UPTEC X 01 020
APR 2001

ISSN 1401-2138

MIKAELA SANDSTRÖM

Evolving optimisation of DNA binding in a novel Cro derived protein

Master's degree project

Molecular Biotechnology Programme
Uppsala University School of Engineering

UPTEC X 01 020 | **Date of issue 2001-04**

Author

Mikaela Sandström

Title (English)

Evolving optimisation of DNA binding in a novel Cro derived protein

Title (Swedish)

Abstract

Interactions between proteins and DNA are essential for biological life processes. However, the understanding of the sequence-specific and high-affinity interaction is incomplete.
Here, *in vitro* evolution was used to study the interactions between helix-turn-helix (HTH) proteins and DNA. A selection system based on Cro, a small HTH protein, was used to select for new DNA binding properties. A phage library was constructed where three of the known five DNA interacting residues were randomly mutated and two residues were subjected to site-directed mutagenisis. Selection was performed at two temperatures (4°C and room temperature) and with a new DNA ligand that is not a natural ligand for wt Cro. After four selection rounds some dominating clones were found. The specificity of these mutants was verified and two clones, DRCLY and WQCSY, were found to be specific for the new ligand. The DRCLY mutant was expressed and purified.

Keywords

Phage display, phage library, single-chain Cro (scCro), DNA ligands and selection.

Supervisors

Mikael Widersten
Department of Biochemistry, Uppsala University

Examiner

Birgitta Tomkinson
Department of Biochemistry, Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **23**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Evolving optimisation of DNA binding in a novel Cro derived protein

Svensk sammanfattning

Interaktioner mellan proteiner och DNA är livsnödvändiga för olika biologiska reaktioner. Dessa bindningar är specifika och starka, men förståelsen för varför är fortfarande ofullständig.

I detta projekt studerades bindningen mellan proteinet Cro och en DNA sekvens som Cro naturligt inte binder till (ORas12). Det är sedan tidigare känt vilka delar av Cro som binder till DNA. Genom att konstruera ett bibliotek av olika Cro proteiner, varierade i just de DNA bindande delarna, så kan ORas12 bindande varianter hittas. Därmed kan en del utrönas om vad det är som gör bindningen mellan proteinet och DNA så specifik och stark.

Bindningen studerades med hjälp av fagpresentation och selektion. Fager är virus som infekterar bakterier. De har vissa proteiner på ytan, s.k. ytproteiner. Genom att sätta in Cro DNA på vissa bestämda ställen i fagens DNA kan Cro presenteras tillsammans med ytproteiner på ytan. Detta utnyttjas då ett s.k. fagbibliotek konstrueras. Ett fagbibliotek består av en mängd fager som alla presenterar varianter av ett och samma protein. Fagbiblioteket utsätts sedan för selektion. Det innebär att biblioteket och DNA sekvensen av intresse (ORas12) blandas, sedan utsätts blandningen för tvätt. De olika fagpresenterade proteinerna kommer att binda olika bra (vissa inte alls) till DNA sekvensen. Endast de som binder bra kommer att sitta kvar efter tvätten och man säger att en selektionsrunda är gjord. Efter flera selektionsrundor kommer bara de varianter av proteinet som binder bra till ORas12 att finnas kvar.

Examensarbete 20 poäng inom Molekylär bioteknikprogrammet
Uppsala universitet, april 2001

TABLE OF CONTENTS

INTRODUCTION

Interactions between proteins and DNA are essential for biological life processes. Many different types of motifs in proteins are known to bind to DNA, e.g. basic leucin-zipper, zinc finger and helix-turn-helix motifs. The first DNA-interacting motif to be recognised was the helix-turn-helix (HTH) (Figure 1). It is constructed from two alpha helices connected by a short chain of amino acids, which constitutes the turn. The two helices are held at a fixed angle, primarily through non-polar interactions between side chain residues in the two helices. The carboxyl-terminal helix is called the recognition helix because it fits into the major groove of DNA. The amino acid side chains in the specific protein play an important part in the recognition of the DNA sequence (Stryer, 1995).

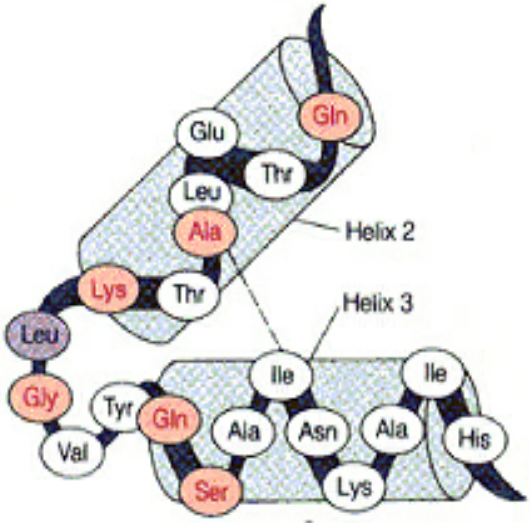

Figure 1. Figure showing a HTH motif with the two alpha helices connected by a turn. This motif is part of one subunit of the Cro protein. The Cro protein consists of two subunits, each containing three alpha helices and three beta strands. The third helix is called the DNA interacting one, where five residues (Gln, Ser, Ala, Asn and Lys) are known to bind to DNA. The second helix is stabilising the protein. (Figure from Mathews et al., Biochemistry.)

Many different types of proteins are included in the group of HTH proteins, e.g. both prokaryotic and eukaryotic transcriptional regulators. Outside the HTH region the structure of the various proteins that contain this motif can vary enormously. However, a feature that is common among the HTH proteins is that they bind to DNA sequences, which are composed of two very similar, symmetrically arranged "half-sites", as symmetric dimers. The tryptophan repressor, the lambda (λ) Cro protein, the λ repressor fragment and the CAP (Catabolite Activator Protein) fragment are all examples of this feature (Harrison al., 1990).

Although information from structural and biochemical studies have revealed a lot about the interactions between HTH proteins and DNA, the understanding of the formation of the sequence-specific and high-affinity protein-DNA complex is still incomplete and additional tools for studying this is of great importance.

Previously interactions between zinc finger containing proteins and DNA have been studied with *in vitro* evolution techniques. Zinc finger proteins, mutated in their DNA interacting amino acid residues, have been expressed by phage display and selected for with "new" target DNA sequences. The protein structures have been analysed and DNA binding specificities have been relieved.

Phage display has proven to be a powerful tool for selecting ligands of defined specificity from large libraries generated by mutagenesis. So, an expression system based on mutated variants of a HTH protein was created in order to study the interactions between HTH proteins and DNA.

Phage Display

Phages are viruses that infect prokaryotic cells. Filamentous phages do not kill their host, they are lysogenic phages, but they extend the generation time of the host with approximately 50% compared to an uninfected cell. M13 is a filamentous phage that infects *Escherichia coli* (*E. coli*). The genome of the M13 phage is small, only 6408 residues, and known to represent ten genes. Gene III and gene VIII code for proteins that are expressed on the surface of the phage, so-called coat proteins (Figure 2). The gene III protein is involved in the adsorption of the phage to the host by attaching to the pili of *E. coli*. By cloning a gene and fusing it to the gene coding for a coat protein (gIIIp or gVIIIp) of the phage, the gene product will be displayed on the surface of the phage. Replication of the genome of the phage can be divided into three stages, conversion of viral ssDNA to replicative, double-stranded form inside the host, multiplication of the replicative form, and synthesis of viral ssDNA from the replicative form (Kornberg, 1974).

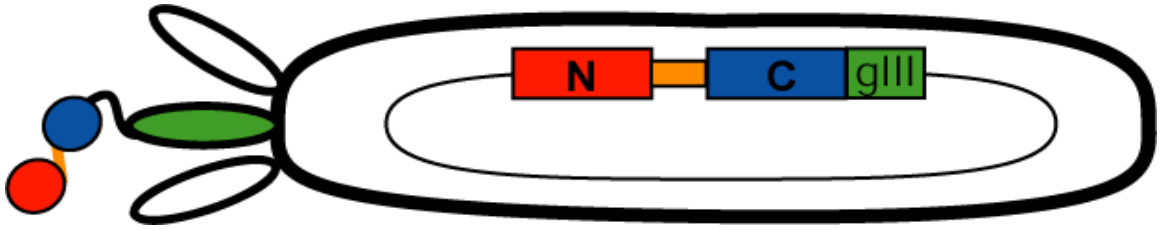

Figure 2. Schematic picture illustrating the phage. The coat proteins are expressed on the surface of the phage. In this figure the gene of interest (red, orange and blue) is fused to gene III of the phage (green) and therefore the protein of interest is expressed together with the protein coded by gene III. (Figure by Mikael Nilsson.)

Phagemids are a particular type of plasmids that contain two distinct origins of replication; colE1ori which promotes replication of dsDNA and M13ori which promotes the formation of ssDNA. The cloned gene of interest is inserted into the phagemid, fused to gene III. The genome of the phagemid can be expressed by the host but the phagemid is not able to make new transfectable phages since the genes coding for the capside proteins are missing. The phagemid needs another phage, a so-called helper-phage, to reproduce. When a cell containing a phagemid is infected by a helper phage, the M13ori on the phagemid promotes ssDNA replication including the gene of interest. The replication origin of the helper phage is defective resulting in that less ssDNA of the helper phage is synthesised, but proteins of the helper phage are expressed. The ssDNA of the phagemid is encapsulated into the capsid proteins of the helper phage and so new phages are produced. By using a phagemid monovalent (one copy) display is promoted. Statistically less than 10% of the phages will display a selectable protein, but less than 1% will display two copies of the fusion protein (Wells et al., 1992). The low amount of phages displaying a fusion protein is not a problem since the phages are in a much higher concentration (10^{11}-10^{12} particles per ml) than what a typical library has the size of (10^{7}-10^{8}).

Small variations of the gene of interest can easily be made with mutagenic PCR, where certain positions (certain bases) of the primers have been randomised. These randomly made variant genes can be inserted into the phagemid (the cloning vector) and used for transformation of the host bacteria. Later when helper phages are added to the bacteria, phages expressing a huge variation of the gene of interest are produced. These phages form a so-called phage library.

The Cro protein

The Cro protein is a small DNA-binding HTH protein that is expressed by the *E. coli* bacteriophage λ. It is a regulatory protein that is involved in the life cycle of the host. After the bacteriophage has infected *E. coli*, the Cro protein and the λ repressor protein determine the life cycle that the host will experience, prophage or lytic growth. In the prophage state, the repressor occupies the operator and thereby blocks the synthesis of several other bacteriophage proteins including the Cro protein whereas it activates its own synthesis. In the lytic state, the Cro protein occupies a different site in the operator, blocking the synthesis of the repressor but allowing its own synthesis. In the lytic state the DNA of the phage is transcribed, replicated, packed into new infecting particles, which are released by lysis of the host (Alberts et al., 1994).

Cro binds to DNA as a homodimer (Jana et al., 1997). Each subunit is 66 residues long and contains three alpha-helices and three antiparallell beta-strands. The third alpha-helix of each subunit of Cro fits the major groove of the DNA and recognises the bases in the sequence (Figure 3). Five residues of each subunit of Cro are known to interact with the target DNA, residues Q27, S28, A29, N31 and K32 of the third helix.

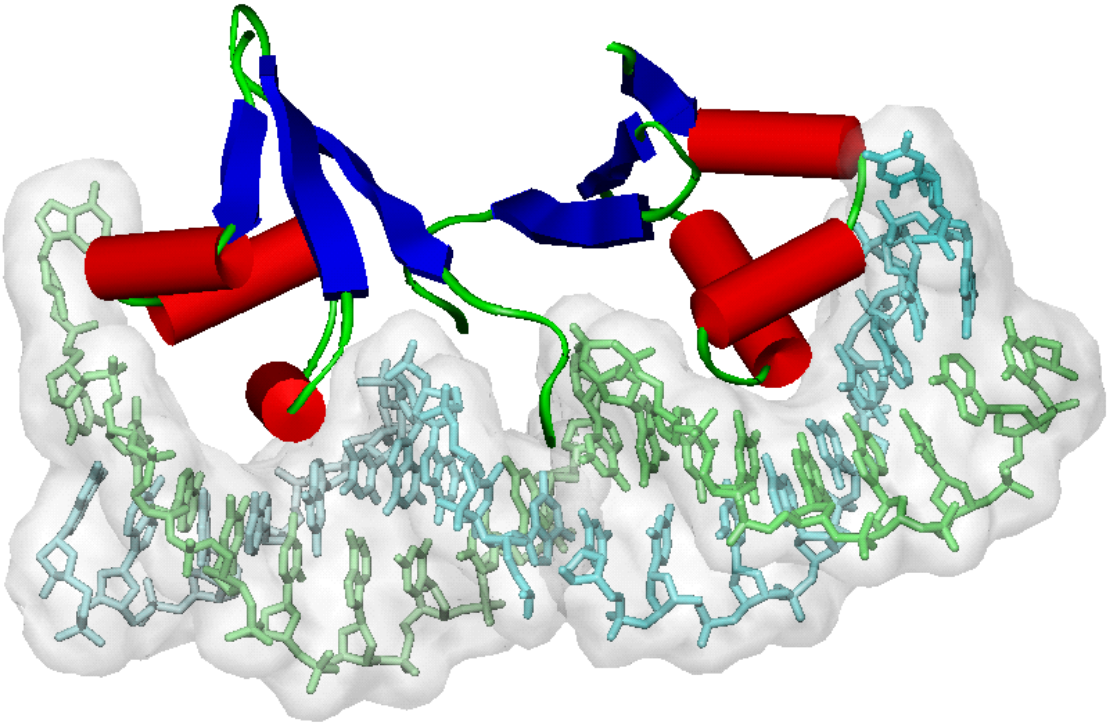

Figure 3. The Cro protein binds to DNA as a homodimer. Each subunit consists of three alpha helices and three beta strands. The third helix in each subunit binds to the major groove of the DNA. Five residues of the third helix (Q27, S28, A29, N31 and K32) are known to interact with DNA. (Figure by Mikael Nilsson.)

The selection system

The selection system was based on the Cro protein mainly because of its small size. Because Cro binds to DNA as a dimer, a derivative of the λ Cro protein was made to facilitate the dimerisation of the Cro protein. The derivative was made of two subunits of Cro covalently connected by an 8 residues long linker. It was called single-chain Cro8 (scCro8) (Jana et al., 1998) (Figure 4).

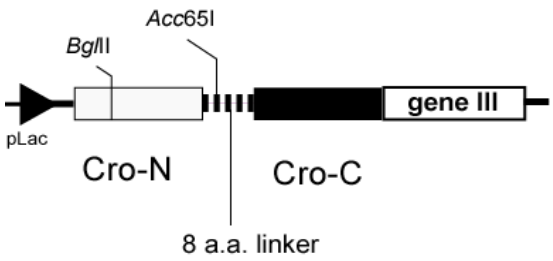

Figure 4. The scCro8 gene construct, where two subunits of Cro were connected by an eight residues long linker. Random mutations were inserted into N-terminal subunit of the Cro gene, downstream of the BglII site. The C-terminal subunit was fused to gene III in order to facilitate phage display (Nilsson et al., 2000).

The wild-type (wt) scCro binds specifically and with high affinity to ORC, an almost perfect palindromic DNA fragment resembling the natural DNA target of Cro. Mutations were done in the first subunit of the scCro protein by using NNS-strategy and site-directed mutagenesis. NNS-strategy means that at the first two positions of the codon all four bases, a, g, t or c, were randomly inserted whereas at the third position only g or c were inserted (Table 1). The total number of different codons is 64, but by using the NNS-strategy, the number of possible codons is reduced to 32. The mutations led to heterodimeric forms of the protein and ORas11 and ORas12 were used to select for heterodimeric mutants (Figure 5). During one selection round the phage library was incubated with biotinylated DNA (ORas11 respectively ORas12) and later with streptavidin-coated magnetic beads (Nilsson et al., 2000). The phages displaying the DNA binding mutants will be trapped on the magnetic beads due to biotin-streptavidin interactions. Not all the proteins displayed will interact with the DNA sequence and therefore the amount of selected phages will be reduced by each selection round. The number of selection rounds varies but when one or a few clones are left the selection is stopped (Figure 6).

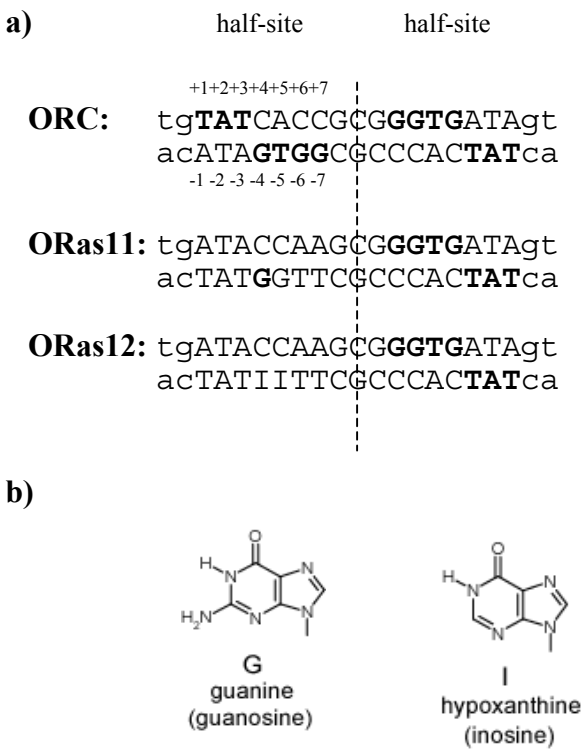

Figure 5. DNA ligands and base structures. (**a**) DNA ligands used in selection of phage-displayed scCro variants. The bold letters indicate wt scCro binding. The wt scCro binds tightly and specifically to ORC, ORas12 was used to select for heterodimeric scCro mutants and ORas11 was used to verify the specificity for ORas12. The only difference between ORas11 and ORas12 is that in ORas12 the guanosines at position –4 and –5 are replaced by inosines. (**b**) The structures of guanine and hypoxanthine.

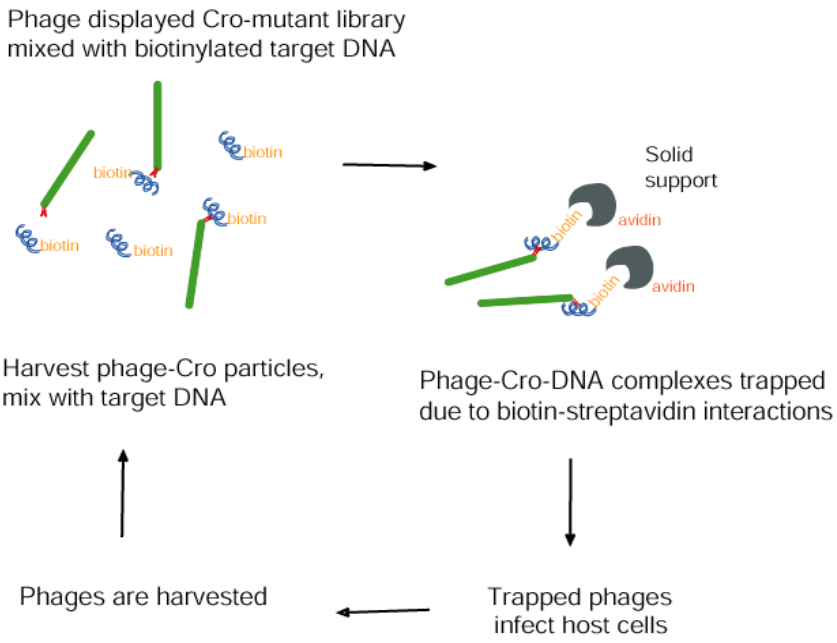

Figure 6. Schematic picture illustrating one selection round. The colours: green indicates phage, red indicates variant Cro protein displayed on the phage, blue indicates DNA and grey indicates streptavidin-coated magnetic beads. (Figure based on a picture from Mikael Widersten.)

Five residues (Q27, S28, A29, N31 and K32) in the third helix of Cro are known to interact with DNA. Residue 30 is an isoleucin, which is crucial for the stability of the protein. In the earlier made selection system the five DNA interacting residues were varied and mutants were selected for using the DNA sequences ORas11 respectively ORas12. In the selection with ORas12 it was evident that cystein at position 29 and tyrosine at position 32 were more frequently noticed than others (Nilsson et al., 2000).

AIM OF THE PROJECT

The aim of the project was to isolate and characterise novel mutant forms of the λ Cro protein and thereby learn more about the interactions between HTH proteins and DNA.

Here, the earlier selection with the ORas12 ligand was evolved. A smaller library was constructed, where three of the five DNA interacting residues (27, 28 and 31) were randomly mutated and the other two residues were changed by site-directed mutagenisis (A29C, K32Y). This gave a phage library of $3.3x10^3$ (32^3) individuals compared to the earlier library (Nilsson et al., 2000) with $3.4x10^7$ (32^5) individuals and the chances of finding a better mutant in this smaller library should have increased. Mutants were selected for by *in vitro* evolution where ORas12 was used as DNA ligand.

STRATEGY

Construction of library of Cro genes randomly mutated in codons for DNA interacting residues
The construction of a three amino acid varied mutant library was started by doing mutagenic PCR where parts of the primers were randomly varied and thereby causing different sequences to be synthesised. The PCR product was the origin of the library to be. A vector (phagemid) and the amplified fragments were cleaved and ligated. Transformation of *E. coli* XL1-Blue by electroporation followed. When the mutation library was constructed the size and quality of it was checked by titration and DNA sequencing.

In vitro selection of DNA binding Cro variants
Initially an antibiotic selection was done in order to get bacteria with wanted properties then helper phages were added to get the phages to reproduce. After the antibiotic selection the cultures were allowed to grow and the produced phages were harvested. Now *in vitro* selection of the phages expressing the different Cro variants followed, using ORas12 as target DNA. Four rounds of selection were done before a chosen number of randomly picked clones were analysed.

Functional characterisation of isolated variant Cro proteins
The specificity of the selected mutants was verified by doing another selection round but with three different target DNAs (ORas12, ORas11 and ORC). The mutant with the highest specificity for ORas12 was expressed, purified and characterised regarding DNA binding properties.

MATERIAL AND METHODS

Bacteria and phages

Two different strains of *E. coli*, XL1-Blue with a tetracycline resistance gene (Tet^R) and JM109(DE3), were used. The filamentous M13 phage was used as helper phage (M13K07, $1x10^{11}$ pfu/ml) and this strain has Kan^R.

Primers

Four different primers were used during the project; they are listed in Table 1.

Primer	Sequence (5´-3´)	T_m	Anneals to
Cro10mut	taa<u>agatct</u>cggcgtgtatNNSNNStgcatc NNStacgccatccatgccggcc	53°C	The pC3scCro8 vector Restriction site for BglII is underlined.
Cro8Lrev	gccgccagagccacc	67.3°C	The pC3scCro8 vector
pC3+1	caggtgaaactgctcgag	56°C	The pC3scCro8
pETfor	caattcccctctagaaataattttg	52°C	The pET vector

Table 1. Table showing the primers used, their sequences, melting temperatures and where they anneal. N=g, a, t or c and S=g or c.

Mutation PCR

The vector pC3scCro8 (1.4 ng) was used as template. The primers Cro10mut and Cro8Lrev were used at a final concentration of 1 µM. The other components were as in a standard PCR protocol; dNTP (0.2 mM), Mg^{2+} (1.5 mM, MBI Fermentas), 5 µl 10xPCR buffer (MBI Fermentas), Taq DNA polymerase (0.5U, MBI Fermentas) and water to a final volume of 50 µl per reaction. PCR cycle: 94°C 5 min, 35x(94°C 30 s, (58-51 gradient)°C 30 s, 72°C 30 s), 72°C 7 min. The PCR product was analysed on an 1.5% agarose gel containing 0.5 µg/ml EtBr. In order to increase the amount of PCR fragments another mutation PCR was done. This time a nested PCR was done, using the PCR product from 56°C and 58°C as template, otherwise the procedure was as previously. PCR cycle: 94°C 5 min, 35x(94°C 30 s, 58°C 30 s, 72°C 30 s), 72°C 7 min.

Precipitation of PCR product

One tenth volume 3 M NaOAc pH 5.2, two volumes 99.5% EtOH and 1 µl (20 mg/ml) glycogen were added to 1 volume PCR product. The solution was incubated in -20°C for at least 20 min and then centrifuged at 16000 g for 30 min in 4°C. The pellet was washed twice with 500 µl ice-cold 70% EtOH and then dried and resolved in a suitable volume, usually 30 µl, H_2O.

Digestion of PCR product and vector

The vector, pC3CroA11:402 (0.5 µg), and the PCR fragment (0.5 µg) were digested with the restriction enzymes Acc65I and BglII. Protocol for digestion of the vector: 5 µl pC3CroA11:402 (120 ng/µl), Acc65I (5 u, MBI Fermentas), BglII (5 u, MBI Fermentas), 3 µl 10xD Buffer (Promega) and H_2O to a final volume of 30 µl. Protocol for digestion of the PCR fragment: 5 µl PCR fragment (147 ng/µl), 84 u Acc65I, 28 u BglII, 15 µl 10xD Buffer and H_2O to a final volume of 150 µl. The two mixtures were incubated at 37°C for 2 hours. The digested PCR fragment was precipitated as above and resolved in a smaller volume.

Gel purification

All of the digested solutions were loaded on an 1.5% agarose gel containing 0.5 µg/ml EtBr and bands of the expected mobility on the gel were excised and purified according to QIAquick Gel Purification Protocol (QIAgen).

Test ligation and transformation (making a test library)

Initially a test ligation was done, using 50 ng vector, together with approximately three times the amount insert, 1 µl 10x Ligation Buffer, Ligase (1 u, MBI Fermentas) and H_2O to a final volume of 10 µl. The sample was incubated at 37°C for an hour and then inhibited at 75°C for 10 min. One microliter of the ligation mix was then used for transformation of 45 µl electrocompetent *E. coli* XL1-Blue. The transformation was done by electroporation. 1.5 ml 2TY [1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl] was added to the electroporated sample and the solution was incubated at 37°C, shaking, for an hour. A titration was performed on ampicillin plates [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 1.5% (w/v) bacto agar and 100 µg/ml ampicillin] and after over-night incubation at 37°C the size of the library was estimated by counting of ampicillin resistant transformants.

Checking quality of library

The quality of the library was checked by first preparing plasmids from 30 of the colonies according to Wizard Plus Miniprep DNA Purification System (SDS) followed by DNA sequencing of the variant Cro genes.

Sequencing

For sequencing the radioactive phosphorous isotope, [^{33}P]-ddNTP (Amersham Pharmacia Biotech), and the pC3+1 primer were used. The procedure was according to Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech). The PCR cycle: 95°C 1 min, 50°C 30 s, 72°C 1 min, 40x(95°C 30 s, 50°C 30 s, 72°C 1 min), 72°C 2 min. The samples were loaded on an acrylamide gel [50 g urea, 15 g 40% acrylamide, 20 g 5xTBE, 27.5 g H_2O, and 800 µl 10% APS and 80 µl TEMED for polymerisation] and run at 65 W for about 2 hours. The gel was fixed with 10% MeOH and 10% HaOAc for 15 min and then dried in a vacuum drier. A BioMax film was exposed to the gel over-night.

Large-scale library ligation and transformation

When the number of individuals in the library exceeded the number of different expected clones ($3.3x10^3$) and the DNA sequences of them were randomly distributed, a large-scale library was done in the same way as previously described. An eight times larger amount of the insert and the vector was used and time of incubation was 13 hours at room temperature. Inhibition of the ligation was done as previously mentioned. The ligation mix was precipitated and dissolved in 10 µl H_2O. Four transformations were done, each with 2.5 µl ligation mix and 100 µl *E. coli* XL1-Blue. The procedure was otherwise as previously written. After incubation the four samples were pooled and 2TY was added to a final volume of 15 ml. A titration was performed on ampicillin plates. The rest of the pool was subjected to selection with antibiotics and later helper phages were added as described below.

Antibiotic selection and addition of helper phages

In order to achieve bacteria with wanted properties the culture was subjected to antibiotic selection. (Selection with ampicillin selects for bacteria that have got the phagemid and tetracycline selects for bacteria that express pili. The final step of the selection is addition of helper phages and later kanamycin. The addition of helper phages is done to produce phages containing the different phagemids and selection with kanamycine selects for bacteria that have been infected by the helper phages.) Ampicillin and tetracycline were added to final concentrations of 20 µg/ml and 10 µg/ml, respectively, and the culture was incubated at 37°C, shaking. After an hour the ampicillin concentration was increased to 50 µg/ml and the incubation was continued for an additional hour. The culture was then transferred to a 500 ml bottle containing 100 ml 2TY with 50 µg/ml ampicillin and 10 µg/ml tetracycline and ~$1x10^{11}$ plaque forming units helper phages were added. The incubation continued for two more hours and then kanamycine was added to a final concentration of 70 µg/ml. The incubation continued over-night.

Phage harvest

The cultures were centrifuged at 2500 g for 15 min to sediment the bacteria. Phage precipitation solution [15% (w/v) NaCl, 20% (w/v) PEG8000] was added and incubation on ice for 15 min followed. Another centrifugation at 9000 g for 30 min followed to precipitate and pellet the phages. The supernatant was removed and the pellet was resolved in 1400 ml Phage Resuspend Buffer [10 mM tris-HCl, pH 7.4, 100 mM KCl, 1% (w/v)

bovine serum albumin (BSA), 0.02% (w/v) NaN_3]. A third centrifugation was done at 16000 g for 5 min to remove insoluble debris.

Affinity selection

Two mixtures, both containing the harvested phages (approximately 10^{10} pfu), biotinylated DNA (1 µM), sonicated salmon sperm DNA (1 ng), 19 µl 5xbinding buffer [50 mM tris-HCl, pH 7.4, 0.5 M KCl] and H_2O to a final volume of 100 µl, were incubated at 4°C and room temperature (RT) for an hour. Thereafter, the mixtures were added to 0.1 mg pre-washed [3x500 µl 1xBinding buffer containing 0.5% BSA] streptavidin-coated magnetic beads (1 mg/ml, SDS) and incubated at the same temperature as previously for another 30 min. After this incubation, the supernatants were removed and the magnetic beads were resolved in 500 µl washing buffer [1xBinding buffer containing 0.3% BSA] and transferred to a tube preblocked with 500 µl Binding buffer with 0.5% BSA for 1 hour. Four more washes were performed (one in the first round) and the magnetic beads were then resolved in 50 ml 2TY. The amount of selected phages were estimated by infecting 2 ml *E. coli* XL1-Blue in log phase, with the selected phages for 15 min at RT. Fifty microliter were then plated on ampicillin plates (two variants, undiluted and 10 times diluted) and incubated at 37°C over-night. The plates were called "out"-plates. The amount of initially added phages were estimated by infecting 100 µl *E. coli* XL1-Blue in log phase, with 2 µl diluted harvested phages (10^{-6}, 10^{-7}, 10^{-8}) for 15 min at RT and all was then plated on ampicillin plates and incubated as previously. These plates were called "in"-plates. The ratio between out-plates and in-plates was used to estimate the efficiency of each round of selection.

Plasmid prep

To find the dominating clones after three and four selection rounds, plasmids from bacteria growing on out-plates after round three and four were prepared according to Wizard plus miniprep DNA purification system (SDS). The mutated parts of the scCro genes were sequenced (Sequencing) using pC3+1 as primer.

Single pass selection

To verify the specificity of the mutants, some of the dominating clones were chosen for further studies. Three types of DNA ligands were used; ORC, ORas11 and ORas12. Another round of selection was done with the dominating clones, now with just one type of clone as in-phages. *E. coli* XL1-Blue was transformed with one type of phagemid, then the antibiotic selection (Antibiotic selection and addition of helper phages) and one round of selection (see Affinity selection, but the in-phages were diluted 1:200 and the work was done only at 4°C) was performed. The in- and out-plates were done as previously (Affinity selection) and the ratio was used as an indication of binding specificity for the different DNAs.

Subcloning into the protein expression system

The phagemids carrying the variants of interest were digested with restriction enzymes Acc65I and BglII, according to protocol; 10 ng plasmid DNA, 10 u Acc65I, 10 u BglII, 3

µl O^{+}Buffer (MGI Ferments) and water to a final volume of 30 µl. The mixtures were incubated at 37°C for 2 hours and then purified on gel (Gel purification).

Ligation

The pETCroA11:402 vector, a previously modified and linear pET21(+) vector (Novagen), was ligated with the digested insert from the gel purification according to protocol; 2 µl pETCroA11:402 (0.6 ng/µl), 1.4 ng insert, 1 µl 10x ligase buffer, 1 u ligase and water to a final volume of 10 µl. Incubation for about an hour followed and the ligation was stopped by heating the samples at 70°C for 10 min.

Transformation

The transformation was performed as previously written, except the amounts that were different; 2 µl ligation mixture and 30 µl *E. coli* XL1-Blue. Plasmids from the colonies were prepared and sequenced to be sure the ligation and transformation were successful. A few changes were made to the sequence reaction mentioned above; the pETfor primer (2 µM) was used and the PCR cycle was different: 95°C 1 min, 48°C 30 s, 72°C 1 min, 40x(95°C 30s, 48°C 30 s, 72°C 1 min), 72°C 5 min.

New transformation (changing host)

The transformation was performed as previously, but with a new *E. coli* strain; JM109(DE3) and with different amounts; 1 µl plasmid DNA (pETCroC12:408 and pETCroC12:414) and 45 µl JM109(DE3).

Protein expression

A 20-ml-ONC (over-night culture) of scCroC12:408, containing 50 µg/ml ampicillin, was grown at 30°C. The next day the culture was diluted 1:100 into 2x750 ml. The growth continued until the OD_{600} was around 1.0 then expression was induced by adding 1 mM IPTG to the culture. After 3 hours the cells were harvested by centrifugation at 2500 g in 15 min. The pellet was resuspended in about 25 ml 50 mM potassium-phosphate, 50mM KCl, pH 7.0 and frozen in -80°C.

Lysis of cells

The cells were gently thawed and then lysed by sonication (3 x 20 s, power setting 7.5). The sample was centrifuged at 18000 g in 20 min and the supernatant was used for protein purification.

Protein purification

The yellowish supernatant was loaded on a Sephadex G-25 column, previously equilibrated with 50 mM potassium phosphate, 50 mM KCl, pH 7.0, in order to desalt the sample. The sample was run through the column and afterwards collected. To avoid clogging of the next columns, the collected volume was filtrated through four filters with

different pore sizes (5 µm, 2 µm, 0.45 µm and 0.2 µm). The filtrated sample was loaded on two 5 ml SP-HiTrap (Amersham Pharmacia Biotech) columns connected in series and connected to an ÄKTA*prime* (Amersham Pharmacia Biotech) system. The flow rate was 0.5 ml/min when the sample was loaded onto the columns and 1 ml/min when the columns were washed with 50 mM potassium-phosphate pH 7.0 and 5% 50 mM potassium-phosphate, 1 M KCl, pH 7.0 until the absorbance reached baseline. The protein was eluted with a salt gradient of KCl from 50 mM to 1 M KCl. Fractions from the different elution peaks were analysed on an SDS-PAGE gel. Two microliter low molecular weight marker and 3 µg previously purified scCro variant were loaded on the gel as guidiance. The gel was run according to manual, stained with Commassie Blue staining and destained with 10% HOAc and 40% MetOH. Fractions of interest were pooled and concentrated using an YM3 filter (Diaflo® Ultrafilters). The concentrated sample was loaded on an Sephacryl S-200 (16/60) gel filtration column equilibrated with 20 mM Tris-HCl, 100mM KCl, 1 mM DL-Dithiothreitol (DTT), pH 7.0, at a flow rate of 0.4 ml/min. Another SDS-PAGE gel was run on fractions of interest according to previously described manners.

RESULTS

Quality check of ligation and transformation

The library of phages expressing scCro variant proteins was estimated to be 72000±8000 individuals, which exceeded the maximum number of variants (33000). Thirty colonies were picked after transformation. The plasmids were prepared and sequenced to check the quality. Twentythree of the thirty colonies picked were sequenced good enough to read. Some of the colonies were religated vectors and in some colonies deletions had occurred so in only eighteen colonies variant scCro mutants were found. The quality seemed good, even though it cannot be statistically verified due to the number of colonies. A slight excess of thymidine was suspected at the first codon at position 27, but it is hard to draw any conclusions from such a small number of colonies.

In vitro selection

The efficiency of the selection round can be estimated by looking at the ratio between the total out-phages and the total in-phages. An increasing ratio indicates that the amount of selected out-phage has increased.

$$\text{RATIO} = \text{number of out-phages/ number of in-phages}$$

The ratios were quite high in the first selection round, this was due to the mild wash (two times). After the second round the ratios had much decreased a lot, an indication of a large reduction of the number of out-phages. This was because of the rougher wash (five times). After the third selection round the ratios were increased, indicating an effective selection procedure. The selection stops when the ratios reach a plateau value. Before the fourth selection round was performed, the in-phages were amplified before use because of some days storage. This might explain the increasing ratios; the numbers of in-phages are increased so the ratios will increase (Table 2).

Selection round	**Ratio (4°C)**	**Ratio (RT)**	**Number of washes**
1	$1.0\text{x}10^{-4}$	$2.2\text{x}10^{-4}$	2
2	$1.5\text{x}10^{-7}$	$3.5\text{x}10^{-7}$	5
3	$1.8\text{x}10^{-6}$	$6.5\text{x}10^{-6}$	5
4	$7.2\text{x}10^{-7}$	$8.1\text{x}10^{-7}$	5

Table 2. The table shows the ratio between out-phages and in-phages at different rounds of selection at different temperatures, 4°C and RT and the number of washes per round.

Plasmids from the third and fourth selection round were prepared and the part of interest sequenced. The sequences showed that after four rounds of selection some clones were more frequent than others, they were dominating. Seven of the dominating clones, scCroC12:401(4°C), scCroC12:316(4°C), scCroc12:405(4°C), scCroC12:414(4°C), scCroC12:408(4°C), scCroC12:404(RT) , and scCroC12:403(4°C), were used for further studies; single pass selection experiments (Table 3).

Clone	**Amino acids** (at position 27, 28, 29, 31, and 32)
scCroC12:401(4°C)	V A **C** R **Y**
scCroC12:316(4°C)	S A **C** R **Y**
scCroC12:405(4°C)	P S **C** L **Y**
scCroC12:414(4°C)	W Q **C** S **Y**
scCroC12:408(4°C)	D R **C** L **Y**
scCroC12:404(RT)	Q R **C** L **Y**
scCroC12:403(4°C)	F C **C** K **Y**

Table 3. Table showing the seven dominating clones from the selection and their amino acid residues at position 27, 28, 29, 31, and 32. These clones were chosen for further studies. Bold letters indicate residues changed by site-directed mutagenisis.

Single pass selection

In the single pass selection experiments the ORas12 specificity of the mutant proteins was studied. Three different DNAs were used, ORas12, ORas11 and ORC otherwise the procedure was as in one selection round. The ratios between out- and in-phages per mutant clone and per DNA showed that two of the seven mutants (WQCSY and DRCLY) were specific for ORas12 (Table 4).

	VACRY	SACRY	PSCLY	WQCSY	DRCLY	QRCLY	FCCKY
ORas12/ORC	0.5	0.3	1.0	2.7	4.7	1.3	0.5
ORas12/ORas11	2.0	1.3	4.4	5.1	3.5	0.8	5.3

Table 4. Table showing the ratios between ORas12 and the two other ligands, ORC and ORas11. The ratios indicate how specific the mutants are for the DNA they were selected for (ORas12).

Mutant protein purification

The DRCLY mutant (scCroC12:408(4°C)) was considered to be the most interesting mutant for protein purification. After expression and sonication of the cells, the lysate was loaded to a G-25 column. Fifty milliliter were collected and filtrated through four filters. Some of the G-25 eluate was lost during the filtration and therefore only 41 ml were loaded onto the two SP-HiTrap (Amersham Pharmacia Biotech) columns. The collected fractions were analysed on an SDS-PAGE gel and fractions 25-27 were found to contain the protein of interest. The fractions were pooled and concentrated from 10 ml to 2 ml by using an YM filter. The concentrated 2 ml were loaded onto the S-200 Sephacryl gel filtration column and the collected fractions were analysed as previously (Figure 7). The mutant was found in two of the peaks and therefore two different pools were made each containing three fractions, 21-23 and 26-28. To estimate the concentration of the protein in each pool a spectrum from 250 nm to 340 nm was made. The second pool, 26-28, was behaving a bit strangely when the spectrum was made. The reason for this is not clear, but the two pools were kept apart even though both contained the same protein. The concentrations were estimated to be 30 mg/l and 22 mg/l. The yield from a 1.5 l-culture was 0.36 mg respectively 0.26 mg.

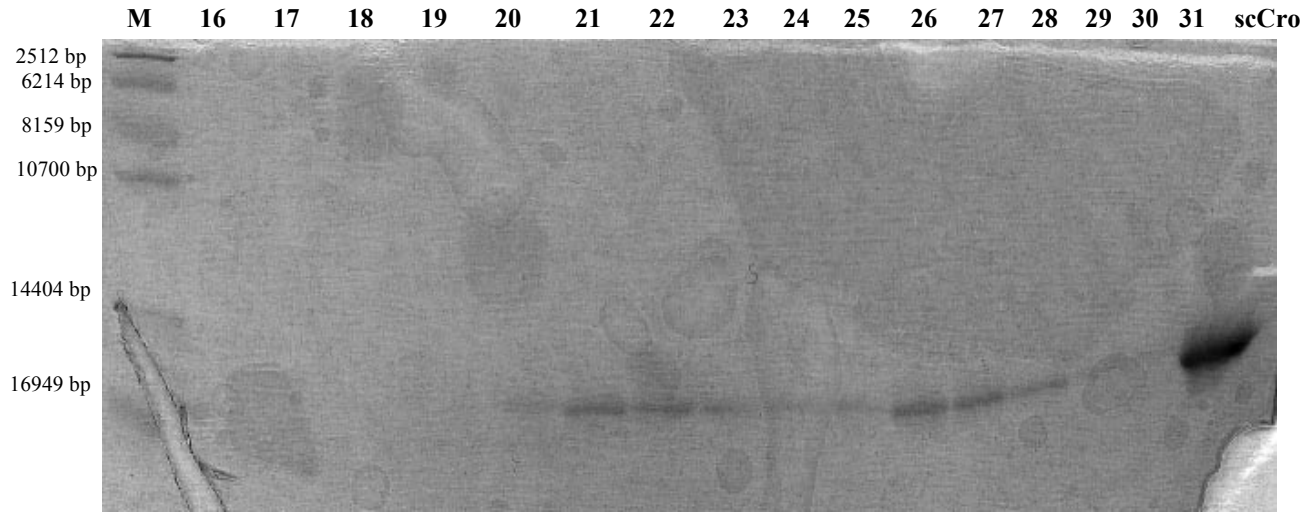

Figure 7. The protein content of the two peaks noticed in the chromatogram from the S-200 Sephacryl gel filtration column was analysed. M stands for low molecular weight marker and scCro for the previously purified scCro8 protein. The numbers to the left indicate the band size of the marker. The mutant protein (DRCLY) was found in some of the analysed samples, especially in samples 21-23 and samples 26-28 and two different pools were made of these.

DISCUSSION

The previous selection was performed with a library constructed of five randomly mutated residues at position 27-29 and 31-32 of the Cro protein. The result after four selection rounds showed a high frequency of 27X, 28X, 29C, 31X, 32Y mutants when ORas12 was used as ligand (Nilsson et al., 2000). Starting from these results a smaller library with three randomly mutated positions (27, 28, and 31) and two site-directed mutations (A29C and K32Y) was created. Selection with ORas12 was performed at two temperatures (4°C and RT) and after three and four selection rounds, respectively, sixteen randomly picked clones from each temperature were sequenced and analysed. It was found that some of the clones that seemed to dominate after three selection rounds decreased after the fourth round (PSCLY) while others seemed to arise or increase after the fourth round of selection (DRCLY, QRCLY) (Table 5). The reason for this is not sure. One reason could be that some clones, e.g. the PSCLY clone, were more efficiently expressed in the phage than others and therefore were dominating in the beginning of the selection and as the selection proceeded the actually better binders were taking over. Studies of the dissociation constants, K_D, of the different clones would clarify the case.

The three dominating clones from the earlier selection were DQCTY, TQCDY and RECKY (Nilsson et al., 2000). To be noticed is that after four selection rounds with the smaller library not one of the selected variants were the same as when the larger library was used. Hopefully indicating that better mutants were found by using the smaller evolved library.

One of the clones, DRCLY that arises after the fourth selection round, resembles the earlier DQCTY clone. Three of the five DNA-interacting residues in these two clones are the same (27, 29 and 32). At position 28, arginine (R) and glutamine (Q) are represented. These two amino acids are quite similar; they both have large side-groups, with a polar head at the same position. However, arginine is positively charged while glutamine is not. At position 31, threonine (T) and leucin (L) are represented. They are not very alike, leucin is hydrophobic whereas threonine carries a hydoxyl group and therefore is hydrophilic.

Another clone seen after selection is the WQCSY clone. This clone also resembles the earlier DQCTY clone and they have three of five residues the same (28, 29 and 32). At position 27, tryptophan (W) and aspartic acid (D) are represented. These two amino acids are very different from one another. Tryptophan is a large, bulky amino acid that contains an aromatic ring whereas aspartic acid is small and negatively charge. At position 31, serine (S) and threonine (T) are represented. These amino acids are very similar; they are both small and hydrophilic amino acid. The only difference is that threonine has an extra methyl group.

QRCLY is another clone that arises after the fourth selection round. This clone differs in only one amino acid from the DRCLY clone. At position 27 this clone has a glutamine (Q) whereas the other clone has an aspartic acid or aspartate (D). These two amino acids are very similar. Glutamine has a polar head group and is slightly larger than aspatric acid that is negatively charged.

Selection of scCroC12:301-316(4°C)

Clone	Position 27	28	31
12:306	P	S	L
12:310	P	S	L
12:314	P	S	L
12:315	P	S	L
12:301	V	A	R
12:311	V	A	R
12:312	V	A	R
12:316	S	A	R
12:303	K	C	R
12:304	G	I	F
12:308	G	I	F
12:309	G	P	Q
12:307	F	K	C
12:302	I	F	C
12:313	R	I	C
12:305	T	G	L

Selection of scCroC12:301-316(RT)

Clone	Position 27	28	31
12:307	P	S	L
12:311	P	S	L
12:314	P	S	L
12:304	G	I	F
12:312	G	I	F
12:315	G	I	F
12:302	W	L	F
12:309	W	Q	S
12:306	W	Q	S
12:303	F	C	L
12:313	F	C	L
12:310	K	W	L
12:305	S	A	R
12:301	R	I	C

Clones 12:308 and 12:316 are missing.

Selection of scCroC12:401-414(4°C)

Clone	Position 27	28	31
12:401	V	A	R
12:407	V	A	R
12:411	V	A	R
12:413	V	A	R
12:403	F	K	C
12:406	F	K	C
12:404	R	I	C
12:402	Q	L	A
12:412	L	L	A
12:408	D	R	L
12:405	P	S	L
12:414	W	Q	S
12:410	G	I	F

Clones 12:409, 12:415 and 12:416 are missing.

Selection of scCroC12:401-415(RT)

Clone	Position 27	28	31
12:402	G	I	F
12:403	G	I	F
12:408	G	I	F
12:411	G	I	F
12:401	V	A	R
12:405	V	A	R
12:407	V	A	R
12:404	Q	R	L
12:409	Q	R	L
12:406	D	R	L
12:412	P	S	L
12:415	I	F	C
12:413	R	I	C
12:414	W	Q	S

Clones 12:410 and 12:416 are missing.

Table 5. The four panels show the various amino acids at the randomly mutated positions (27, 28 and 31) after three respectively four selection rounds at two different temperatures (4°C and RT). The frequency of PSL clones is decreasing while the frequency of DRL and QRL are increasing.

Seven of the dominating clones were chosen for further studies, VACRY, SACRY, PSCLY, WQCSY, DRCLY, QRCLY and FKCCY. The specificity for the ORas12 ligand was verified by single pass selection studies (Figure 8 and Figure 9). The specificity for the inosine-containing ORas12 over the guanosine-containing ORas11 was well recognised in the earlier selection study, which means that the mutants were discriminating between the two ligands. In this single pass selection study two of the seven chosen clones, DRCLY and WQCSY, were found to be specific for ORas12, two clones, SACRY and VACRY, were more specific for ORC (like the wild-type scCro8) and one clone, PSCLY, was equally specific for both ORC and ORas12. The most striking result was that the QRCLY clone, which differs in only one amino acid (Q instead of D) from the DRCLY clone, was found to be more specific for ORas11 than for ORas12, which it was selected for. Hence, this mutant (QRCLY) showed no discrimination between the two ligands. The varied results found with the DRCLY and the QRCLY clones indicate that aspartic acid is important for sequence specificity. The results from this study can be compared to the results from the earlier study, which showed that only the DQCTY and TQCDY clones were specific for ORas12. The wild-type Cro (QSANK) binds specific to ORC but when comparing its binding properties for ORas12 and ORas11, the wt Cro is more specific for ORas12 than for ORas11.

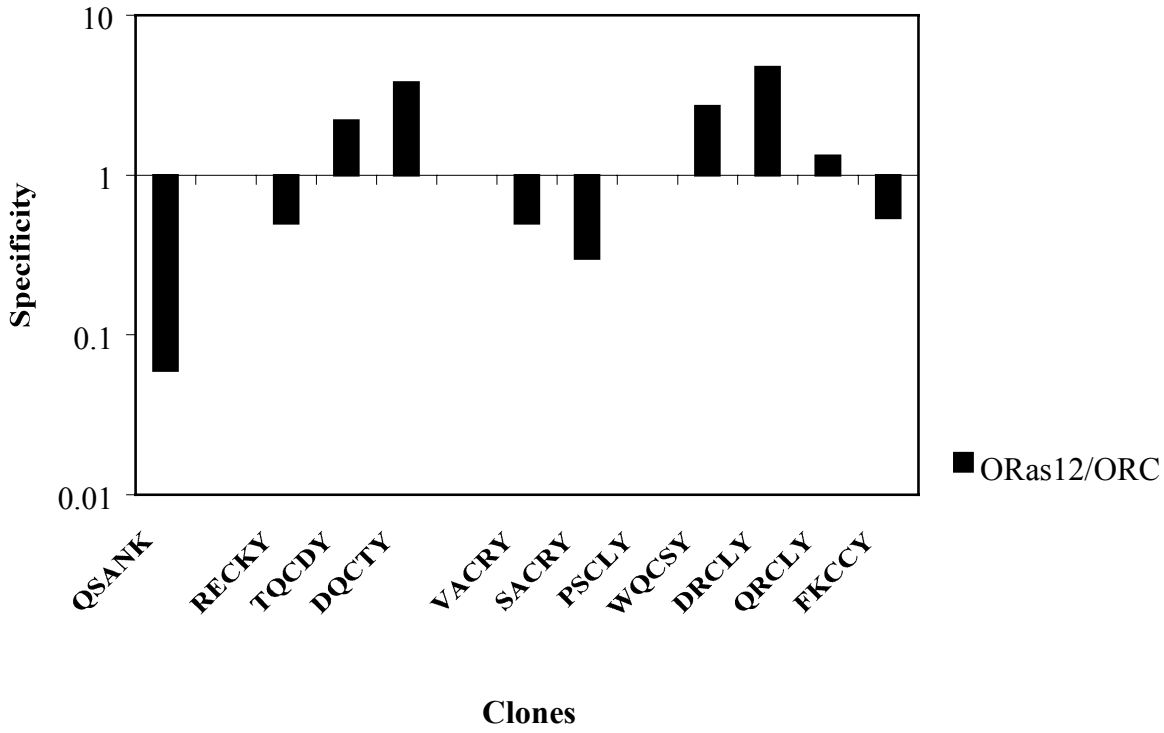

Figure 8. Results from the single pass selection study. The figure shows the specificity for the ORas12 ligand compared to the ORC ligand on a logarithmic scale. The studied clones were wild-type scCro8 (QSANK), three clones from the earlier study (RECKY, TQCDY and DQCTY) (Nilsson et al., 2000) and seven clones from this study (VACRY, SACRY, PSCLY, WQCSY, DRCLY, QRCLY and FKCCY). Five of these clones, TQCDY, DQCTY, WQCSY, DRCLY, and QRCLY, show specificity for ORas12 compared to ORC.

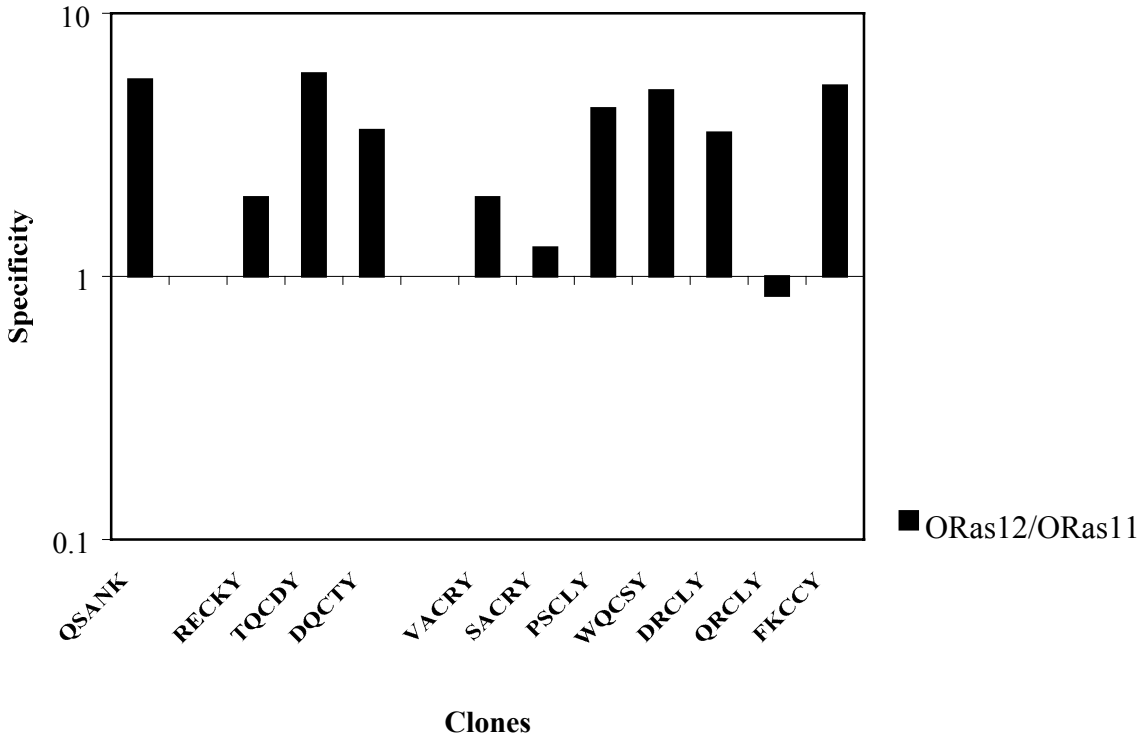

Figure 9. Results from the single pass selection study. The figure shows the specificity for the ORas12 ligand compared to the ORas11 ligand on a logarithmic scale. The studied clones were wild-type scCro8 (QSANK), three clones from the earlier study (RECKY, TQCDY and DQCTY) (Nilsson et al., 2000) and seven from this study (VACRY, SACRY, PSCLY, WQCSY, DRCLY, QRCLY and FKCCY). To be noticed is that one clone, QRCLY, binds more specifically to ORas11 than to ORas12, which it was selected for.

The DRCLY mutant was expressed and purified but due to time limitations of the project the characterisation of the mutant was not done.

Future studies would be to determine the dissociation constant, K_D of the DRCLY mutant, by electrophoretic mobility shifts (EMS) and also to express and purify the QRCLY mutant and compare the K_D of the both mutants. This would hopefully reveal the importance of aspartic acid for ORas12 specificity. It would also be interesting to determine the K_D of the other ORas12 specific mutant, WQCSY.

CONCLUSIONS

This project was part of a larger work where a goal is to construct homodimers with new, pre-defined, DNA binding properties. Today specific and high-affinity heterodimers are found by using the selection system. To study the protein/DNA interactions more precise, efforts to produce information on the 3-D structure of a previously purified mutant are made through X-ray crystallography. The aim for the future is to use mutant subunits and make homodimers and in that way resemble the native Cro protein even more. A thrilling idea is to get such knowledge of the protein/DNA interactions that prediction of which DNA sequence the protein binds could be made just by looking at the sequence of the protein.

ACKNOWLEDGEMENTS

First I would like to thank my supervisor Mikael Widersten for letting me do my master's degree project at the Dept. of Biochemistry and for guiding me through this project with such interest and enthusiasm. I would also like to thank Mikael Nilsson for introducing me to the exiting world of phages and for always having time and patient to show and explain things. I am grateful to Birgitta Tomkinson for being my examiner and for reading this report in such a short time. Last but certainly not least, I would like to thank all the people working at the Dept. of Biochemistry for always being so helpful and friendly and thereby making this time so enjoyable and fun.

REFERENCES

1. 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. *Prot. Engin.* **13:** 519-526.

2. Wells J.A. and Lowman H.B. (1992). Rapid evolution of peptide and protein binding properties *in vitro*. *Curr. Opin. Struc. Biol.* **2:** 597-604.

3. Harrison S.C. and Aggarwal A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* **59:** 933-69.

4. Jana R., Hazbun T.R., Mollah A.K.M.M. and Mossing M.C. (1997). A folded monomeric intermediate in the formation of lambda Cro dimer-DNA complexes *J. Mol. Biol.* **273:** 402-416.

5. Stryer L. *Biochemistry.* (1995) Fourth Edition. Freeman.

6. Alberts B., Bray D., Lewis J., Raff M., Roberts K. and Watson J.D. (1994) *Molecular Biology of the Cell.* Third Edition. Garland.

7. Kornberg A. (1974) *DNA Replication.* Freeman.

8. Mathews, C.K., van Holde K.E. and Ahern K.G. (1999*) Biochemistry.* Third edition. Addison Wesley Longman.

APPENDIX

The genetic code

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Symbols for Amino Acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid or aspartate
E	Glu	Glutamic acid or glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine