Stabilizing C-Terminal Tails on AraC

Madhushree Ghosh and Robert F. Schleif*
Department of Biology, Johns Hopkins University, Baltimore, Maryland

ABSTRACT **We examined the effects of the metabolic stability of random sequences appended to the C-terminus of the dimerization domain of the regulatory protein of the *Escherichia coli* arabinose operon, AraC. Genetic scoring utilized the *trans* dominant negative effect of the dimerization domain on the activity of intact AraC, and physical scoring used sodium dodecyl sulfate (SDS) gel electrophoresis. We confirmed previous results obtained with Arc and lambda repressors that C-terminal charged residues tend to be stabilizing and that hydrophobic residues are destabilizing. Additionally, we found that the provision of a single, charged C-terminal residue conferred significant stability that was independent of interior sequence. Hence, it appears that in the engineering of proteins, flexible tails may be freely added, with only the identity of the C-terminal amino acid being restricted. Proteins 2001;42:177–181.**
© 2000 Wiley-Liss, Inc.

Key words: protein stability; AraC; complementation

INTRODUCTION

A number of factors affect the *in vivo* lifetimes of proteins in bacteria, which vary from minutes to hours. [1–3] These include the state of the protein, as damaged and abnormal proteins are metabolically unstable, the physiological state of the cell, and the intrinsic properties of the proteins themselves.[1,3–6] Extensive studies have shown that both N-terminal and C-terminal sequences can dramatically affect proteins' metabolic stabilities by affecting the proteins' sensitivity to intracellular proteases.[4,7–11]

In *Escherichia coli*, N-terminal L, F, Y, W, R, and K destabilize proteins,[4,6,12,13] and at the C-terminus, Sauer and coworkers showed with studies on Arc repressor and lambda repressor that nonpolar residues are destabilizing and polar residues are stabilizing,[4,7,10,11,14,15] with the stabilizing effect of a single, charged residue diminishing with increasing distance away from the C-terminus.[4,10] Internal sequences have also been shown to affect protein stabilities.[16,17,18]

Our work on the regulation mechanism employed by the AraC protein in *E. coli* shows that an arabinose-regulated arm–domain interaction plays a key role.[19,20] Because the use of phage display libraries facilitates the identification of peptides that bind to proteins and enzymes in the presence or absence of ligands, it is tempting to attempt to add specific N-terminal or C- terminal arms to proteins.[16,19,20] These arms would produce specific protein–protein interactions or regulate enzyme action analogous to the light-switch mechanism seen in AraC.[19–21]

Because of our interest in designing arms to proteins, we have explored more deeply the effect of sequence at the C-terminus on protein stability.[19–21] In particular, we wished to determine whether the provision of a single polar C-terminal residue is sufficient to provide high protein metabolic stability and whether the stability rules as determined by Sauer and coworkers[9–11,15–18] would hold true for the dimerization domain of AraC.

MATERIALS AND METHODS

Enzyme Assays and DNA Sequencing

Arabinose isomerase levels were assayed as described.[19,22] The results shown are averages of at least three assays. Cell cultures were grown at 37 °C in an M10 medium[22] to an apparent optical density (OD) of 0.4–0.9 in 0.4% glycerol (v/v), casamino acids (0.4% w/v), and thiamine (10 μg/mL) in the presence and absence of 1.3% arabinose.[22]

All candidates with tail sequences linked at the carboxy terminal were sequenced with ^{35}S radioactive-labeled primers as described in the SequiTherm EXCEL II DNA Sequencing Kit from Epicentre Technologies (WI).

Oligonucleotides

Oligonucleotides used as primers (Table I) for cloning, sequencing, and polymerase chain reaction (PCR) were synthesized on an Applied Biosystems 381A synthesizer, deprotected,[23] and purified as described previously.[24] Oligonucleotides A and B were commercially synthesized with standard purification at a 200 nmole concentration by Integrated DNA Technologies, Inc. (IA), were dissolved in 100 mM Tris HCl buffer (pH 7.5), and were used without further manipulation or purification.

Plasmids

Plasmid pWR03 is pSE380 (Invitrogen, San Diego, CA) containing the entire *araC* gene.[19] This was used to construct the AraC dimerization domain in plasmids pMG04 and pMG08. Plasmid pMG04, containing residues 1–170 of the dimerization domain gene inserted between *Nco*I and *BamH*I sites of pSE380, was constructed via PCR

Grant sponsor: NIH; Grant number: GM18277

*Correspondence to: Johns Hopkins University. Department of Biology, 3400 North Charles Street, Baltimore, MD 21218. E-mail: bob@gene.bio.jhu.edu

Received 23 June 2000; Accepted 3 October 2000

© 2000 WILEY-LISS, INC.

TABLE I. Oligonucleotide Primers Used for Cloning, PCR Amplification, and Cloning†

Primer	Sequence	Sequence Homology/Use
2328	CATGCCATGGATGAAGCGCACCAGCCGCAT	pWR03: 367–387; sequencing, PCR
2327	ATAACAATTTCACACAGGAAACAG	pSE380: 249–273; cloning
2310	CGCGGATCCGGTGCCTTA	*BamH*I in A and B; PCR extension; sequencing
2325	CGCGGGATCCTTACGACTAGTTAATCGC	pWR03: 496–511; cloning stop codon with *BamH*I site in pWR03
2331	CGCGGATCCCGACTCGTTAAT	pWR03: 496–511; cloning *BamH*I site in pMG04/pWR03 without stop codon
A	CGCGGATCCGGGTTA$(NNN)_{12}$TAAGGCACCGGATCCGCG	Primer for random amino acid introduction beyond residue 170 in C protein
B	CGCGGATCCGGGTTA$(NNN)_{11}$CGTTAAGGCACCGGATCCGCG	Primer for random peptide generation similar to A ending with codon for arginine at C terminal end of C dimerization domain

†All sequences are listed 5′ through 3′.

amplification of pWR03 as a template with primers 2031 and 2325. These primers introduced an *Nco*I and a stop codon preceding a *BamH*I site. pMG08 was constructed similarly with primers 2031 and 2331. Oligonucleotides A and B were ligated into the C terminus of pMG08 after *BamH*I digestion.

Strains

Dimerization domain candidates with tails of various stabilities were selected by the transformation of pMG08 attached to A or B into strain SH322 [*ara*$(CBAD)^+$ *leu* Δ*lac*74 *galKStr*^{r}].[3] Arabinose isomerase levels of the candidates were determined in SH321 (Δ*araC-leu1022*, *araB*$^+$*A*$^+$*D*$^+$ Δ*lac*74 *galKStr*^{r}).[24]

Generation of AraC Dimerization Domain Gene Fused to Oligonucleotides Generating Tails of Various Stabilities

A 3-μg aliquot of plasmid pMG08 was digested with 15,000 units of *BamH1* for 90 min at 37 °C. The digested plasmid was run on a 1.5% agarose gel, excised, and purified by phenol-chloroform extraction, ethanol precipitation, and finally a Geneclean II kit from Bio101. Oligonucleotides A and B were PCR-extended (one cycle) with primer 2310 in a 1/4 molar ratio. The entire PCR reaction was digested overnight with 20,000 units of *BamH*I, phenol-chloroform-extracted, ethanol-precipitated, and then ligated to *BamH*I-digested pMG08.

Ligation reactions were performed at room temperature for 3 h via the ligation of a 2.5 molar excess of fragment with plasmid. Reaction products were then transformed in SH322 and plated on tetrazolium/ampicillin plates containing 0.03 mM isopropyl-thio-β-galactopyranoside (IPTG). This concentration of IPTG was chosen as the minimum that would generate negative complementation with a dimerization domain lacking an appended C-terminal tail. Control ligation reactions were also plated on yeast extract tryptone (YT)/ampicillin plates for corroboration of the extent of ligation of the plasmid and PCR-digested product.

Identification of Dimerization Domains with Ligated Tails of Various Stabilities

To isolate candidates containing inserts of oligonucleotides generating peptides of various lengths and stabilities, colonies were selected and tested for the presence of inserted oligonucleotides by PCR amplification with primers 2310 and 2328. Only candidates with insertions can yield PCR products. All candidates were isolated on the basis of this selection method and were further tested and sequenced with ^{35}S sequencing methods.

After the sequencing, the candidates were tested for the solubility of AraC containing the randomized C-terminal tail. Candidate proteins were overexpressed through the growth of cells with 1 mM IPTG. One milliliter of the cells of each candidate was centrifuged, resuspended in 300 μL of 100 mM phosphate buffer (pH 7.4), 50 mM KCl, 10% glycerol, 1 mM EDTA, 1 mM dithioerythritol (DTE), 0 and .5 M phenyl methyl-sulfonyl fluoride, and lysed by sonication. The cell debris was pelleted via centrifugation of the lysed cells at 13,000 X g for 5 min. For each candidate, a 10-μL aliquot of the supernatant was then electrophoresed on a 14% resolving SDS gel.

RESULTS

To perform these studies, we took advantage of the fact that a high synthesis level of just the dimerization domain of AraC, the regulatory protein of the arabinose operon, has a *trans* dominant negative effect on the inducibility of the arabinose operon (Fig. 1).

Intact dimers of AraC are required to induce the arabinose catabolite genes. The presence of an excess of AraC dimerization domains forces most of the full-length wild-type protein made from a chromosomal gene copy into inactive heterodimers. Hence, cells with a wild-type chromosomal AraC gene and an overproducing AraC dimerization domain encoded by a plasmid do not contain functional AraC protein. If the addition of a C-terminal tail to the dimerization domain relieves its *trans* dominant negative effect, the resulting protein is no longer able to form the inactive heterodimers. This can result either from instability of the protein or from its insolubility.

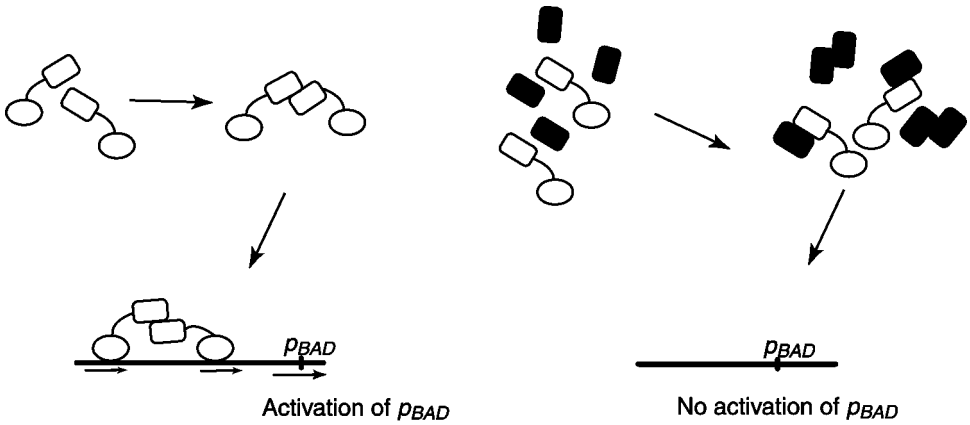

Fig. 1. Negative complementation of wild-type AraC. The activation of p_{BAD} is promoted by the AraC protein binding as an intact functional protein to DNA. An excess of dimerization domain forces intact AraC into inactive dimers. Thus, the *trans* dominant negative effect of the heterodimers prevents p_{BAD} activation.

TABLE II. Red Candidates Obtained from Tetrazolium/IPTG/Ampicillin Screening of the AraC Dimerization Domain With a Stabilizing Peptide Tail of Random Sequence[†]

Candidate[a]	Sequence	Polar C-Terminal[b]	Isomerase Levels	
			− ara	+ ara
1	-GSGALIWRPALNTAFAQ	+	0	300
2	-GSGLSAPRLIMNRSKAPDPMQ	+	0	2,000
3	-GSRLLIQRVPRPRINKISQ	+	30	40
4	-GSGALLNLTCLLDALTTH	+/−	0	100
5	-GSGALGPSCSTQQLLTSS	+/−	0	200
6	-GSGYRSCRPFFMKCI	−	0	250
7	-GSGL	−	0	200
8	-GSGLALRSGLI	−	0	20

[†]Isomerase levels of the dimerization domains are also listed as units in the absence and presence of arabinose (see the Discussion). − and + ara refer to the assays of cells grown in the absence and presence of arabinose.
[a]Stabilized domains show the formation of inactive heterodimers and negative complementation except for Candidate 2, which shows units similar to white candidates; see the Discussion.
[b]The polarity/hydrophobicity/neutral character of the final residue was assessed as the sole determinant of stability for the entire protein and was compared to the data obtained by Sauer and coworkers.[10]

By cloning DNA fragments, coding them for random-sequence C-terminal extensions to the dimerization domain, transforming them into $AraC^+$ cells, and screening transformed colonies for their ability to catabolize arabinose, we isolated candidates whose peptide extensions led either to uninducibility or inducibility of the arabinose operon. The screening for the ability of colonies to catabolize arabinose was performed with tetrazolium arabinose color-indicating plates. Colonies can grow on such plates whether or not they catabolize arabinose. The $AraC^+$ (able to catabolize arabinose) colonies were white, and the $AraC^-$ colonies were red. $AraC^-$ or uninducible candidates would likely result from arms or tails that conferred sufficient metabolic stability to the dimerization domains that would form heterodimers with wild-type AraC. $AraC^+$ or inducible candidates would result from added arms or tails whose sequences led either to instability of the dimerization domain or insolubility of the dimerization domain.

Isolation of Candidates With Stabilizing Tails

Colonies containing plasmids encoding stabilizing or destabilizing peptides attached to the dimerization domain of AraC were selected on the basis of their color on the tetrazolium/ampicillin plates: red for uninducible, *trans* dominant negative, and, hence, stabilizing peptide, and white for inducible, the peptide conferring instability or insolubility. Candidates were then screened by PCR amplification of the regions; we took advantage of the fact that the first twelve nucleotides on the tail were fixed to code for -GSGL and the twelve nucleotides after the final twelfth randomized codon were also fixed. Although we provided for a randomized peptide of twelve amino acids, seven of the candidates possessed stop codons within the twelve and were, therefore, shorter. Tables II and III show the candidates we found with stabilizing and destabilizing or desolubilizing tails.

To obtain quantitative measures of the activities of the wild-type AraC remaining after negative complemen-

TABLE III. White Candidates Obtained via Tetrazolium/IPTG/Ampicillin Screening of the AraC Dimerization Domains with a Destabilizing or Desolubilizing Peptide Tail of Random Sequence[†]

Candidate[a]	Sequence	Nonpolar C-Terminal[b]	Soluble	Isomerase Levels − ara	Isomerase Levels + ara
1	-GSGALYVVIVTYL	+	+	0	10,000
2	-GSGLIIVCFG	+/−	−	0	20,000
3	-GSGALIISMYIKMTHA	+/−	−	0	2,000
4	-GSGIIIVCFG	+/−	+	0	1,000
5	-GSGLIWRSGY	+/−	+	0	1,300
6	-GSGLHALKVLRIGNFH	+/−	−	0	8,000
7	-GSGLVCFVIPVYNPDT	+/−	−	0	4,000
8	-VVNVSPATYYTPEY	+/−	−	50	16,000
9	-GSGLLIRYIMLLTVLR	−	−	0	5,000

[†] − and + ara refer to the assays of cells grown in the absence and presence of arabinose.
[a]All overexpressed candidates in the supernatant were considered soluble, and all candidates present in the pellet were considered insoluble. See Figure 2 and the Discussion.
[b]The polarity/hydrophobicity/neutral character of the final residue was assessed as the sole determinant of stability for the entire protein and was compared to the data obtained by Sauer and coworkers.[10]

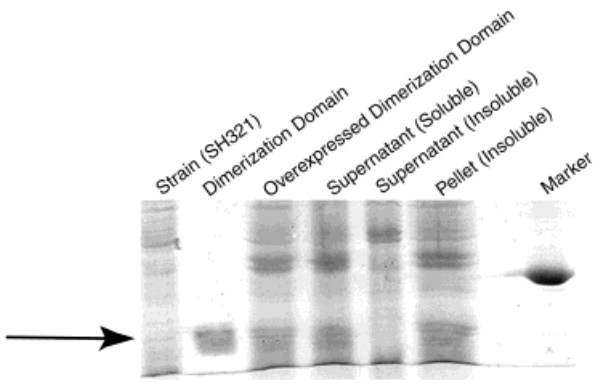

Fig. 2. SDS–PAGE gel (14%) showing soluble and insoluble dimerization domains. Approximately 3 μg of protein was loaded in each lane. Lane 7 shows carbonic anhydrase with a molecular weight of 32,000. The arrow indicates the position of the C dimerization domains in each lane.

tation with the dimerization domain plus tail, we measured the inducibility of the arabinose operon by assay of arabinose isomerase levels. Tables II and III also show these results. With the exception of the red candidate 2, all candidates found to be negatively complementing showed low levels of arabinose isomerase. All candidates that rendered the domain metabolically unstable or insoluble showed high isomerase levels in the presence of arabinose.

Isolation of Candidates With Tails Promoting Instability or Insolubility

As the dimerization domain of AraC can be made unavailable for negative complementation either by being metabolically unstable or by being insoluble, we also examined the solubility of the candidates that did not negatively complement. A representative SDS gel is shown in Figure 2. Without any exceptions, all red candidates promoting negative complementation yielded soluble dimerization domains. White candidates 2, 3, 6, 7, 8, and 9 were insoluble and, therefore, unavailable for negative complementation. Candidates 1, 4, and 5, however, were soluble and yet did not promote negative complementation; see the Discussion.

Tails with Single Polar Residue at Carboxy Terminal End

Is a single C-terminal polar residue sufficient to stabilize the dimerization domain of AraC, independent of the preceding residues? The published work on the metabolic stability of proteins and the work reported in this article do not explicitly address this question.[4,10] We tested this question by fixing the final residue as arginine, randomizing the preceding eleven residues, and selecting red and white colonies as described previously. Three candidates were obtained, and all were red, indicating that the dimerization domain was stable.

DISCUSSION

We used two approaches to investigate factors relevant to the addition of C-terminal tails of arbitrary sequence to the dimerization domain of AraC. First, we added twelve amino acids of random sequence and examined additions that conferred metabolic stability and solubility or instability or insolubility. Second, we examined whether a C-terminal polar residue was sufficient to confer stability.

We used oligonucleotides to generate tails of random sequence, negative complementation with wild-type AraC, and tetrazolium indicator plates to distinguish candidates with stabilized and destabilized or insoluble AraC protein. Initial investigations focused on studying the candidates that allowed negative complementation by stabilizing the domain. As shown in Table II, the first three candidates possess glutamine (charged/polar) as their final residue. This contributes to the stability of the tail, subsequently preventing protease action as reported earlier.[4] Candidates 4 and 5 have histidine and serine (neutral/moderately polar) as their final

residues, whereas Candidates 6, 7, and 8 have isoleucine (nonpolar/hydrophobic) as their final residues. Candidates 6, 7, and 8, however, also possess polar and charged residues near the final nonpolar/hydrophobic residue. These data confirm previously determined results that neutral and even nonpolar residues at the extreme end of the flexible carboxy terminal tail may stabilize a protein, if polar residues precede the final residue.[4,10] Possibly the more polar amino acids adjacent to the C-terminus confer sufficient stability on the dimerization domain such that it can negatively complement wild-type AraC protein.

Next, we isolated candidates that failed to generate negative complementation with wild-type AraC. These would be candidates that are either so metabolically unstable or so insoluble that too little of the dimerization domain would be available to dimerize with wild-type AraC and generate negative complementation. Most of these candidates indeed possess destabilizing C-terminal residues or are found in inclusion bodies indicative of insolubility. We do not understand the origin of the higher molecular weight species also seen on the SDS gels (Fig. 2), whose presence parallels that of the dimerization domain band. Candidates 1, 4, and 5 are perplexing, however, as they appeared to be soluble according to our inclusion body test, which simultaneously shows that they are not highly unstable, but they failed to generate negative complementation. Perhaps these proteins are insoluble but form aggregates that are not readily pelleted, or the proteins are solubilized by the cell lysis sonication steps preceding electrophoresis.

Finally, by fixing the final residue, arginine, as polar, we determined whether this constraint could stabilize the protein, independent of the preceding eleven residues. For unknown reasons, we encountered considerable difficulty in this experiment and obtained only three candidates. The three candidates were domains with the tails -GSGLQVGTCFIPFFMTR, -GSGALRISD, and -GSGLVSAEN. All of them ended with a final residue that was polar and/or charged in nature (arginine, aspartic acid, and asparagine, respectively). All negatively complemented, suggesting, but certainly not proving, that a C-terminal polar residue such as arginine is sufficient to stabilize the dimerization domain of AraC.

These data indicate, but do not definitively prove, that the provision of a C-terminal arginine is sufficient to confer stability on the dimerization domain of AraC. As this result extends work previously reported by others on other proteins, it is reasonable to assume that one may rather freely engineer proteins via the addition of C-terminal tails as long as they end in highly polar amino acids. Undoubtedly, however, in certain cases interior amino acids may generate metabolic instability.

ACKNOWLEDGMENTS

The authors thank Tara Harmer for helpful discussions and Urszula Gryczynski for technical assistance. This work was supported by an NIH grant to R.F. Schleif.

REFERENCES

1. Bachmair A, Finley D, Varshavsky A. *In vivo* half-life of a protein is a function of its amino terminal residue. Science 1986;234:179–186.
2. Goldberg AL, St John AC. Intracellular protein degradation in mammalian and bacterial cells: Part 2. Annu Rev Biochem 1976;45:747–803.
3. Levchenko I, Smith CK, Walsh NP, Sauer RT, Baker TA. PDZ-like domains mediate binding specificity in the Clp/Hsp100 family of chaperones and protease regulatory subunits. Cell 1997;91:939–947.
4. Parsell DA, Sauer RT. The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. J Biol Chem 1989;264:7590–7595.
5. Shrader TE, Tobias JW, Varshavsky A. The N-end rule in *Escherichia coli*: Cloning and analysis of the leucyl, phenylalanyl-tRNA protein transferase gene *aat*. J Bacteriol 1993;175:4364–4374.
6. Varshavsky A. The N-end rule. Cell 1992;69:725–735.
7. Gottesman S, Roche E, Zhou YN, Sauer RT. The ClpXP and ClpAP proteases degrade proteins with C-terminal peptide tails added by the SsrA tagging system. Genes Dev 1998;12:1338–1347.
8. Keiler KC, Silber KR, Downard KM, Papayannopoulos IA, Biemann K, Sauer RT. C-terminal specific protein degradation: Activity and substrate specificity of the Tsp protease. Protein Sci 1995 4:1507–1515.
9. Keiler KC, Sauer RT. Sequence determinants of C-terminal substrate recognition by the Tsp protease. J Biol Chem 1996;27:2589–2593.
10. Parsell DA, Silber KR, Sauer RT. Carboxy-terminal determinants of intracellular protein degradation. Genes Dev 1989;4:277–286.
11. Silber KR, Keiler KC, Sauer RT. Tsp: A tail-specific protease that selectively degrades proteins with nonpolar C termini. Proc Natl Acad Sci U S A 1992;89:295–299.
12. Johnson ES, Gonda DK, Varshavsky A. *Cis–trans* recognition and subunit-specific degradation of short-lived proteins. Nature 1990;346:287–291.
13. Tobias JW, Shrader TE, Rocap G, Varshavsky A. The N-end rule in bacteria. Science 1991;254:1374–1377.
14. Gottesman S, Maurizi MR, Wickner S. Regulatory subunits of energy-dependent proteases. Cell 1997;91:435–438.
15. Jubete Y, Maurizi MR, Gottesman S. Role of the heat shock protein DnaJ in the Lon-dependent degradation of naturally unstable proteins. J Biol Chem 1996;271:30798–30803.
16. Bowie JU, Sauer RT. Identification of C-terminal extensions that protect proteins from intracellular proteolysis. J Biol Chem 1989;264:7596–7602.
17. Smith CK, Baker TA, Sauer RT. Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. Proc Natl Acad Sci U S A 1999;96:6678–6682.
18. McDonald JP, Peat TS, Levine AS, Woodgate R. Intermolecular cleavage by UmuD-like enzymes: Identification of residues required for cleavage and substrate specificity. J Mol Biol 1999;285:2199–2209.
19. Reed WL, Schleif RF. Hemiplegic mutations in AraC protein. J Mol Biol 1999;294:417–425.
20. Schleif RF. Arm–domain interactions in proteins: A review. Proteins 1999;34:1–3.
21. Saviola B, Seabold R, Schleif R. Arm–domain interactions in AraC. J Mol Biol 1998;278:539–548.
22. Schleif RF, Wensink P. Practical methods in molecular biology. New York: Springer-Verlag; 1981. p 46–50.
23. Sawadogo M, Van Dyke M. A rapid method for the purification of deprotected oligonucleotides. Nucleic Acids Res 1991;19:674.
24. Cubellis M, Marino G, Maynol L, Piccialli G, Sannia G. Use of fast protein liquid chromatography for the purification of synthetic oligonucleotides. J Chromatogr 1985;329:406–414.