Chemical and Enzymatic Synthesis of Gene Fragments

A Laboratory Manual

Edited by H. G. Gassen and Anne Lang

Chemical and Enzymatic Synthesis of Gene Fragments

A Laboratory Manual

Edited by H. G. Gassen and Anne Lang

verlag chemie

Weinheim · Deerfield Beach, Florida · Basel · 1982

Prof. Dr. Hans G. Gassen
Dipl.-Chem. Anne Lang
Institut für Organische Chemie und Biochemie
der Technischen Hochschule
Petersenstr. 22
D-6100 Darmstadt

Publisher's editor: Hans F. Ebel

Production manager: Peter J. Biel

This book contains 69 figures and 15 tables

Deutsche Bibliothek Cataloguing-in-Publication Data

Chemical and enzymatic synthesis of gene fragments:
a laboratory manual / ed. by Hans G. Gassen and Anne Lang. –
Weinheim; Deerfield Beach, Florida; Basel:
Verlag Chemie, 1982
ISBN 3-527-26063-3
NE: Gassen, Hans G. [Hrsg.]

Printing: betz-druck gmbh, D-6100 Darmstadt 12
Bookbinding: Josef Spinner GmbH, D-7583 Ottersweier
Printed in the Federal Republic of Germany

Preface

In the past the chemical and enzymatic synthesis of oligonucleotides of defined sequence had to be left to a few experts. Now, however, with the triester approach, the phosphite method and the solid-support techniques gene fragment synthesis has turned into an easy procedure even for a non-chemist. Due to the elegant chemistry involved, all methods work without sophisticated equipment and are prone to mechanisation and eventual automation. It is hoped that combined chemical-enzymatic gene synthesis may become a standard technique in a molecular biology laboratory, such as DNA sequencing or in-vitro recombination of nucleic acids.

We omitted chemical RNA synthesis, since this field is developing so rapidly at the moment that one has to refer to the original publications. However, we included enzymatic synthesis of RNA fragments, procedures which already have obtained a high degree of standardisation.

Most of the contributions are revised versions of the protocols supplied for the EMBO sponsored course on "Automated Chemical and Enzymic Gene Synthesis", held in Darmstadt, March 21 to April 3, 1982. The protocols were improved on the basis of the experience of 30 student scientists with chemical, biological or medical backgrounds. Previously omitted procedures, such as the wandering spot method for oligonucleotide analysis, were included. In editing the manuscript we encountered problems with the nomenclature of nucleic acid components. In unambiquous cases we favoured a simple description, hoping for example, that oligodeoxynucleotide is always understood to mean oligo-2′-deoxyribonucleotide.

This book aims to provide those interested in DNA/RNA research with state-of-the-art methods in the synthesis, purification, and analysis of DNA and RNA fragments. The editors wish to thank the authors for their efforts in preparing manuscripts from the revised laboratory protocols. We gratefully acknowledge the skill and the patience of Mrs. E. Rönnfeldt in typing the manuscripts.

We express our thanks to Verlag Chemie for the friendly and very efficient cooperation.

Darmstadt, in July 1982

H.G. Gassen
A. Lang

List of Contributors

Blöcker, Helmut see Frank, R.

van Boeckel, C.A.A. see van Boom, J.H.

van Boom, Jacques H., G.A. van der Marel, C.A.A. van Boeckel, G. Wille, and *C.F. Hoyng* (see extra entry)
Department of Organic Chemistry
Gorlaeus Laboratories
University of Leiden
P.O. Box 9502
2300 RA Leiden
The Netherlands

Caruthers, Marvin H.
Department of Chemistry
University of Colorado
Boulder, CO 80309
U.S.A.

Cook, Ronald M., D. Hudson, E. Mayran, and *J. Ott*
Biosearch
1281-F Andersen Drive
San Rafael, CA 94901
U.S.A.

Dörper, T. see Winnacker, E.-L.

Eckert, V., see also Ohtsuka, E.
Fachgebiet Biochemie
Technische Hochschule Darmstadt
Petersenstr. 22, D-6100 Darmstadt
Federal Republic of Germany

Eiband, J. see Seliger, H.

Eick, D. see Fritz, H.J.

Frank, Ronald, and *H. Blöcker*
GBF, Gesellschaft für Biotechnologische Forschung mbH
Mascheroder Weg 1, D-3300 Braunschweig
Federal Republic of Germany

Fritz, Hans J., W.-B. Frommer, W. Kramer, and *W. Werr*
Institut für Genetik, Universität zu Köln
Weyertal 121, D-5000 Köln 41
Federal Republic of Germany

Fritz, Hans J., D. Eick, and *W. Werr*
Institut für Genetik, Universität zu Köln
Weyertal 121, D-5000 Köln 41
Federal Republic of Germany

Frommer, Wolf-Bernd see Fritz, H.J.

Gait, Michael J., H.W.D. Matthes, M. Singh, B.S. Sproat, and *R.C. Titmas*
Laboratory of Molecular Biology
Medical Research Council Centre
University Medical School
Hills Road, Cambridge, CB2 2QH
U.K.

Gassen, Hans G. (several contributions), *V. Eckert, C. Gatz, W. Hillen, G. Klock, A. Lang,* and *M. Schmitt*
Fachgebiet Biochemie
Technische Hochschule Darmstadt
Petersenstr. 22, D-6100 Darmstadt
Federal Republic of Germany

Gatz, Christiane, and *W. Hillen*
Fachgebiet Biochemie
Technische Hochschule Darmstadt
Petersenstr. 22, D-6100 Darmstadt
Federal Republic of Germany

Hauel, N. see Seliger, H.

Hillen, Wolfgang
Fachgebiet Biochemie
Technische Hochschule Darmstadt
Petersenstr. 22, D-6100 Darmstadt
Federal Republic of Germany

Hoyng, C.F.
Genentech, Inc.
South San Francisco, CA 94080
U.S.A.
(*see also* van Boom, J.H.)

Hudson, D. see Cook, R.M.

Klein, S. see Seliger, H.

Klock, Gerd
Fachgebiet Biochemie
Technische Hochschule Darmstadt
Petersenstr. 22, D-6100 Darmstadt
Federal Republic of Germany

Kramer, W. see Fritz, H.J.

Krusche, Jörg U., see also McLaughlin, L.W.
Dupont de Nemours (Deutschland) GmbH
Abteilung Analytische Instrumente
Dieselstr. 18, D-6350 Bad Nauheim
Federal Republic of Germany

Lang, Anne, and *H.G. Gassen*
Fachgebiet Biochemie
Technische Hochschule Darmstadt
Petersenstr. 22, D-6100 Darmstadt
Federal Republic of Germany

van der Marel, G.A. see van Boom, J.H.

Matthes, H.W.D. see Gait, M.J.

Mayran, E. see Cook, R.M.

McLaughlin, Larry W., and *J.U. Krusche*
(see extra entry)
Max-Planck-Institut für experimentelle Medizin
Hermann-Rein-Str. 3, D-3400 Göttingen
Federal Republic of Germany

Narang, C.K. see Seliger, H.

Ohtsuka, Eiko, and *V. Eckert*
(see extra entry)
Faculty of Pharmaceutical Sciences
Osaka University
1–6 Yamadaoka, Suita
Osaka, Japan 565

Ott, J. see Cook, R.M.

Schmitt, Marion, and *H.G. Gassen*
Fachgebiet Biochemie
Technische Hochschule Darmstadt
Petersenstr. 22, D-6100 Darmstadt
Federal Republic of Germany

Schott, Herbert
Institut für Organische Chemie
Universität Tübingen
Auf der Morgenstelle 18, D-7400 Tübingen
Federal Republic of Germany

Seemann-Preising, B. see Seliger, H.

Seliger, Hartmut, S. Klein, C.K. Narang,
B. Seemann-Preising, J. Eiband, and *N. Hauel*
Sektion Polymere, Universität Ulm
Oberer Eselsberg, D-7900 Ulm
Federal Republic of Germany

Singh, M. see Gait, M.J.

Sproat, B.S. see Gait, M.J.

Titmas, R.C. see Gait, M.J.

Uhlenbeck, Olke C.
Department of Biochemistry
School of Chemical Sciences
University of Illinois
Urbana, Illinois 61801
U.S.A.

Werr, W. see Fritz, H.J.

Wille, G. see van Boom, J.H.

Winnacker, Ernst-Ludwig, and *T. Dörper*
Institut für Biochemie der Ludwig-Maximilians-Universität
Karlstr. 23, D-8000 München 2
Federal Republic of Germany

Contents

Chemical Synthesis of Polydeoxyribonucleotides of Defined Sequence

Biochemical Modification of Gene Fragments

SYNTHESIS OF OLIGODEOXYRIBONUCLEOTIDES BY A CONTINUOUS FLOW, PHOSPHOTRIESTER METHOD ON A KIESELGUHR/POLYAMIDE SUPPORT

Michael J. Gait, Hans W.D. Matthes, Mohinder Singh, Brian S. Sproat, and Richard C. Titmas

Laboratory of Molecular Biology
Medical Research Council Centre
Cambridge

SUMMARY

A new composite polydimethylacrylamide/kieselguhr support is used in a continuous flow assembly apparatus for the efficient solid phase synthesis of a deca- and a nonadecadeoxyribonucleotide using phosphotriester intermediates.

1 Introduction

There are few scientists these days interested in DNA who have not yet heard of solid phase synthesis of DNA (if nowhere else in advertisements for DNA synthesis machines in Nature). In 1975 when Hartmut Seliger gave an EMBO Course entitled "Synthesis of oligonucleotides on polymer-supports" it was a different story. Then few DNA chemists gave solid phase synthesis much chance of success and few would contemplate its use beyond tri- or tetranucleotides. Automated synthesis was merely a distant dream.

However, the principle of solid phase synthesis has been recognised since its inception in the mid-1960s [1] as an extremely elegant one. It is particularly suitable for preparation of biopolymers of defined sequence such as oligonucleotides and peptides. In essence one end of a growing chain is protected by an insoluble, macromolecular protecting group, whilst the other end is reacted with the next unit to be coupled (i.e. Ⓟ -A-B + C-x ⟶ Ⓟ -A-B-C-x). Any excess of C-x can be removed by mere filtration and washing. After selective removal of protecting group x the process can be repeated by addition of the next unit D-x and so on (Ⓟ -A-B-C-X ⟶ Ⓟ -A-B-C + D-x ⟶ Ⓟ -A-B-C-D-x). At the end of the synthesis the bond between the chain and support Ⓟ is broken and the chain is released into solution and purified by chromatography.

Whereas by the mid-1970s short peptides in many cases could be assembled well, and indeed commercial peptide synthesis machines were already available, the coupling of nucleotides on solid phase did not seem to be as efficient as in conventional solution reactions, such that yields of desired products became vanishingly small. A full discussion of the various polymer-supports and methods tried can be found in some recent review articles [2]. It is sufficient here to state that there were two basic reasons for these failures: 1. the supports chosen were not appropriate to the chemical reactions used for DNA synthesis and 2. the chemical reactions used were insufficiently selective and reliable to meet the two extra demands of the solid phase approach.

First these are heterogeneous reactions which are liable to be slower than their solution counterparts. The minimum increase in $t_{1/2}$ is probably about 2-3 times under optimum conditions but this can be much worse if there is any incompatability between support, solvent and chemistry. Thus the solid phase method works best for relatively fast reactions ($t_{1/2}$ measures in sec or a few min), where small rate changes do not significantly increase the assembly time and can be easily al-

lowed for. Secondly, it is common to push reactions to completion on solid phase by use of a substantial excess of added nucleotide in solution (i.e. obtaining essentially pseudo-first order kinetics). In this situation the slightest reactive impurity in the nucleotide in solution gets magnified by virtue of the excesses used (e.g. 1% impurity in a nucleotide used in 10-fold excess over growing chain can give up to 10% side reaction).

The classical phosphodiester approach to DNA synthesis, despite some considerable success in conventional solution methods, turned out not to be particularly suited to solid phase, even though many attempted it. Partly this was due to poor choice of polymer supports. For example, polystyrene gel is inappropriate because its non-polar nature does not match the highly polar nature of oligonucleotide chains with their negatively charged phosphodiesters. In addition phosphodiester chemistry is not sufficiently selective to give consistently high yields in coupling reactions, an important necessity for successful solid phase synthesis, and monomer units were difficult to obtain in high purity. Even so, some reasonable syntheses of oligodeoxyribonucleotides up to about 12 units were obtained using in particular polyamide supports [3,4].

Phosphotriester chemistry (including the companions "phosphite-triester" and "phosphoroamidite") is much more suitable for solid phase synthesis and in the last two years or so has proved to be highly successful. The oligonucleotide chains as they are assembled contain uncharged, fully protected phosphates and are found to be compatible in solvation properties with a variety of polymer supports (polyamide, polystyrene, silica). Also coupling reactions are much more selective and give higher yields.

2 EXPERIMENTAL OUTLINE

This contribution is designed to be an introduction to all the important techniques necessary to make oligodeoxyribonucleotides by the phosphotriester approach on solid phase. The methods described are those in current use in the author's laboratory and maximises the use of commercially available materials and equipment.

Two syntheses will be described: the decanucleotide, d(C-C-G-A-T-A-T-C-G-G). This is a self-complementary sequence that has a central six nucleotides which are the recognition sequence for the restriction endonuclease EcoRV. The oligonucleotide is wanted for an nmr study of its interaction with the endonuclease and also for crystallisation and X-ray analysis. The synthesis will demonstrate the use of monomer units

in couplings to solid phase on moderate scale (10 μmol) for the preparation of a few mg of decanucleotide.

The nonadecanucleotide, d(T-G-G-T-C-A-T-A-G-C-T-G-T-T-T-C-C-T-G): this sequence is complementary to a region of the single-stranded bacteriophage DNA M13 upstream of the normal DNA insertion sites used in the cloning and sequencing procedures of Sanger's group. The oligonucleotide is designed as a primer for the preparation of partial duplexes of M13 that leave inserted DNA as essentially single-stranded. This can be used as a hybridisation probe for the selection of new clones containing oppositely orientated inserts and to obtain overlap information that would otherwise be difficult to obtain. The synthesis will be on small scale (5 μmol) and will demonstrate the use of dinucleotide blocks as coupling units.

3 DISCUSSION OF METHODS

3.1 Polymer support and its functionalisation

3.1.1 The support

Most recently published methods of DNA synthesis by the phosphotriester route have used gel resins of either polystyrene [5], polyacryloylmorpholide [6] or our own polydimethylacrylamide copolymers [7,8,9]. These are solvent-swollen polymers that are generally handled in glass reaction vessels where solvents are added batchwise, followed by some sort of agitation and then filtration by application of partial vacuum or nitrogen pressure. Such techniques, although perfectly acceptable, suffer from the disadvantage that because of solvent hold-up in the gel, several washings are needed to remove a previous solvent or reagent. This batchwise technique is cumbersome to automate and can be time-consuming and expensive on solvents.

The most efficient way to wash a polymer-support is to pack it into a column and pass solvents through continuously. Unfortunately low crosslinked, gel resins, such as those suitable for synthesis, tend to pack down under pressure of solvent leading to irregular flow characteristics and high back pressure, and they are therefore unsuited to this approach. Porous inorganic materials, such as silica, can be used for solid phase synthesis by phosphotriester [10] or phosphite [11] methods. However, there is still tendency, in our experience, to generate back pressures especially with the finer mesh varieties, leading to inconsistent flow and other technical problems. It is also difficult to obtain reproducible nucleoside loadings on silica and care must be taken to cap off unfunctionalised silanol groups.

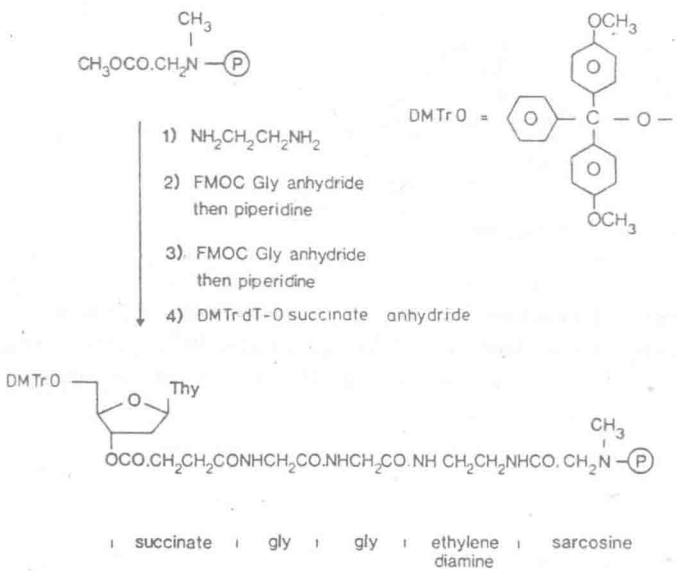


Figure 1 Functionalisation of polydimethylacrylamide/kieselguhr

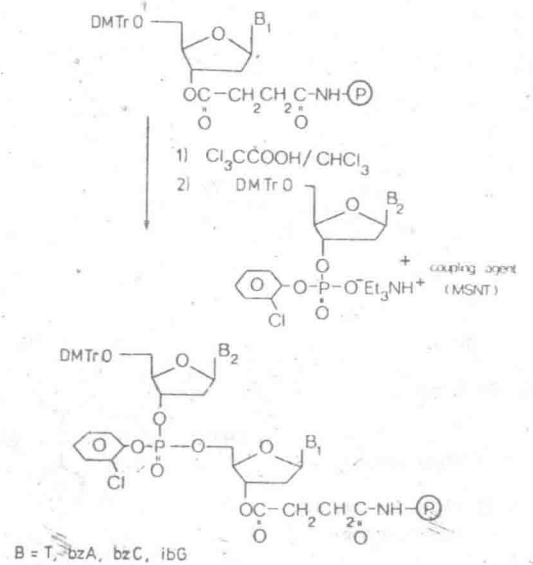
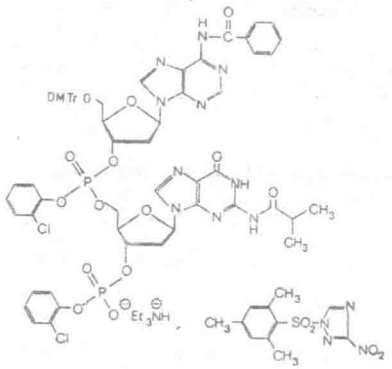


Figure 2 Coupling reaction for the first protected nucleotide yielding a polymer-bound dinucleotide. Mesitylene-2-sulfonyl-3-nitro-1,2,4-triazole (MSNT) is used as condensing agent (coupling reagent)



The coupling agent MSNT

Figure 3
Structure of the fully protected dimer d(ApG) and the coupling reagent mesithylene-2-sulfonyl--3-nitro-1,2,4-triazole

Recently a new composite support has been developed consisting of fabricated beads of inorganic material (e.g. kieselguhr) that contain large pores, several 100 nm in diameter, in which polydimethylacrylamide gel resin has been prepared in situ [12]. The support combines the excellent flow characteristics of the beads, that generate negligible back pressures even at fairly high flow rates, with the proven value of the polyamide gel. Rapid diffusion of solvent takes place in and out of the beads so that no agitation of the material is required during synthesis. The material has been used successfully in flow systems for both peptide [12] and oligonucleotide synthesis [13] and is now commercially available (LKB Biochrom, Ltd., Cambridge, U.K. and Victor Wolf, Ltd, Clayton, Manchester, U.K.).

3.1.2 Functionalisation

The support contains a controlled amount of ester functionalities introduced in the polymerisation reactions used in its preparation as acryloyl sarcosine methyl ester. Batches of support are assayed by amino acid analysis for sarcosine content. This represents the nominal loading value of nucleoside obtainable for the particular batch. So far batches containing ca 0.1 $mmol \cdot g^{-1}$ sarcosine have been used and through a series of precycling reactions (Fig. 1) nucleoside loadings of 85 - 90 $\mu mol \cdot g^{-1}$ have been consistently obtained. The support is treated with ethylene diamine overnight, washed well with dimethylformamide (DMF) and then coupled under standard peptide coupling conditions with a protected glycine derivative (Fmocgly) prepared as its symmetrical anhydride. The fluorenylmethoxycarbonyl (Fmoc) protecting group is removed by treatment of the support with piperidine/DMF and a second protected glycine derivative added as before. The double glycine spacer has been found necessary in order to obtain consistent rates of internucleotidic coupling in DNA assembly.

To attach the first nucleoside the support is once again treated with piperidine/DMF and then coupled with the symmetrical anhydride derivative of one of the four 5'-O-dimethoxytrityl-2'-deoxynucleoside-3'-O-succinates in an analogous way to a peptide coupling reaction.

3.2 Assembly of oligonucleotide chains

Oligonucleotides are synthesised from 3' to 5' ends in a series of assembly cycles involving only two chemical reactions per cycle (Fig. 2). In common with other solid phase methods the extremely acid labile dimethoxytriphenylmethyl (dimethoxytrityl) group is used for protection of the 5' position (the pixyl [14] protecting group is an acceptable alternative). Of all the protic acids which can be used for

deprotection we have found that trichloroacetic acid (TCA) is the best [8]. Oligonucleotide chains containing deoxyadenosine are susceptible to slight depurination (loss of adenine) under acidic conditions, and upon removal of protecting groups at the end of the synthesis the chain fragments at the depurinated site. We have observed, however, that depurination is negligible in most cases (much less than 1% per deprotection step per adenine residue) using brief TCA treatment for the removal of terminal dimethoxytrityl groups. The alternative reagent, zinc bromide, which gives rise to little or no depurination, has been used in conjunction with polystyrene [15] and silica supports [16]. Unfortunately this reaction is prone to inhibition by other chelating agents including amino groups of amide bonds. Increasingly slow reactions have been observed particularly for highly functionalised resins [17] as the oligonucleotide chain is extended and this requires careful monitoring. As might be anticipated zinc bromide deprotects too slowly when used in conjunction with polyamide supports [17,18]. TCA is not subject to such inhibition and gives consistent deprotection within 3 min.

The acidic deprotection step generates free hydroxyl groups on the support, which are coupled in a second step with an appropriately protected monomer or dimer unit (Fig. 3). These contain dimethoxytrityl groups on the 5' position, acyl groups protecting the exocyclic amino groups of adenine, cytosine and guanine, and 3'-O-2-chlorophenyl phosphodiesters. These latter moieties react with hydroxyl groups to give fully protected phosphotriesters. The dehydration (coupling) reaction is carried out in pyridine with the coupling agent mesitylene-2-sulphonyl-3-nitro-1,2,4-triazole (MSNT) [19]. The two reactions of deprotection and coupling are separated by appropriate polymer washing steps. In contrast to other published solid phase methods we find no particular benefit in incorporating a third reaction of treatment of the support with a capping reagent after coupling.

The polymer support is contained in a small glass column connected to a manually operated solvent delivery system operated by slight pressure of an inert gas such as argon or nitrogen. All fittings need only be low pressure of Teflon and glass. An inexpensive kit is commercially available (Omnifit, Ltd., Cambridge, U.K.), but any reliable flow system, high performance liquid chromatography (hplc) or automated synthesiser could potentially be adapted for the purpose. The cycle time is 60-80 min and involves very little manual work.

3.3 Deprotection

After the appropriate number of assembly cycles three stages of deprotection are required but without intermediate purifications.

(1) The support is treated with 1,1,3,3-tetramethylguanidinium-syn-2-nitrobenzaldoximate [20] in aqueous dioxan to remove 2-chlorophenyl protecting groups from internucleotide phosphates and to break the succinyl linkage and liberate oligonucleotides into solution. The use of this oximate reagent gives rise to very little unwanted cleavage of internucleotide bonds, which for many years was a major side reaction resultant from the use of inferior reagents for this step.
(2) The oligonucleotide is treated with concentrated ammonia solution in a sealed tube at 50°C for 5 h to remove acyl protecting groups.
(3) The final step is treatment with acetic acid:water (8:2, v/v) for 30 min to remove terminal dimethoxytrityl groups.

3.4 Purification by high performance liquid chromatography (hplc)

hplc is recommended for most oligonucleotides up to 20 units and when the highest purity is required. Initial purification is best carried out by ion exchange chromatography which separates oligonucleotides principally (but not exclusively) on the basis of their length. Oligonucleotides are polyanions having formal negative charges increasing integrally as a function of chain length. Most likely impurities in solid phase synthesis are either truncated (prematurely terminated) or failure sequences (missing one or more units), or arise from chain cleavage reactions. Such impurities will be shorter in length and hence are less strongly retained on ion exchange columns. This makes identification of the desired peak on chromatograms usually straightforward. For longer chains or those which have some self-complementarity or are rich in guanine, disaggregants (ethanol or formamide) and elevated temperatures are necessary in this chromatography step.

After gel filtration to remove salts a second purification step can be carried out by reversed phase chromatography. Here separation relies on differences in hydrophobicity and particularly good resolution can be obtained of minor impurities with the same or similar length (e.g. base modified chains). Oligonucleotides having gone through both chromatography steps are reproducibly 99% pure or better as determined by analysis of ^{32}P-labelled samples. The second step (reversed phase) can often be omitted for the shorter chains.

3.5 Purification by polyacrylamide gel electrophoresis (page)

This technique can be used in place of ion exchange hplc as a first line purification method. It is particularly useful for long chains which cannot be resolved by hplc or for purification of identical length chains synthesized as a mixture for gene probes. Once again separation