In Vitro Evolution of the Binding Specificity of Neocarzinostatin, an Enediyne-Binding Chromoprotein†

Bernadette Heyd,[§,#] Frédéric Pecorari,[§,#] Bruno Collinet,[§] Elisabeth Adjadj,[¶] Michel Desmadril,[§] and Philippe Minard*[,§]

Laboratoire de Modélisation et Ingénierie des Protéines, UMR 8619, CNRS - Université de Paris-Sud. 91405 ORSAY, France, and Laboratoire de biophysique moléculaire INSERM U350, Université de Paris-Sud 91405 ORSAY, France

Received December 16, 2002; Revised Manuscript Received March 14, 2003

ABSTRACT: Neocarzinostatin is the most studied member of the enediyne-chromoprotein family, and is clinically used as an antitumoral agent. Neocarzinostatin could be a promising drug delivery vehicle if new binding specificities could be conferred to its protein scaffold. We used in vitro evolution methods to demonstrate that this approach is feasible. We created large libraries containing between 1.7×10^8 and 1.4×10^9 independent clones, where up to 13 side chains pointing toward the binding crevice were randomly substituted. We then used phage display to select variants that bind to a model ligand (testosterone) which is unrelated to the natural ligand of neocarzinostatin. Several different binders were selected from each library. The corresponding proteins were expressed in *Escherichia coli* and their affinities and specificities were characterized in detail. K_D values of about 20 nM were obtained for streptavidin-bound testosterone. The K_D of selected proteins for free soluble testosterone are between 7 and 55 μM and therefore higher than the K_D for streptavidin-bound testosterone. The spacer and streptavidin used during selection contributed to the high affinity of the selected binders for the target. Binding studies of 15 different steroids related to testosterone allowed us to determine that C3, 4, 5, 6, and 7 on cycles A and B and the conjugated 3 oxo group of the steroid molecule were essential for molecular recognition. Other testosterone analogues substituted on C1, 2, 9, 11, 15, and 17 were not discriminated from testosterone. These results demonstrate that the binding specificity of this protein family can be extended to compounds that are completely unrelated to the natural enediyne chromophore family. This type of highly expressed, stable proteins with tailored binding properties have a wide potential range of applications.

As the number of known protein structures grows, it is becoming increasingly clear that many protein architectures are to some extent functionally versatile (*1*). Very similar protein structures are regularly discovered in proteins that are involved in the molecular recognition of unrelated molecular partners. Although evolutionary convergence may have occurred in some cases, most single folds with different binding capacities result from divergent evolutionary processes. Thus, natural evolution seems to have commonly made extensive use of each structural innovation, suggesting that diversification from a preexisting fold is often a more economical evolutionary solution than the emergence of a new fold.

The wide-spread nature of this process suggests that in vitro evolution experiments mimicking the natural process of divergent evolution could be used to extend the natural repertoire of molecular partners of a protein family. In vitro evolution methods associates in vitro-generated genetic diversity with selection procedures such as phage, ribosome, RNA, or cell display (see ref *2* and references therein for a recent review). These strategies have been applied with much success to the exceptionally versatile immunoglobulin fold (*3*−*5*). Immunoglobulin-based phage display and ribosome display libraries have been successfully used to select in vitro-evolved antibodies with new specificity or tailored properties.

The main objective of this work was to apply directed evolution methods to create a new specific binding site in neocarzinostatin. The enediyne chromoprotein neocarzinostatin (NCS) has elicited considerable interest from chemical, biological, and medical perspectives, and is currently the center of intense research activity (*6*−*10*). The neocarzinostatin chromoprotein was originally identified as a cytotoxic and antibiotic compound secreted by *Streptomyces* species (*11*). This secreted compound was subsequently shown to be a tight but noncovalent complex ($K_D \sim 10^{-10}$ M) between a protein (hereafter referred to as NCS) and an organic compound called the neocarzinostatin chromophore (*12*). The potent antitumoral and antibiotic activity of this compound are due to the DNA cleaving activity of the enediyne ring of the neocarzinostatin chromophore. Although it has been proposed that the protein part of the complex could be a histone specific protease, recent results from our laboratory (*13*) and others (*9*) do not support this view. The function of the protein is to stabilize the otherwise very

† This work was supported by Institut Curie (Programme Incitatif et Cooperatif).

* Corresponding author. Tel: 33 1 69 15 71 39. Fax: 33 1 69 85 37 15. E-mail: philippe.minard@mip.u-psud.fr.

# These authors contributed equally to this work and should be considered as co-first authors.

§ Laboratoire de Modélisation et Ingénierie des Protéines, UMR 8619, CNRS - Université de Paris-Sud.

¶ Laboratoire de biophysique moléculaire INSERM U350, Université de Paris-Sud.

10.1021/bi0273664 CCC: $25.00
Published on Web 04/24/2003

reactive chromophore and to release the chromophore to the target DNA. Although the chromophore without its carrier protein displays some activity in vitro, all clinical trials and current applications of neocarzinostatin rely on the complex neocarzinostatin/neocarzinostatin chromophore as an active principle. A form of this product substituted with Styrene-Maleic anhydride copolymer (SMA NCS) is currently being used in Japan for the treatment of hepatic tumors (*14*).

The ability of NCS to act as a transporter and its very low immunogenicity suggest that this protein could be a useful and general transporter if other binding specificity could be conferred to the protein scaffold. It was very recently demonstrated that wild-type NCS can accommodate substituted naphthoate compounds related to the neocarzinostatin chromophore within its binding cleft (*9*). However, it remains unclear whether engineering the binding pocket of this carrier protein could result in completely new binding specificity, for target compounds not related to the NCS chromophore. The creation of a new binding specificity for a drug-like compound to a non-IG protein scaffold is currently a difficult challenge that has very rarely been successful (*15*).

Structural comparisons of NCS with other related proteins suggest that this approach could be feasible. NCS is a member of the chromoprotein family. The four other known members of this family (actinoxanthin, macromomycin, kedarcidin, and C1027 protein) are close homologues to NCS and all tightly and specifically bind to their own specific chromophore. The structures of these proteins are almost identical (PDB codes: 1ACX, 2MCM, 1AKP, 1HZK, 1NOA) to that of NCS (*16*−*18*). These proteins form a beta sandwich, the topology of which is similar to the immunoglobulin fold (*17*). However, there is no significant homology between this prokaryotic chromoprotein family and the eukaryotic immunoglobulin fold family. The chromophore is bound in a deep crevice that is formed by an extension of two loops that fold up along one side of the barrel. The side chains protruding into this crevice are not conserved within the protein family, suggesting that they are not directly involved in fold stability. *Streptomyces* species seem to have taken advantage of several variants derived from this protein architecture to bind a range of different chromophores tightly and specifically. In each case, the side chains are arranged in a specific pattern in the crevice.

We used in vitro evolution methods to extend this putative evolutionary process and to create a new binding site for a small molecule unrelated to the neocarzinostatin chromophore in this protein scaffold. We first developed an integrated display and expression system taking advantage of the very favorable expression properties of NCS. We created large libraries that were used to select variants that bind a model "hapten": testosterone. The proteins produced by the selected clones were expressed and their specificity analyzed.

EXPERIMENTAL PROCEDURES

Construction of Libraries. The synthetic gene encoding NCS originated from the pNCS.sec (*13*) vector was amplified by PCR and subcloned into the NdeI and NheI sites of a new vector called pHDi.Ex. designed to allow phage display and direct high level expression of proteins variants previously selected by phage display (see results and Figure 1).

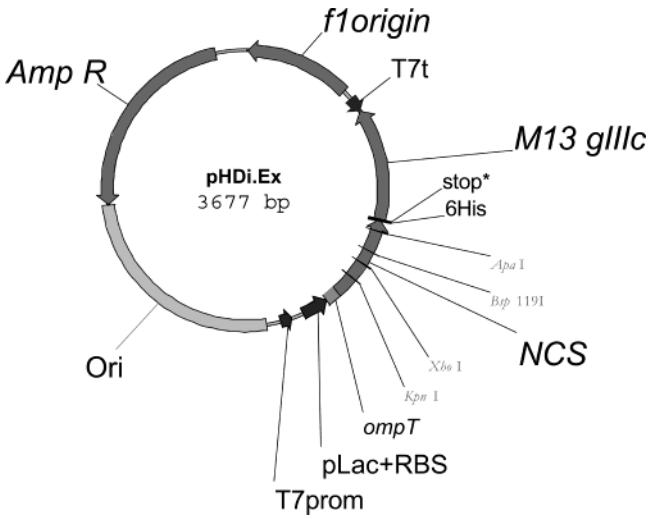

FIGURE 1: Schematic diagram of the phage display and expression vector. AmpR: gene encoding ampicillin resistance. Ori: origin of replication from ColE1. T7 prom: ϕ10 promoter specific to T7 RNA polymerase. pLac+RBS: lactose promoter and ribosome binding site. OmpT: sequence for periplasmic targeting. NCS: sequence encoding NCS fused to six histidines. Stop: suppressible TAG stop codon. M13gIIIc: sequence encoding the C terminal domain of gene III from M13 phage. T7t: T7 transcription terminator. f1 origin: origin of replication and packaging signal of f1 bacteriophage.

Libraries were prepared by replacing the regions of NCS to be randomized with synthetic cassettes encoding mutations at the defined positions. To minimize recircularization of the vector that could introduce a wild-type NCS bias, we devised a counter selection strategy. The regions to be randomized, comprised between KpnI and XhoI sites or between Bsp119I and ApaI sites, were independently replaced with a dummy cassette containing a unique SmaI site, generating vectors pHDi.ExΔKpnXho and pHDi.ExΔApaBsp. Two oligonucleotides pairs were used to produce these counter selection cassettes (ΔKpnXho:5′-CTCACCCGGGAATC-3′, ΔKpnXho.rev: 5′-TCGAGATTCCCGGGTGAGGTAC-3′, ΔApaBsp: 5′-CGAAGTACCCGGGAATGGGCC-3′ and ΔApaBsp.rev: 5′-CATTCCCGGGTACTT-3′).

Synthetic cassettes encoding random residues were prepared with HPSF grade oligonucleotides from MWG Biotech. For library 1, Lib1 (5′-GCAGGTACCGCCTATNNKGTTNNKCAGNNKGCCNNKGTCGACACCGGCGTA-3′) and Lib1.rev (5′-GGTGACGCTCGAGNNATCCGCGGGATTMNNCGCMNNTACGCCGGTGTCGAC-3′) oligonucleotides were annealed at 60 °C and filled-in at 72 °C with pfu DNA polymerase (Stratagene). The cassette for library 3 was produced in the same way using Lib3 (5′-CAGGCAGGTACCGCCTATNNSGTTRVYCAGNNSGCCNNSGTCGACACCGGCGTA-3′) and Lib3.rev (5′-GGTGACGCTCGASNNATCCGCGGGATTSNNCGCSNNTACGCCGGTGTCGAC-3′) oligonucleotides. The cassette for library 2 was produced by PCR using pHDi.ex as a template and Lib2 (5′-CGTAGCTTCGAAGGCTTCCTGNNSGACGGTACGCGTTGG-3′) and Lib2.rev (5′-ACCTTCGGGCCCGTTSNNSNNAGCATCSNNTAARBYCACSNNGCACGCGGCCGTGG-3′) as primers.

Cassettes fragments were digested with KpnI and XhoI, or Bsp119I and ApaI, and ligated into pHDi.exΔKpnXho, or pHDi.exΔApaBsp that had been cut with the same enzyme. Finally, T4 DNA ligase (New England Biolabs) was

heat inactivated, and the ligation products were digested with SmaI, and electroporated into the *Escherichia coli* XL1 Blue MRF′ (Stratagene) strain. After 10 electroporations of approximately 10 μg of phagemid vector for each library, libraries 1, 2, and 3 were obtained. These libraries contained 6.8×10^8, 1.7×10^8, and 2.2×10^8 independent clones, respectively. Library 4 was constructed via in vitro recombination of libraries 2 and 3. Plasmid DNA samples from libraries 2 and 3 were prepared and digested with EcoRI and XhoI. The resulting 742 bp fragment from library 2 was ligated into the large fragment (2935 bp) from library 3. The ligation product obtained was desalted, concentrated, and electroporated into the *E. coli* XL1 Blue MRF′ strain. Forty electroporations were performed with approximately 40 μg of DNA, yielding library 4 which contained 1.4×10^9 independent clones.

For all four libraries, electroporated cells were plated on 2YT plates containing 1% (w/v) glucose and 100 μg/mL ampicillin and incubated overnight at 30 °C. Plates were covered with 2YT medium containing 1% (w/v) glucose, 100 μg/mL ampicillin, and 20% glycerol, and the colonies were scraped off and stored in aliquots at −80 °C. At least 20 randomly clones were picked from each library and sequenced. No bias was detected in the distribution of amino acids at randomized position.

Phage Selection. To prepare phage stocks for selection, 2YT broth containing 1% (w/v) glucose and 100 μg/mL ampicillin was inoculated with the glycerol library stocks. The number of viable cells in the inoculum was approximately 10 times the estimated diversity for the corresponding library. Phage were rescued with VCSM13 helper phages (Stratagene) essentially as described (*19*) except that phages were allowed to replicate at 30 °C.

The phagemid libraries were panned using streptavidin magnetic beads (Roche diagnostics) and avidin agarose beads (Sigma) coated with biotinylated testosterone (Sigma). Beads were coated in TBS (20 mM Tris-HCl, 150 mM NaCl, pH 7.4 at 20 °C) containing 0.1% BSA for 30 min at RT The beads were then washed four times with TBS containing 0.5% Tween-20 and 0.1% BSA (TBST-B) before being used for the selection procedure. For the first round of selection, $10^{12}-10^{13}$ phages were incubated for 1 h under agitation at RT in TBST-B with 50 μL of the bead suspension in a total volume of 1 or 2 mL. Unbound phages were removed by washing 10 times with TBST-B. Bound phages were eluted by adding 200 μL of 0.1 M glycine pH 2.4 containing 1 mg/mL BSA for 10 min and then neutralized with 100 μL of 1 M Tris-HCl (pH 7.4). Eluted phages were used to infect exponentially growing XL1 Blue MRF′ cells. The infected cells that were spread onto 2YT plates containing 1% (w/v) glucose and 100 μg/mL ampicillin. An aliquot was used for the titration of eluted phages. Cells were coinfected with VCSM13 helper phages, and phages were allowed to replicate overnight at 30 °C. Culture supernantants were precipitated with PEG (20% PEG 6000, 2.5 M NaCl) and solubilized in TBS for use in the next round of selection. Three to five rounds of selection were carried out, using successively avidin neutravidin and streptavidin beads. The selection process was monitored by following the enrichment of eluted phages or by phage ELISA.

Phage ELISA. ELISA plates Maxisorp (Nunc) were coated for 1 h at RT with a solution of 5 μg/mL streptavidin (Sigma) containing 5 μM biotinylated testosterone in TBS. The plates were blocked for 1 h at RT with TBS containing 1% (w/v) BSA and 0.1% Tween-20 (TBST-B), and 10^8-10^{11} phages in TBST-B were added for 1 h. Bound phages were revealed with a horseradish peroxidase conjugated anti-M13 monoclonal (Amersham Biosciences), and detected at 390 nm using BM Blue as a substrate (Roche diagnostics). All steps were interspersed by three washes with TBS containing 0.1% Tween-20 (TBST). For the negative controls, ELISA wells were coated with streptavidin alone (i.e., without biotinylated testosterone) or with BSA.

Expression and Purification of Selected NCS Variants. E. coli BLR(DE3) pLysS strain (Novagen) was used as a host for the expression of NCS variants. Cells transformed with the corresponding phagemid were grown in 500 mL of 2YT medium in a 2 L flask for 70 h at 30 °C, 200 rpm, in the presence of 100 μg/mL ampicillin and 30 μg/mL chloramphenicol. Bacteria were pelleted by 15 min centrifugation at 9000*g* and the soluble proteins that had been directly secreted into the culture medium were precipitated with 650 g/L of ammonium sulfate. Proteins were pelleted by a 20 min centrifugation step at 17000*g*, solubilized in TBS, and then dialyzed overnight at 4 °C against 1 L of TBS to eliminate any remaining ammonium salts. Insoluble material was removed by centrifugation for 20 min at 17000*g*.

Proteins were purified as follows at RT. Twenty-five millimolar imidazole was added to the protein solution, which was applied to a 3 mL Ni-agarose (Ni-NTA agarose, Qiagen) column that had been equilibrated with TBS containing 25 mM imidazole at a flow rate of 1 mL/min. The column was washed with 30 mL of TBS containing 25 mM imidazole, and the protein was then eluted with TBS containing 250 mM imidazole. The eluted fraction was dialyzed two times overnight at 4 °C against 1 L of TBS. Samples of purified NCS variants were loaded onto SDS−PAGE gels (14% acrylamide-bisacrylamide) to check that they had been homogeneously purified. Protein concentrations were determined by spectrometry using extinction coefficients of 14180, 14180, 25560, 25560, 22430 M^{-1} cm^{-1} at 280 nm, for 1a.15, 1b.6, 3.24, 4.1, and 4.4, respectively (calculated from their sequences according to Gill and von Hippel[27]) Molecular masses of purified proteins were checked by Maldi-TOF mass spectrometry.

ELISA Analysis of Selected Clones and Competition ELISA Analysis. The functionality of the purified NCS variants was checked by ELISA in conditions similar to those used for phage ELISA, except that phages were replaced with 200 nM or less of NCS variants, and that detection was carried out with an anti-$(his)_6$ tag horseradish peroxidase conjugate (Roche diagnostics). Negative controls were done as for phage ELISA.

Competition ELISA was carried out by the same way with steroids, but NCS variants samples were incubated with 10 and 100 μM competitors for 1 h before being added to ELISA plates. The concentrations of the NCS variants used were 100, 100, 100, 50, and 2 nM for clones 1a.15, 1b.6, 3.24, 4.1 and 4.4, respectively. This gave approximately the same development times for chromogenic reaction with all NCS variants. The steroid competitors (Steraloids) used were [1] 5-androsten-3β-ol-17-one, [2] 1,3,5(10)-estratrien-3,17β-diol, [3] 4-androsten-3,17-dione, [4] 5α-androstan-17β-ol-3-one, [5] 4-androsten-3β,17β-diol, [6] 4,9(11)-androstadien-

17β-ol-3-one, [7] 4,6-androstadien-17β-ol-3-one, [8] 1,4-androstadien-17β-ol-3-one, [9] 4-androsten-7β,17β-diol-3-one, [10] 4-androsten-6α,17β-diol-3-one, [11] 4-androsten-11α,17β-diol-3-one, [12] 4-androsten-11β,17β-diol-3-one, [13] 4-androsten-15α,17β-diol, [14] 4-androsten-2α,17β-diol-3-one, [15] 4-androsten-15α,17β-diol-3-one.

Measurement of Dissociation Equilibrium Constants. The binding affinities (K_D values) of the NCS variants for testosterone and biotinylated testosterone were determined by quenching intrinsic NCS fluorescence. The excitation wavelength was set at 295 nm and the emission was measured at 342 nm on a Jasco FP777 spectrofluorimeter using a quartz cuvette equipped with a magnetic stirrer. Increasing concentrations of ligands were added to a 1 μM purified NCS variant solution in 20 mM Tris, pH 7.4 at 25 °C. After each addition of the ligand, the intensity of the fluorescence was measured after equilibrium had been reached. All measurements were corrected for the fluorescence of the ligand itself. K_D values were calculated by a nonlinear regression method using SigmaPlot software.

SPR Analysis. All experiments were carried out on a BIAcore 2000 instrument. The ligand was immobilized on a SA chip by injecting 25 μL of a 5 μM biotinylated testosterone solution in Tris buffer (Tris-HCl 20 mM, pH 7.4 at 25 °C) at a flow rate of 5 μL/min. A streptavidin surface without biotinylated testosterone was used as a control. The K_D values of NCS variants for testosterone were determined by injecting 250 μL of protein samples prepared in Tris buffer at concentrations ranging from 100 to 400 nM with a flow rate of 40 μL/min. To check there was no rebinding artifact, 40 μL of a 70 μM testosterone solution in Tris buffer was injected at a flow rate of 40 μL/min during the dissociation phase for each binding analysis. The kinetic data were evaluated with BIAevaluation 3.0.1 software.

The dissociation constant of 1a.15 for testosterone was determined by competition BIAcore under mass transport limitation as described previously (*20*, *21*) using a sensor chip SA saturated with biotinylated testosterone (~1600 resonance units). A streptavidin surface without biotinylated testosterone was used as a control. Each binding-regeneration cycle was performed with a constant flow rate of 25 μL/min of 20 mM Tris buffer pH 7.4 at 25 °C. Samples of 200 μL of 250 nM NCS variant in Tris buffer containing various amounts of testosterone, preincubated at least 1 h at 4 °C, were injected through the sample loop of the system. The surface was regenerated by injecting of 10 μL of 5 M guanidinium chloride in Tris buffer. Data were evaluated by use of the BIAevaluation and SigmaPlot softwares. The slopes of the association phase of linear sensorgram during injection were plotted against the corresponding total testosterone concentrations to calculate the dissociation constant.

RESULTS

Display and Expression System. The previously described *E. coli* expression system using a synthetic gene and a T7 derived promoter (*13*) allowed very efficient secretion of NCS directly in the culture medium. We designed a new phage display phagemid (Figure 1) called pHDi.Ex. This combined genetically stable monovalent display system and high level expression system made it possible to directly express selected clones from a strong promoter without the need to subclone the expressed sequences. In this construct, a synthetic gene coding for NCS was controlled by two overlapping promoters: a "weak" promoter (pLac) for the relatively low expression level required for phage display and a very strong promoter (T7) for the high level expression of the unfused protein. In both cases, the NCS coding sequence was fused to the ompT presequence which is required for exportation to the periplasm. The NCS coding sequence was fused to a hexa histidine tag and followed by a suppressible amber stop codon. This coding unit is followed by the C terminal region of the M13 Gene III (249−406) (*22*). A T7 transcription termination sequence was located on the 3′ side of the transcription unit. These elements were assembled on a phagemid backbone, taking care to minimize unnecessary sequences so as to maximize the electroporation efficiency for library construction. A control experiment showed that this construction allows very high levels (>100 mg/L) of the his-tagged wild-type NCS to be secreted as a folded protein by *E. coli* strains expressing T7 RNA pol (e.g., BL21). Furthermore, when transformed into a supE strain, NCS was expressed fused with the C domains of the gene III protein and could be efficiently assembled on phages when cotransfected in F′ strains with a helper phage. Titration experiments with anti-NCS polyclonal antibodies show that under these conditions, NCS was displayed on approximately 0.3% of the phage population. This level is similar to that obtained with single promoter phagemid constructs (*22*). The genetic stability of the display system is an important practical consideration when working with large libraries and was therefore checked in control experiments. These controls have shown that this construct is genetically stable, as no deletions or genetic drift toward rearranged constructs was observed during phage propagation, whereas fast genetic drift was observed when NCS libraries were displayed with other phage display constructs.

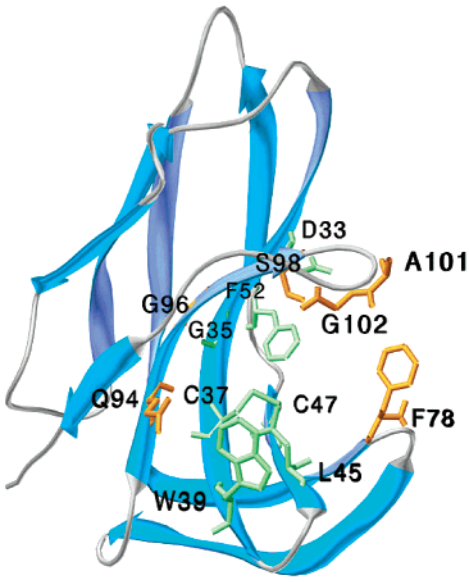

FIGURE 2: Structure of NCS (PDB code: 1NOA): The residues pointing toward the natural binding site of NCS that were randomized to construct libraries 1 and 3 are shown in green, and those randomized in library 2 are shown in orange.

Design and Construction of the Libraries. Like other chromophores from this chromo-protein family, the NCS chromophore is tightly bound in a deep crevice (Figure 2).

Table 1: Sizes of Libraries and Mutagenesis Schemes Used for Their Constructions[a]

library		randomized residues												
number	size[b]	D33	G35	C37	W39	L45	C47	F52	F78	Q94	G96	S98	A101	G102
1	6.8×10^8	NNK	NNK	NNK	NNK	NNK	NNK	NNK						
2	1.7×10^8								NNS	NNS	RVY	NNS	NNS	NNS
3	2.2×10^8	NNS	RVY	NNS	NNS	NNS	NNS	NNS						
4	1.4×10^9	NNS	RVY	NNS	NNS	NNS	NNS	NNS	NNS	NNS	RVY	NNS	NNS	NNS

[a] N is A or C or G or T; K is G or T; S is C or G; Y is C or T. [b] Expressed as colony forming units (cfu).

The structure of the protein suggested that the 13 side chains that point toward this crevice could be randomly substituted to create new concave surfaces and possible new binding sites. However, the size of a library required to exhaustively explore this sequence space (20^{13} different side chains combinations) is much higher than the size of practically accessible libraries. Thus, when designing the libraries, we chose a near exhaustive exploration of a reduced sequence space on one side, and a sparse sampling of a wider space on the other side. Recent results have shown that very sparse sampling of a potentially very large sequence space can be efficient (*23*). Conversely, analysis of any positive results obtained during of an exhaustive search of a reduced sequence space can reveal important features of the selected sequences, as more than one independent solution can occur in the same fitness well. We therefore used several strategies to construct libraries (Table 1). In library 1, only seven positions were simultaneously randomized. These seven positions were all located in the long twisted hairpin made by beta strands 3, 4, and 5. In the larger library 4, all 13 positions pointing toward the crevice were randomly substituted. Library 4 was built by combining two smaller independent sublibraries (libaries 2 and 3). In an attempt to limit the sequence complexity of this larger library, we considered that positions 35 and 96 correspond to glycine residues located at the bottom of the crevice. It seemed possible that a frequent occurrence of bulky side chains in these positions might interfere with the binding ability of the library. These positions were therefore only partially randomized in libraries 2, 3, and 4, so that these two positions only contained small side chains (i.e., G, A, D, N, S, T). In libraries 1, 2, and 4, the two cysteines corresponding to the 37−47 disulfide bond in wild-type protein were randomized, as previous results (not shown) suggested that this disulfide bond can be mutated without preventing the protein to fold. These libraries were constructed by ligating appropriate randomized cassettes in the expression/display construct. Large libraries (from 1.7×10^8 to 1.4×10^9 independent clones) were obtained. Randomly picked clones were sequenced and showed the expected variability, with no apparent bias in amino acid composition for the randomized residues.

Selection Procedure. Variants that were able to bind to testosterone were selected from the library. Testosterone was chosen as a target "hapten" for two reasons. First, many testosterone analogues useful to analyze the specificity of selected binders are available, and, second, the structures of several protein/testosterone complexes are known, which could prove useful for comparing naturally and artificially evolved modes of interaction.

Variants from libraries 1, 2, and 4 were selected and used to compare the efficiency and the variety of solutions found in these differently designed libraries. Furthermore, two identical but independent selections were successively conducted with library 1 to test the robustness of the selection procedure as stochastic dynamics may occur during the selections procedure (*24*).

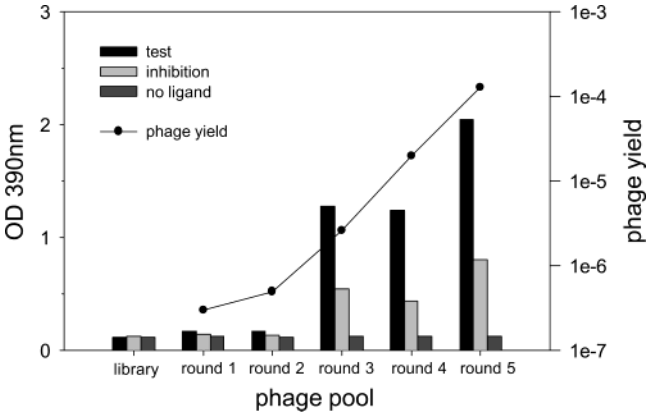

FIGURE 3: Monitoring of the selection process using library 1 (selection 1a). The specificity of the phages obtained from each round of selection was characterized by ELISA; phage pools were tested for their ability to fix to immobilized testosterone (test), to fix to immobilized testosterone in the presence of 50 μM free testosterone as a competitor (inhibition), and to fix in the absence of immobilized testosterone (no ligand). The phage yield is the ratio phage output/phage input at a given selection step as determined by phage titration.

A phage titration and a phage ELISA procedure were used to monitor the selection progress on phage pools at the end of each round of selection. Phage ELISA is based on the NCS-mediated binding of phages to testosterone coated ELISA plates. Bound phage is revealed by anti-phage immunoconjugates. An example of a selection procedure is shown in Figure 3. The output/input ratio exceed the background after the third round of selection. The selection of a specific phage population was confirmed by phage ELISA as a testosterone-dependent signal was clearly observed after the third round of panning. Analysis of individual randomly selected phage clones revealed that a significant proportion was able to bind testosterone. The results of the different selections and the sequences of the selected clones are presented in Table 2.

The four independently selected sequences from library 1 are clearly convergent solutions. Five of the seven randomized positions appear to be strictly conserved, and one position can be either a serine or a threonine. Only one of the selected positions showed variability in the selected clones. One of the clones was amplified twice and recovered from two completely independent selection process (1a.15 and 1b.13). Conversely, each selection process also produced a sequence that was only found in one of the two selections (1a.24 from selection 1a and 1b.6 from selection 1b). The

Table 2: Results for the Different Selections and Sequences of Selected Clones

				sequence[a]												
selection	library	% positive (at round)	selected clone	D33	G35	C37	W39	L45	C47	F52	F78	Q94	G96	S98	A101	G102
1a	1	28/38 (4)	1a.15	W	Y	R	A	H	S	L						
				TGG	TAT	AGG	GCG	CAT	TCT	CTC						
			1a.24	W	Y	R	R	H	S	L						
				TGG	TAT	CGG	CGT	CAT	TCG	CTC						
1b	1	2/24 (3)	1b.6	W	Y	R	S	H	T	L						
				TGG	TAT	CGT	AGT	CAT	ACG	CTC						
			1b.13	W	Y	R	A	H	S	L						
				TGG	TAT	AGG	GCG	CAT	TCT	CTC						
3	3	5/96 (3)	3.10	W	**A**	W	R	W	Y	H						
				TGG	GCC	TGG	CGC	TGG	TAC	CAC						
			3.24	W	**A**	W	R	W	Y	N						
				TGG	GCT	TGG	CGC	TGG	TAC	AAC						
			3.74	W	**T**	N	W	W	Y	N						
				TGG	ACT	AAC	TGG	TGG	TAC	AAC						
			3.92	W	**S**	N	W	W	Y	N						
				TGG	AGC	AAC	TGG	TGG	TAC	AAC						
4	4	32/48 (5)	4.1	W	**A**	S	W	Y	S	I	W	H	**T**	R	L	S
				TGG	GCC	AGC	TGG	TAC	TCC	ATC	TGG	CAC	ACC	AGG	TTG	TCC
			4.4	Y	**A**	W	S	H	Y	S	W	L	**A**	N	Y	A
				TAC	GCT	TGG	TCC	CAC	TAC	AGC	TGG	CTG	GCC	AAC	TAC	GCC

[a] Residues 35 and 96 were substituted in libraries 3 and 4 with RVY codon, so that only G, A, D, N, S, T were allowed for these positions (in bold).

most striking result was that all of the sequences independently selected from this library were virtually identical. Thus, the library clearly appears to contain enough diversity to sample this region of the sequence space with several independent but almost identical sequences.

Four independent selected sequences were selected from library 3. As before, the selection process was clearly convergent as these sequences can be clustered into two subfamilies: clones 3.10 and 3.24 on one side and clones 3.74 and 3.94 on the other, each with six common residues out of seven. These two subfamilies have three side chains in common, but the combination of the 37, 39, and 52 side chains differ. Interestingly, the randomized codon at position 35 in library 3 did not allow the presence of the tyrosine side chain that appeared to be conserved in selections 1a and 1b. This resulted in a new set of solutions from library 3 that were presumably out-competed by a different group of sequences in selections from library 1.

The two independent sequences selected from library 4 were not closely related to each other or to the sequences selected in library 3. The W pairs found in sequence 4.1 (33−45) might be reminiscent of the same pair found in clones selected from library 3. Library 4 sampled the sequence space much less exhaustively, and it is therefore not surprising that nonconvergent solutions emerged from this library.

Altogether, these results clearly show that a range of different proteins derived from NCS with new binding site for testosterone can be selected. The design of the library directly influenced the nature of the selected clones, but positive clones were identified whatever the assumptions or the constraints introduced in the library design.

Properties of the Selected Proteins. Five clones were selected for further studies, and the proteins produced from these clones were purified. The expression levels of the isolated proteins were 10, 5, 4, 16, and 8 mg of pure protein per liter of culture medium for variants 1a.15, 1b.6, 3.24, 4.1, and 4.4, respectively. These levels are clearly lower than for wild type NCS (>100 mg/L culture medium) but still compare favorably with most described alternative scaffolds. These purified proteins were stable as they were fully functional, without any sign of degradation or aggregation, after one year at 4 °C for concentrations over 100 μM.

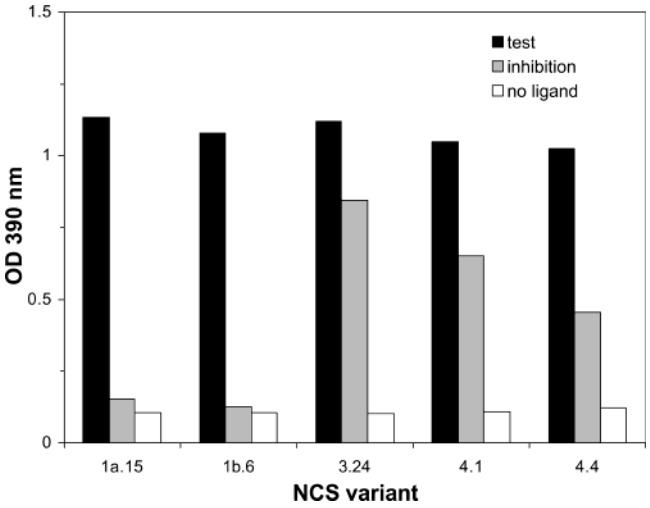

FIGURE 4: Binding specificity of five NCS variants. The purified proteins expressed by each clone was tested by ELISA for their ability to bind to immobilized testosterone (test), to bind to immobilized testosterone in the presence of 50 μM free testosterone as a competitor (inhibition), and to bind in the absence of immobilized testosterone (no ligand).

ELISA showed that the purified proteins were clearly able to bind testosterone specifically (Figure 4). No signal was observed in the absence of immobilized testosterone. Soluble testosterone competed with immobilized testosterone for binding to each of the purified proteins. The efficiency of soluble testosterone as a competitor was higher with some of the selected proteins than with others, suggesting that the mode of interaction differs between the different subgroups of selected proteins (see below for further discussion on this

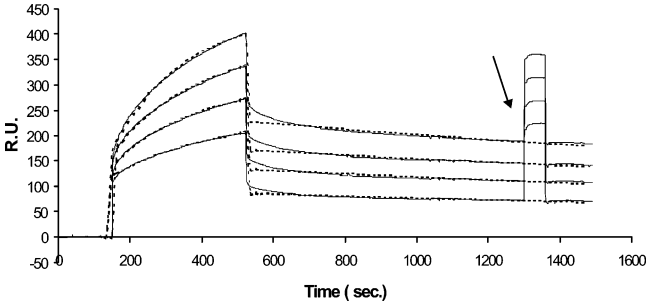

FIGURE 5: Determination of the affinity of variant 1a.15 for streptavidin bound testosterone by surface plasmon resonance. Purified protein 1a.15 was injected (100, 200, 300, and 400 nM) over a streptavidin chip surface coated with biotinylated testosterone. Dotted lines: optimally fitted theoritical curves. (parameters are $k_{on} = 1.18 \times 10^4$ M^{-1} s^{-1}, $k_{off} = 2.5 \times 10^{-4}$ s^{-1}, $K_D = 21$ nM). Injection of (70 μM) soluble testosterone (arrow) had no effect on the dissociation phases. The dissociation phase was therefore not affected by rebinding artifact.

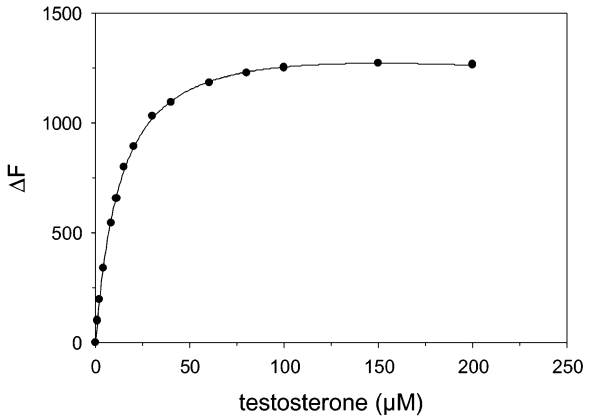

FIGURE 6: Determination of the affinity of variant 1a.15 for soluble testosterone by fluorescence titration. ΔF is the fluorescence intensity of protein 1a.15 in the absence of testosterone minus the fluorescence intensity after each addition of free testosterone. The fitted curve corresponds to a K_D of 14.8 $\pm$ 0.4 μM.

point). Dissociation constants were estimated using several different procedures. Biacore experiments were first conducted with the proteins from selection 1a. Biotinylated testosterone was immobilized on a streptavidin chip and the signal observed in the presence of different concentrations of circulating protein 1a.15 was measured (Figure 5). The observed kinetics for protein 1a.15 correspond to a K_D in the range of 20 nM for immobilized testosterone. Similar values were obtained with the four other clones (data not shown). Control experiments were conducted to detect potential artifacts, which are known to introduce bias in Biacore measurements. Similar K_D values were observed with two different flow rates and with two different ligands densities on the chip. This rules out a mass transport limitation effect. Competition with circulating testosterone slowed the apparent binding constant as expected but did not affect the dissociation phase (Figure 5). This demonstrates that the dissociated protein did not rebind, which is a common cause of underestimation of dissociation constant with this method (*20*). Finally, the purified protein was found to be strictly monomeric (data not shown) and the determined binding constants were therefore free of avidity effects.

Fluorescence titrations were also conducted with the purified proteins, as at least one tryptophan residue is present in the binding sites of the selected proteins (Figure 6). A clear fluorescence signal was observed in each case allowing an accurate determination of the binding constant for testosterone and for biotinylated testosterone. The binding constants determined (Table 3) with, respectively, testosterone and soluble biotinylated testosterone, show that in all the selected proteins, the spacer between testosterone and biotin contributed to the binding affinity. However, these K_D values were surprisingly 2 orders of magnitude higher than the binding constants determined from Biacore measurements for immobilized biotin testosterone.

To address this problem, competitive ELISA experiments were designed to evaluate the relative affinity of the purified proteins for testosterone, biotinylated testosterone, and the streptavidin−biotinylated testosterone complex (Figure 7A). The results clearly showed that the relative affinity for the complex was 2 orders of magnitude higher than that for soluble, free testosterone. Thus, the apparent discrepancy between the Biacore and fluorescence titration methods is

Table 3: Dissociation Constants of the Testosterone-Binding NCS Variants Determined by Fluorescence Titration

NCS variant	K_D (μM) for testosterone	K_D (μM) for biotinylated testosterone
1a.15	14.8 $\pm$ 0.4	6.5 $\pm$ 0.1
1b.6	6.9 $\pm$ 0.2	3.1 $\pm$ 0.1
3.24	13.0 $\pm$ 0.8	2.1 $\pm$ 0.1
4.1	55.0 $\pm$ 5.6	15.3 $\pm$ 0.5
4.4	18.1 $\pm$ 0.3	4.1 $\pm$ 0.1

due to contributions of streptavidin to the binding affinity. Controls showed that competitive ELISA can be used to determine the true binding constant, as in the conditions used the fraction of bound protein is a negligible fraction of the total protein concentration (*25*). Furthermore, Biacore experiments were conducted in competition mode to determine the binding constant for free soluble testosterone.[24] This clearly confirms that the K_D of protein 1.a15 for free testosterone is about 10 μM (Figure 7B) and therefore higher than the K_D for streptavidin-bound testosterone, which is close to 20 nM. The binding contribution of the steroid was dominant in all of the selected proteins, but the contribution of streptavidin and the spacer relative to that of the steroid seemed to be higher in clones selected from libraries 3 and 4.

Specificity of Steroid Binding. We used competitive ELISA to analyze the recognition specificity between the selected proteins and the steroid moiety. We compared the binding efficiency of each proteins with 15 different steroids related to testosterone (Figure 8). All of the proteins bound to the steroid with some level of specificity. Some of the steroid analogues that are very close to testosterone such as compound 1 (i.e., DHEA) were discriminated (Table 4). This clearly demonstrates that, although no specific pressure for molecular discrimination was introduced into the selection procedure, a fair degree of specificity was achieved. Proteins 1a.15 and 1b.6, with a relatively strong binding contribution from the steroid part, were more specific than others in their steroid binding spectrum. These studies also give an indication of which part of the steroid is discriminated (Table 4 and Figure 9). The conjugated carbonyl in C3 is clearly important, as are substitutions on C3, C4, C5, C6, and C7. However, these proteins do not discriminate testosterone from

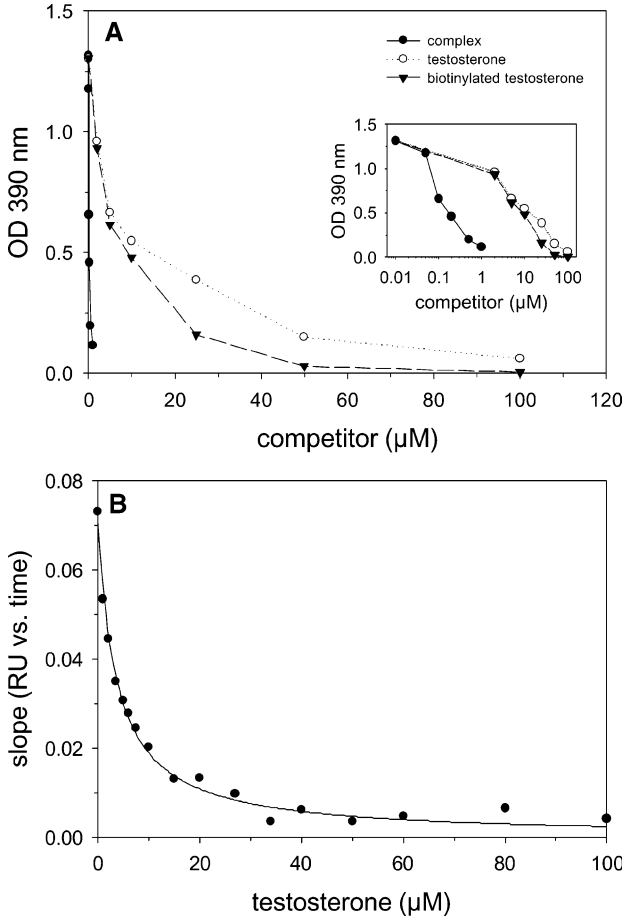

FIGURE 7: Competition experiments with protein 1a.15. (A) Competitive ELISA was used to study the relative effect of the biotinylated testosterone−streptavidin complex, free testosterone, and biotinylated testosterone on the binding of protein 1a.15 to immobilized testosterone. The inset shows the same data plotted on a logarithmic scale for competitor concentration. The K_D of protein 1a.15 for the biotinylated testosterone-streptavidin complex was around 100 nM and around 10 μM for free testosterone and biotinylated testosterone. (B) Determination of the affinity of protein 1a.15 for free testosterone by competitive BIAcore. The fitted curve corresponds to a K_D of 3.6 ± 0.4 μM.

analogues substituted in C2, C11, C15, and C17. Similarly, the 17-hydroxyl group of testosterone is not discriminated from the 17-oxo group of androstenedione (compound 3). This is probably related to the mode of selection, as the spacer in biotinylated testosterone is substituted on the C17-hydroxyl group. It seems likely that the selection procedure using C17-subtituted testosterone as a "hapten" induces a bias in the selected binders for an orientation of the steroid with the substituted D ring pointing toward the outside the binding crevice, and consequently C3, C4, C5, and C6 of cycles A and B are deeper within the binding site.

DISCUSSION

Our first aim was to create a new specific binding site for a small organic molecule using neocarzinostatin as a protein scaffold. Our results demonstrate that this protein and its variants can be displayed on phages and efficiently produced in *E. coli*. Large libraries were constructed and the side chains oriented toward the crevice were randomized. Three different libraries, differing by the number of mutated side chains and the nature of the residues allowed in some positions, were produced. Although the restraints applied when designing the libraries directly affected the nature of the selected clone families, all three libraries successfully produced proteins that bound the target molecule with a high affinity. Different families of binding solutions for the same target "hapten" were effectively found. These results clearly demonstrate, for the first time, that NCS is a protein scaffold that can be engineered and has a potential for molecular recognition that goes beyond the natural enediyne chromophore family. It should be emphasized that the creation of a protein with the ability to recognize a small molecule is a rather difficult problem, as the protein scaffold must provide the correct spatial positioning rather than simply fitting into a preorganized binding site.

Libraries with fewer randomly substituted side chains tended to produce convergent solutions, suggesting that libraries 1 and 3 were sufficiently diverse to sample the same region of the sequence space several times independently. Although the output of successive selections showed that not all selectable clones initially present in the library were effectively recovered (1a.24 and 1b.6 were only selected once), the same sequence family appeared to be reproducibly selected. Note that the sequence space of library 3 was included in that of library 1. However, type 3.10 or 3.74 sequences were not selected from library 1. It is currently unknown whether this is because the sequence family from library 1 has some selective advantage or whether this is due to partial sampling or undetected bias that differs between these two libraries. The sequence of the selected clones showed a high frequency of aromatic residues. Although this peculiarity has also been noted for phage display peptide libraries and antibody combining site, in our study this might be related to structural constraints due to the shape of testosterone.

For all of the selected clones, the steroid is the main but not the unique contributor to binding energy. Indeed, none of the selected proteins were able to bind to streptavidin without testosterone (Figure 4). However, the spacer and streptavidin contributed to the high affinity of the selected binders for the target used during selection. Streptavidin was used alternatively with neutravidin and avidin during the selection cycles, and therefore the structural elements of streptavidin contributing to the binding energy might be common to neutravidin and avidin. How streptavidin and the spacer contribute to the interactions with the selected proteins is not currently known and will only be revealed by a structural analysis of the ternary complex. The selection procedure could be adapted (e.g., competitive elution or a different mode of testosterone immobilization) to evolve these proteins further or to select different binder families from the initial libraries. The present report focuses on one type of target molecule. Work is currently in progress to address further issues such as the diversity, nature, and molecular properties of other targets that could be bound by these libraries

Although primarily designed as a first step toward the development of drug-targeting complexes, engineered neocarzinostatin with new binding specificities could also prove useful for other applications as an antibody-like recombinant molecule. Phage display technology provides a much faster and more convenient route of creating recombinant antibodies than hybridoma technology. However, the practical applica-

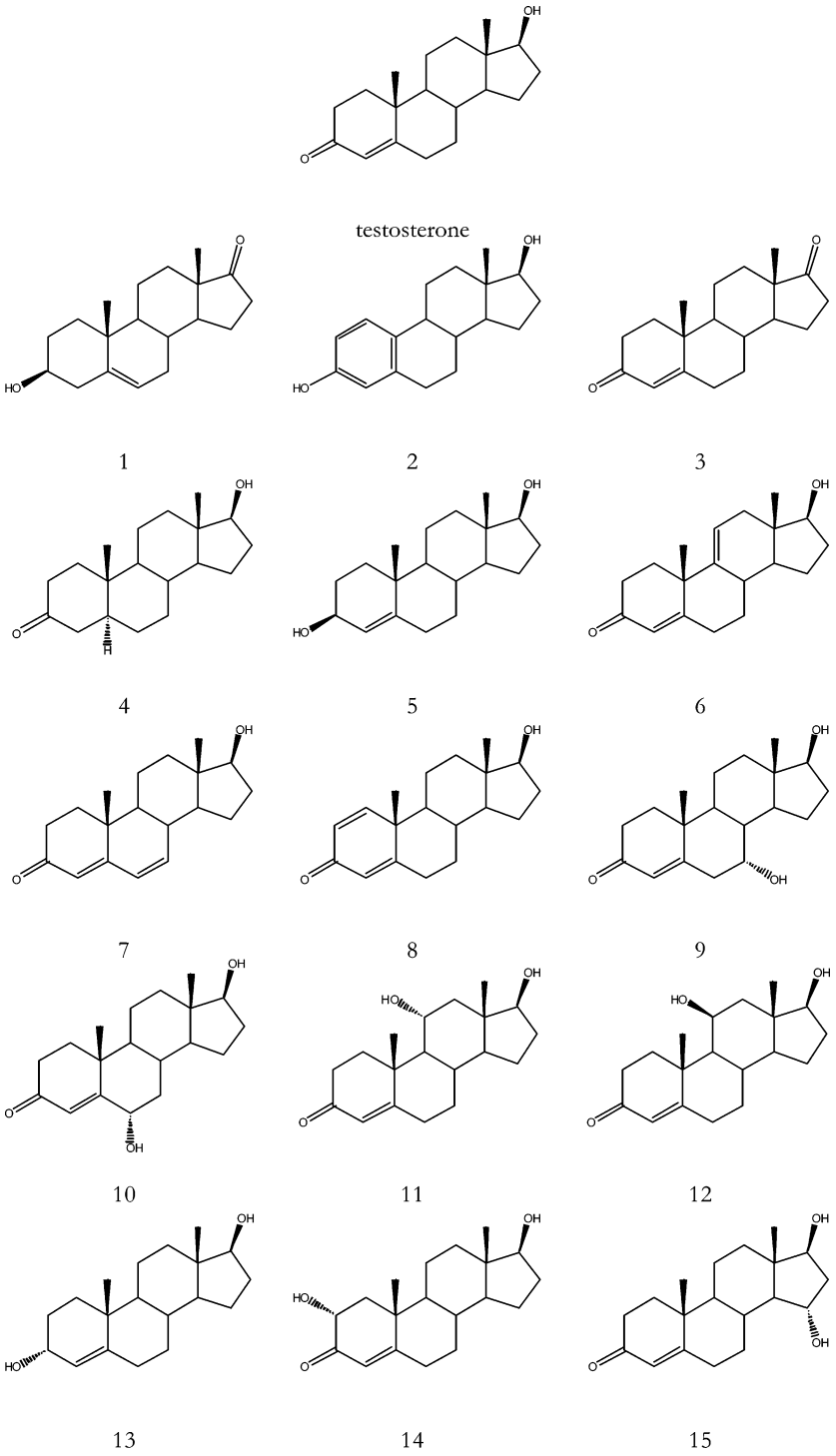

FIGURE 8: Chemical structures of testosterone and the other related compounds used as competitors in binding studies (see Experimental Procedures for correspondence between compound name and numbering).

tions of in vitro evolved recombinant antibodies as specific binding reagents are in many cases limited by their poor folding properties and low level of expression in bacterial expression systems. Several strategies have recently been explored to overcome this difficulty. For example, antibody frameworks with favorable expression properties have been selected (*26*) or evolved in vitro (*27*). These fragments can be used to generate libraries with improved expression properties. Another way of overcoming the unfavorable expression properties of antibodies is to develop stable protein scaffold with favorable expression but that are not related to antibodies (see for reviews refs *28*−*30*). Related studies have successfully selected "affibodies" from helicoidal protein libraries (*31*). A lipocalin scaffold protein with new specific binding sites for small molecule targets unrelated to the natural ligand (*15*) was recently created, as was a binding site that can modulate the properties of an active enzyme (*32*) making it possible to detect directly the interaction. This work demonstrates that NCS fulfills the criteria that are required for a good scaffold candidate. Its relatively small size (113 AA), high expression level, stability, and specific architecture are important and favorable features. It is noteworthy that the overall fold of NCS is similar to the immunoglobulin fold, but the region that was

Table 4: Steroid Binding Specificity of NCS Variants as Determined by Competition ELISA[a]

ligand	NCS variant				
	1a.15	1b.6	3.24	4.1	4.4
testosterone	±	±	±	±	+
biotinylated testosterone	±	+	++	++	++
1	−	−	±	±	−
2	−	−	−	±	−
3	±	+	+	+	+
4	−	±	−	+	±
5	−	−	−	±	−
6	±	±	−	+	+
7	±	±	±	+	±
8	±	±	−	+	+
9	-	-	−	±	±
10	±	-	±	+	+
11	±	±	−	+	±
12	±	+	−	+	±
13	±	−	−	±	−
14	±	±	±	+	+
15	±	±	−	+	±

[a] Competitive ELISA was performed with 1a.15 and 1b.6 binding proteins (10 μM) and 3.24, 4.1 and 4.4 (100 μM). Levels of inhibition are expressed relative to the signal observed in the absence of the competitor and are reported in the table using the following scale: (++) inhibition superior to 75%, (+) inhibition comprised between 50 and 75%, (±) inhibition comprised between 25 and 49%, and (−) inhibition comprised between 0 and 24%.

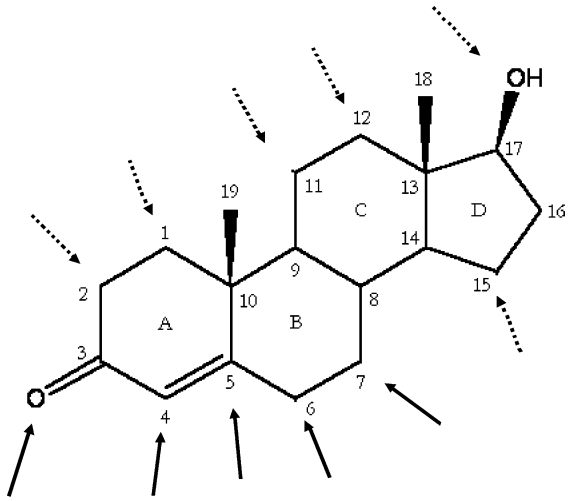

FIGURE 9: Binding specificity of selected binders. Summary of the results from competitive ELISA with proteins 1a.15 and 1b.6 (Table 4). The parts of the testosterone structure that are important for recognition by binding proteins are indicated with plain arrows, and those that are not important for recognition are indicated with dotted-line arrows.

randomly substituted in this study is the natural chromophore-binding site and is not equivalent to the CDR region of immunoglobulins. However, the similarity between the immunoglobulin fold and the NCS fold suggests that the loops equivalent to CDR in NCS can be randomized (either in combination with or independently from the crevice). This could provide a second region of this scaffold for the generation of new binding specificities. This second region appears to be well adapted to the recognition of macromolecular surfaces. This could be particularly relevant for the development of drug targeting as the NCS/chromophore complex is a powerful antitumoral agent that is currently used in human therapeutics. The current limits to NCS therapy are related to the lack of tissue specificity and the toxicity of the NCS/chromophore complex. It should be possible to select individual clones with tumor-associated antigen specificities that could help chromophore targeting on specific tissue from NCS libraries randomized in CDR-like regions.

ACKNOWLEDGMENT

We thank Chistophe Marchand for mass spectrometry experiments and Joel Mispelter for helpful discussions.

REFERENCES

1. Todd, A. E., Orengo, C. A., and Thornton, J. M. (2001) *J. Mol. Biol. 307*, 1113−43.
2. Amstutz, P., Forrer, P., Zahnd, C., and Pluckthun, A. (2001) *Curr. Opin. Biotechnol. 12*, 400−405.
3. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Pluckthun, A., and Virnekas, B. (2000) *J. Mol. Biol. 296*, 57−86.
4. Winter, G., Griffiths, A. D., Hawkins, R. E., and Hoogenboom, H. R. (1994) *Annu. Rev. Immunol. 12*, 433−55.
5. Griffiths, A. D., and Duncan, A. R. (1998) *Curr. Opin. Biotechnol. 9*, 102−8.
6. Nozaki, S., Tomioka, Y., Hishinuma, T., Inoue, M., Nagumo, Y., Tsuruta, L. R., Hayashi, K., Matsumoto, T., Kato, Y., Ishiwata, S., Itoh, K., Suzuki, T., Hirama, M., and Mizugaki, M. (2002) *J Biochem (Tokyo) 131*, 729−38.
7. Schaus, S. E., Cavalieri, D., and Myers, A. G. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 11075−80.
8. Sudhahar, G. C., Balamurugan, K., and Chin, D. H. (2000) *J. Biol. Chem. 275*, 39900−6.
9. Urbaniak, M. D., Muskett, F. W., Finucane, M. D., Caddick, S., and Woolfson, D. N. (2002) *Biochemistry 41*, 11731−9.
10. Kappen, L. S., Xi, Z., and Goldberg, I. H. (2001) *Biochemistry 40*, 15378−83.
11. Maeda, H., Kumagai, K., and Ishida, N. (1966) *J. Antibiot. (Tokyo) 19*, 253−9.
12. Kappen, L. S., Napier, M. A., and Goldberg, I. H. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1970−4.
13. Heyd, B., Lerat, G., Adjadj, E., Minard, P., and Desmadril, M. (2000) *J. Bacteriol. 182*, 1812−8.
14. Maeda, H. (2001) *Adv. Drug Deliv. Rev. 46*, 169−85.
15. Skerra, A. (2001) *J. Biotechnol. 74*, 257−75.
16. Kim, K. H., Kwon, B. M., Myers, A. G., and Rees, D. C. (1993) *Science 262*, 1042−6.
17. Adjadj, E., Quiniou, E., Mispelter, J., Favaudon, V., and Lhoste, J. M. (1992) *Eur. J. Biochem. 203*, 505−11.
18. Gao, X. (1992) *J. Mol. Biol. 225*, 125−35.
19. Schier, R., Bye, J., Apell, G., McCall, A., Adams, G. P., Malmqvist, M., Weiner, L. M., and Marks, J. D. (1996) *J. Mol. Biol. 255*, 28−43.
20. Nieba, L., Krebber, A., and Pluckthun, A. (1996) *Anal. Biochem. 234*, 155−65.
21. Karlsson, R. (1994) *Anal. Biochem. 221*, 142−51.
22. Lowman, H. B., Bass, S. H., Simpson, N., and Wells, J. A. (1991) *Biochemistry 30*, 10832−8.
23. Schlehuber, S., Beste, G., and Skerra, A. (2000) *J. Mol. Biol. 297*, 1105−20.
24. Levitan, B. (1998) *J. Mol. Biol. 277*, 893−916.
25. Friguet, B., Chaffotte, A. F., Djavadi-Ohaniance, L., and Goldberg, M. E. (1995) *J. Immunol. Methods 182*, 145−50.
26. Martineau, P., Jones, P., and Winter, G. (1998) *J. Mol. Biol. 280*, 117−27.
27. Hanes, J., Jermutus, L., and Pluckthun, A. (2000) *Methods Enzymol. 328*, 404−30.
28. Nygren, P., and Uhlen, M. (1997) *Curr. Opin. Struct. Biol. 7*, 463−469.
29. Smith, G. (1998) *Trends Biochem. Sci. 23*, 457−460.
30. Vita, C. (1997) *Curr. Opin. Biotechnol. 8*, 429−34.
31. Nord, K., Gunneriusson, E., Ringdahl, J., Stahl, S., Uhlen, M., and Nygren, P. A. (1997) *Nat. Biotechnol. 15*, 772−7.
32. Legendre, D., Soumillion, P., and Fastrez, J. (1999) *Nat. Biotechnol. 17*, 67−72.

BI0273664