelated DNA

	DNA ligand			
	IC_{50}[a] (µg/mL)		IC_{90}[a] (µg/mL)	
scCro protein	ORas11	ORas12	ORas11	ORas12
QSANKA (wt)	7	3	90	41
RQCRYA (A12:407)	25	6	410	340
RSCYNW (A12:415)		<0.5	<0.5	3
TQCDY (B12:401)			<0.5	<0.5
RECAY (B12:406)	<0.5	<0.5	40	37
DQCTY (B12:407)		<0.5	<0.5	4
RECKY (B12:416)	2	1	48	42
MPCTY (B12:501)	<0.5	2	170	46
EPCSY (B12:508)			3	<0.5
SACRY (C12:316)	19	4	330	250
VACRY (C12:401)	28	7	410	380
FKCCY (C12:403)	<0.5	<0.5	230	47
PSCLY (C12:405)	1	3	130	94
DRCLY (C12:408)	29	7	390	320
WQCSY (C12:414)	<0.5	<0.5	270	160
QRCLY (C12:404*)	48	17	>450	430

[a] IC_{50} and IC_{90} are the concentrations of competitor DNA (sonicated salmon sperm DNA) that inhibit complex formation between scCro variants and target DNA to 50 or 90%, respectively.

Variant C12:404* ($Q_{27}R_{28}C_{29}L_{31}Y_{32}$) showed the least sensitivity for competition by unrelated DNA in the presence of either ORas11 or ORas12. Other variants with comparably high IC_{50} values in the presence of either of the ORas ligands

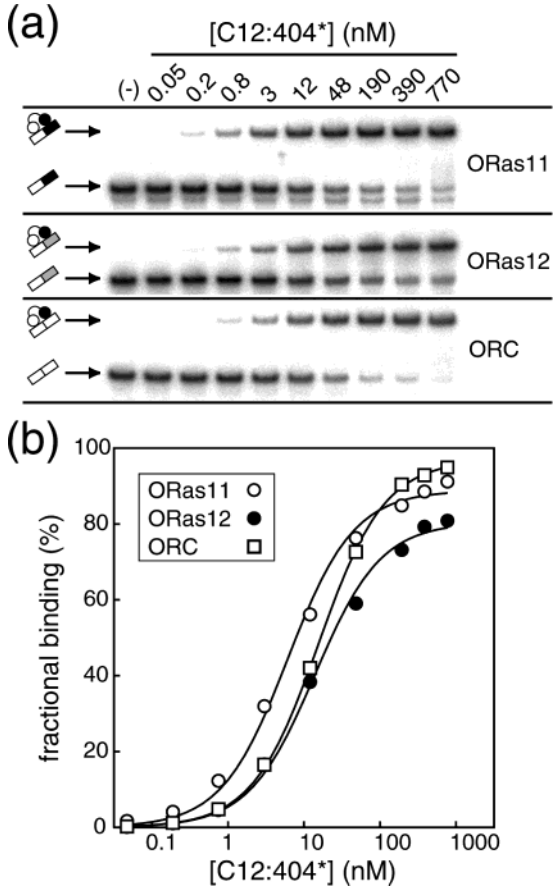

FIGURE 5: Titration of complex formation between scCro variants, exemplified by C12:404*, and DNA ligands. (a) Titration by increasing concentrations of protein in the presence of <20 pM dsDNA ligand resolved by EMSA. Band identities are indicated by cartoons. (b) Least-squares fit (solid lines) of % fractional binding = $[\text{protein}]_{\text{total}} \times B_{\text{max}}/(K_D + [\text{protein}]_{\text{total}})$ to the experimental data, where % fractional binding was calculated from (amount of radioactivity detected in protein/DNA complex)/(radioactivity in free DNA + protein/DNA complex) × 100 and B_{max} is maximum complex formation. (□) ORC; (○) ORas11; (●) ORas12.

were C12:408 ($D_{27}R_{28}C_{29}L_{31}Y_{32}$), C12:401 ($V_{27}A_{28}C_{29}R_{31}Y_{32}$), A12:407 ($R_{27}Q_{28}C_{29}R_{31}Y_{32}A_{33}$), and C12:316 ($S_{27}A_{28}C_{29}R_{31}Y_{32}$). In general, the variants selected from library C showed less sensitivity to the competitor as compared to variants isolated from library B, indicating that the presence of the $X_{27}X_{28}C_{29}X_{31}Y_{32}$ core sequence in all clones of this library facilitated enrichment of higher affinity clones.

B. Dissociation Constants of ScCro Protein/DNA Complexes. Dissociation constants were determined after fitting a 1:1 binding isotherm to the experimental data (Figure 5). The relative selectivity favoring ORas11 over ORC, observed in variant C12:404* in the phage survival ratio assays, was confirmed. The K_D value for the ORas11 complex was 3-fold lower than the value for ORC complex dissociation (Table 3). Comparing the affinity for the different ORas ligands also confirmed the previously observed relationship; ORas11 is the preferred DNA for this variant. The binding properties of the structurally related C12:408 were similar to those of C12:404*, albeit displaying slightly lower affinity for the ORas ligands. When tested for its affinity toward ORC, however, variant C12:408 showed a tendency to from higher molecular weight aggregates, observed as supershift in EMSA. The same observation was made with other variants in bacterial lysates when incubated with ORC (data not shown).

Table 3: Equilibrium Dissociation Constants of ScCro/DNA Complexes[a]

	K_D (nM)		
scCro protein	ORas11	ORas12	ORC
QSANKA (wt)	147 ± 9 (1)	228 ±10 (1)	0.014 ± 0.003[b] (1)
RQCRYA (A12:407)	50 ± 3 (1)	23 ± 1 (1)	73 ± 3 (1)
DRCLY (C12:408)	15 ± 4 (2)	16 ± 5 (2)	31 ± 17[c] (2)
QRCLY (C12:404*)	5.7 ± 0.9 (3)	13 ± 1 (2)	16 ± 1 (4)

[a] Dissociation constants were determined after fitting a binding equation for 1:1 binding to the experimental data. Numbers in parentheses indicate the number of EMSA experiments performed to determine K_D. [b] Determined in the presence of 1 pM ORC. [c] K_D was determined by linear regression analysis using the least-squares fit $[\text{DNA}]_{\text{free}}/[\text{complex}] = K_D \times 1/[\text{protein}]_{\text{total}}$ to the experimental data.

Variant A12:407, selected from the A library for binding to ORas12 and putatively assigned as capabable of ORas recognition, did indeed demonstrate relatively high affinity for these ligands (Table 3). Although this variant displayed a lower overall affinity toward all three ligands tested, A12:407 retained the selectivity toward the ORas12 ligand to which it was originally selected for binding to. The relative specificity was the highest observed (3.2-fold) among the tested variants. Hence, as judged from the K_D values, the two Arg residues in the recognition helix of A12:407 do not appear to have been selected for as a consequence of affording unspecific electrostatic interactions with DNA backbone phosphates.

Overall, when the ligand specificity of isolated proteins, not linked to phage, was tested in binding assays, the previously observed selectivity between the ORas ligands was in principle lost (Table 2). A possible reason for these discrepancies may be the differences in structure of the ligands used during phage-display selection as compared to those applied in the EMSA studies (Figure 1b). The radiolabeling step, which introduces eight extra A·T base pairs in the EMSA ligands, increases the DNA melting temperature, which is anticipated to cause differences in flexibility. Hence, if differences in conformational properties of the biotinylated ORas ligands were a determinant during selection, a change could affect the binding interactions between protein and DNA.

Structure Modeling. Modeling of three-dimensional structures of wild-type scCro and variants thereof was conducted as an approach to assess the structural consequenses by the replacements in helix 3. Via in silico mutagenesis, energy minimizations, and molecular dynamics simulations, modeled structures were obtained. In general, the overall modeled structures of these variants were highly similar as compared to the model of the wild-type structure, with a higher degree of similarity when the unmutated "subunit" was compared. The most striking difference observed in all three variants was the change in conformation of Phe-58 in the unmutated subunit, which contributes to the hydrophobic core of the varied subunit through intersubunit interactions with residue 29.

DISCUSSION

The biochemistry of protein/DNA interaction works at several levels: (i) sequence complementarity, that is, weak

interactions between chemical groups such as hydrogen bonding and electrostatic interactions (*21*−*23*); (ii) surface (shape) complementarity (*24*) for efficient desolvation and realization of a maximum number of van der Waals interactions; and (iii) macromolecular flexibility, DNA bending, and structural changes in the protein, upon complex formation (*15*, *25*). In an attempt to address these issues in bulk, we have explored the route of functional selection of variant proteins with new properties that afford DNA binding to unnatural target sequences. Through random substitutions of residues critical in specific DNA recognition, combined with an affinity-based selection targeting predesigned DNA ligands, a limited number of variants with recognition preferences distinct from the wild-type counterpart have been identified and analyzed for their structure/function relationships.

The gene construct encoding the single-chain Cro variants consists of two identical *cro* genes linked in tandem via an eight-codon linker. For library generation the 5′-proximal gene (*cro*-N) was subjected to randomized mutagenesis, whereas the 3′-gene module (*cro*-C) was left unmutated. This gene module was coupled via another linker to the phage gene III. The corresponding encoded proteins were therefore asymmetric with an N-terminal domain encoding a variant sequence, whereas the C-terminal domain retained the wild-type structure (Figure 1a). In parallel, one half-site of the ORas DNA ligands was derived from ORC, a high-affinity ligand recognized by Cro, whereas the sequence of the other half-site was altered in order to probe protein libraries for scCro variants of adequate binding affinities. Anticipated advantages with such a design for selection were that (i) fewer numbers of randomized codons allowed inclusion of a larger portion of all possible variants in the constructed libraries and (ii) low basal levels of affinity between the DNA ligand and a given variant would exist, lowering the risk of false negatives caused by the law of mass action of nonfunctional proteins disturbing the selection procedure. An expected disadvantage, however, would be a case in which the native domain (the C-terminal "subunit" in the scCro selections) of the variant protein binds with an affinity substantially higher than that of the substituted protein portion. The binding contribution of the variant domain may be masked, and hence the significance of its structure in the contribution to overall binding affinity would be difficult to assess. The finding that the majority of selected variants showed poor or even negative discrimination between the ligand selected against, ORas12, as compared to the wild-type ligand ORC may be a reflection of this notion (Figure 3). Similar results have also been reported in a related study applying in vivo screening for altered specificity of a single-chain derivative of the phage 434 cI DNA-binding domain (*26*).

During initial design of the ligands, ORas11 and ORas12 were expected to interact with scCro variants of similar primary structure; both of these ligands presumably present the same chemical groups to a bound scCro protein (Figure 2). The finding that not a single identical clone was selected from the ORas11 and ORas12 ligands implied that intrinsic differences in the flexibility of these ligands, caused by the guanosine to inosine substitutions in ORas12, were important factors in protein/ORas discrimination. This assumption was supported by the survival ratio assays (Figure 3) in which scCro-phage clones isolated after selection toward ORas12 favored binding of ORas12 over ORas11. The reverse behavior has previously been observed for ORas11-selected phage clones from library B (*6*).

A common feature of the primary structure of ORas12-favoring variants selected from the B library was the occurrence of an $X_{27}X_{28}C_{29}X_{31}Y_{32}$ sequence motif. These combined results implied a role for this Cys-29/Tyr-32 residue pair in recognition of ORas12. In wild-type Cro the same positions are occupied by Ala-29 and Lys-32. Ala-29 makes van der Waals interactions through its peptide backbone amide with the methyl group of T(−5) in ORC (*15*) (Figure 2). Because this base interaction involves a repetitive unit of the protein, conservative residue changes in position 29 were not a priori expected to result in major changes in DNA specificity. Introduction of Cys at position 29, however, appears to be one of the most important features favoring selection of scCro phage variants when challenged for binding to biotinylated ORas12. The majority of Cys-29-containing clones are also accompanied by a Tyr residue in position 32, suggesting a structural linkage of these two residues: a motif-based selection rather than an independent readout of individual residues. In wild-type Cro the corresponding Lys-32 makes three hydrogen-bond interactions through its side-chain ammonium with the bases of both G(−6) and G(−7) (Figure 2). Because these nucleotides are adjacent to T(−5), it is reasonable to assume that the nucleotide triplet from (−5) to (−7) is primarily recognized by amino acid residues 29 and 32 as a motif and, consequently, changes in these nucleotide positions would require a new protein sequence motif for specific interaction to occur. Although structurally unrelated, a motif, or syllable-based protein−DNA recognition, has also been concluded for zinc-finger proteins (*9*, *27*).

In ORas11 and ORas12, nucleotides (−5) to (−7) are GTT or ITT, respectively. In none of the clones analyzed from libraries selected toward the ORas11 ligand (77 sequenced clones) have the Cys-29/Tyr-32 pair of residues been observed (*6*) (M. T. I. Nilsson and M. Widersten, unpublished data), suggesting the Cys-29/Tyr-32 motif as favoring recognition of ITT over GTT. If the level of recognition is through direct interaction between these residues and the nucleotides or if the differences are due to other structural alterations, promoting new specific contacts is at present not clear.

To study the structural consequences of the Ala-29 to Cys replacement, calculated molecular models were constructed. Variants A12:407 ($R_{27}Q_{28}C_{29}R_{31}Y_{32}$), B12:407 ($D_{27}Q_{28}C_{29}T_{31}Y_{32}$), and C12:405 ($P_{27}S_{28}C_{29}L_{31}Y_{32}$) were chosen as representatives for Cys-29/Tyr-32-containing variants carrying a range of different substitutions in the mutated positions. As discussed above, the side-chain methyl group of Ala-29 is in van der Waals contact with the phenyl ring of Phe-58 protruding from the other domain ("Cro-C subunit"). This interaction contributes to the hydrophobic core and is important for stabilizing the folded state of the native protein (*28*, *29*). To fit the bulkier Cys side chain of a variant, the modeled proteins take on an alternative overall local topology. If there is a functional role for this observed change in topology during protein−DNA interactions, a more detailed structural and functional analysis would be required.

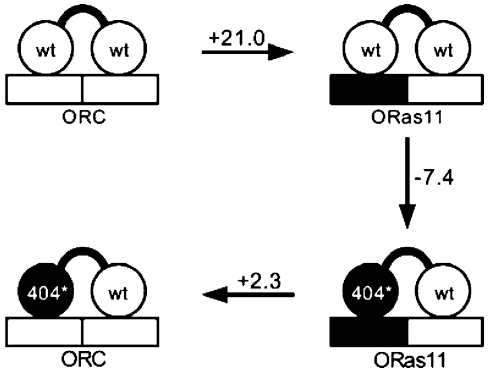

FIGURE 6: Changes in free energy between different scCro protein/DNA complexes. Replacing the wild-type N-terminal scCro domain with that of variant C12:404* restores 7.4 kJ/mol of the binding energy lost (21 kJ/mol) when changing DNA ligands from ORC to ORas11. The energy loss when substituting ORC for ORas11 as ligand bound to C12:404* amounts to +2.3 kJ/mol. Values of $\Delta\Delta G$ were calculated from $\Delta\Delta G = -RT \ln K_{D1}/K_{D2}$, where $T = 273$ K, K_{D1} is the dissociation constant for the initial complex, and K_{D2} that for the final complex. (White sphere) wild-type scCro domain; (black sphere) variant C12:404* domain; (white box) ORC half-site; (black box) ORas11 half-site.

The free energy change upon complex formation of wild-type scCro and ORC, as calculated from the determined K_D values, is on the order of 60 kJ/mol of independent binding sites, that is, "subunits". This includes nonspecific interactions as well as specific base interactions. Replacing ORC with ORas11 lowers the free energy change upon complex formation by 21 kJ/mol (Figure 6). The numbers of base-specific interactions made by Cro bound to ORC, as deduced after analysis of an available X-ray structure (*15*), are seven hydrogen bonds and two van der Waals interactions (Figure 2). Assuming a similar structure of the scCro/ORC complex, interactions predicted to be lost in the scCro/ORas11 complex are five of the hydrogen bonds [bidental hydrogen bonds between Gln-27 and A(+2) and Lys-32 interacting with G(−6), G(−7)] and one van der Waals interaction [Asn-31 interacting with T(+3)]. Substituting residues 28, 29, 31, and 32 in the N-terminal "subunit" of scCro, to create variant C12:404* restores 7.4 kJ/mol of lost energy in the protein/ORas11 complex (Figure 6). Hence, these residue replacements (Arg-28, Cys-29, Leu-31, and Tyr-32) restore approximately one-third of the energy lost when the DNA ligand is changed from ORC to ORas11. Indications of whether the interactions between C12:404* and ORas11 are base-specific or unspecific can be deduced from comparisons of the $\Delta\Delta G$ values obtained when the ORas11 ligand in complex with C12:404* is replaced by ORC, which results in an energy loss of 2.3 kJ/mol (Figure 6). The difference between these $\Delta\Delta G$ values indicates that at least one interaction made between C12:404* and a bound DNA ligand is of a nonspecific nature, corresponding to −5.1 kJ/mol. To improve on the recognition specificity of a variant protein such as C12:404*, such nonspecific interactions would be targeted for further mutagenesis and characterization.

From the structure/function data compiled for all analyzed proteins, the sequence $X_{27}R/Q_{28}C_{29}X_{31}Y_{32}$ appears to accommodate affinity and/or specificity for the different DNA ligands. Variants A12:407, C12:404*, and C12:408 all contain this core sequence. C12:408 and C12:404*, which both carry an $X_{27}R_{28}C_{29}-\text{aliphatic}_{31}-Y_{32}$ motif, displayed higher binding affinities but showed lesser degrees of selectivity between the ORas12 ligand and ORC as compared to A12:407. It therefore appears that discrimination between ORC and ORas12 is better accommodated by a structure resembling an $X_{27}Q_{28}C_{29}X_{31}Y_{32}$ motif. The same core sequence can be found in variants B12:401 ($T_{27}Q_{28}C_{29}D_{31}Y_{32}$), B12:407 ($D_{27}Q_{28}C_{29}T_{31}Y_{32}$), and C12:414 ($W_{27}Q_{28}C_{29}S_{31}Y_{32}$). These proteins were among the clones that in the phage survival assays showed the highest degrees of selectivity in favor of ORas12 over ORC. Attempts to determine K_D values for B12:401 and B12:407, however, have been unsuccessful due to protein aggregation at the conditions required during the EMSA (data not shown).

In summary, the work describes the successful isolation of a repertoire of structurally related scCro variants with functional properties that can be linked to observed differences in corresponding primary structure.

ACKNOWLEDGMENT

The contribution of Mikaela Sandström during the initial stages of the project is gratefully acknowledged.

REFERENCES

1. Wintjens, R., and Rooman, M. (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes, *J. Mol. Biol. 262*, 294−313.
2. Rosinski, J. A., and Atchley, W. R. (1999) Molecular evolution of helix-turn-helix proteins, *J. Mol. Evol. 49*, 301−309.
3. Aravind, L., and Koonin, E. V. (1999) DNA-binding proteins and evolution of transcription regulation in the archaea, *Nucleic Acids Res. 27*, 4658−4670.
4. Huffman, J. L., and Brennan, R. G. (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif, *Curr. Opin. Struct. Biol. 12*, 98−106.
5. Jana, R., Hazbun, T. R., Fields, J. D., and Mossing, M. C. (1998) Single-chain lambda Cro repressors confirm high intrinsic dimer-DNA affinity, *Biochemistry 37*, 6446−6455.
6. Nilsson, M. T. I., Mossing, M. C., and Widersten, M. (2000) Functional expression and affinity selection of single-chain Cro by phage display: isolation of novel DNA-binding proteins, *Protein Eng. 13*, 519−526.
7. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, *Science 228*, 1315−1317.
8. Smith, G. P., and Petrenko, V. A (1997) Phage display, *Chem. Rev. 97*, 391−410.
9. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) Design and selection of novel Cys_2His_2 zinc finger proteins, *Annu. Rev. Biochem. 70*, 313−340.
10. Segal, D. J., Dreier, B., Beerli, R. R., and Barbas, C. F., III (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2758−2763.
11. Kim, J. G., Takeda, Y., Matthews, B. W., and Anderson, W. F. (1987) Kinetic studies on Cro repressor-operator DNA interaction, *J. Mol. Biol. 196*, 149−158.
12. Travers, A. A (1995) DNA bending by sequence and proteins, in *DNA-Protein Structural Interactions* (Lilley, D. M. J., Ed.), pp 49−75, Oxford University Press, Oxford, U.K.
13. Widersten, M., and Mannervik, B. (1995) Glutathione transferases with novel active sites isolated by phage display from a library of random mutants, *J. Mol. Biol. 250*, 115−122.
14. Ohlendorf, D. H., Tronrud, D. E., and Matthews, B. W. (1998) Refined structure of Cro repressor protein from bacteriophage λ suggests both flexibility and plasticity, *J. Mol. Biol. 280*, 129−136.
15. Albright, R. A., and Matthews, B. W. (1998). Crystal structure of lambda-Cro bound to a consensus operator at 3.0 Å resolution, *J. Mol. Biol. 280*, 137−151.
16. Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) Protein-DNA interactions: a structural analysis, *J. Mol. Biol. 287*, 877−896.

17. Nadassy, K., Wodak, S. J., and Janin, J. (1999) Structural features of protein-nucleic acid recognition sites, *Biochemistry 38*, 1999−2017.
18. LeFevre, K. R., and Cordes, M. H. J (2003) Retroevolution of λ Cro toward a stable monomer, *Proc. Natl. Acad. Sci. U.S.A. 100*, 2345−2350.
19. Takeda, Y., Ross, P. D., and Mudd, C. P. (1992) Thermodynamics of Cro protein-DNA interactions, *Proc. Natl. Acad. Sci. U.S.A. 89*, 8180−8184.
20. Garner, M. M., and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system, *Nucleic Acids Res. 9*, 3047−3061.
21. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) Amino acid−base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level, *Nucleic Acids Res. 29*, 2860−2874.
22. Luscombe, N. M., and Thornton, J. M. (2002) Protein-DNS interactions: amino acid conservation and the effects of mutations on binding specificity, *J. Mol. Biol. 320*, 991−1009.
23. Jayaram, B., McConnell, K., Dixit, S. B., Das, A., and Beveridge, D. L. (2002) Free-energy component analysis of 40 protein-DNA complexes: a consensus view on the thermodynamics of binding at the molecular level, *J. Comput. Chem. 23*, 1−14.
24. Yeh, C.-S., Chen, F.-M., Wang, J.-Y., Cheng, T.-L., Hwang, M.-J., and Tzou, W.-S. (2003) Directional shape complementarity at the protein-DNA interface, *J. Mol. Recognit. 16*, 213−222.
25. Dickerson, R. E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality, *Nucleic Acids Res. 26*, 1906−1926.
26. Simoncsits, A., Tjörnhammar, M.-L., Wang, S., and Pongor, S. (1999) Isolation of altered specificity mutants of the single-chain 434 repressor that recognize asymmetric DNA sequences containing the TTAA and TTAC subsites, *Nucleic Acids Res. 17*, 3474−3480.
27. Wolfe, S. A., Grant, R. A., Elrod-Erickson, M., and Pabo, C. O. (2001) Beyond the "Recognition code": structures of two Cys_2His_2 zinc finger/TATA box complexes, *Structure 9*, 717−723.
28. Mossing, M. C., and Sauer, R. T. (1990) Stable, monomeric variants of lambda Cro obtained by insertion of a designed beta-hairpin sequence, *Science 250*, 1712−1715.
29. Mollah, A. K. M. M., Aleman, M. A., Albright, R. A., and Mossing, M. C. (1996) Core packing defects in an engineered Cro monomer corrected by combinatorial mutagenesis, *Biochemistry 35*, 743−748.

BI049122K