Volume 14 Number 5 1986 Nucleic Acids Research

The MURFI linker for multiple reading frame insertion of a sense or nonsense codon into DNA

Daniel Perlman and Harlyn O.Halvorson

Rosenstiel Center, Brandeis University, Waltham, MA 02254, USA

Received 17 September 1985; Revised and Accepted 5 February 1986

ABSTRACT

Blunt-end palindromic DNA linkers with a central restriction site have been designed for the multiple reading frame insertion (abbreviated MURFI) of a sense or nonsense codon into DNA. We have utilized an amber MURFI linker, $^{5'}$CTAG TCTAGA CTAG$^{3'}$ to disrupt the lacZ gene, yielding truncated β-galactosidase proteins. Conditional disruption of the tet^r gene in E. coli has also been demonstrated. Nonsense codon MURFI linkers permit conditional fusion of multiple gene products while sense codon linkers can add structural elements (e.g. β-turn, cationic segment, hydrophobic segment) or a desired amino acid to a protein (e.g. methionine, cysteine). Shotgun or alternatively site-directed insertion of the symmetric linkers is possible. The over-all length of the linker may be adjusted to retain the original reading frame, matching nucleotide additions or subtractions at recipient DNA sites. If a linker restriction site occurs elsewhere in the target DNA, single linker copies may still be inserted using non-phosphorylated linkers.

INTRODUCTION

DNA "linkers" are oligonucleotides of a defined sequence that typically contain a restriction site and that can be ligated to the termini of blunt-end DNA segments. Usually DNA linkers have a sequence which is self-complementary so that self-annealing two single strands produces blunt-end double-stranded DNA segments. Such "palindrome" double stranded segments have 2-fold symmetry allowing the linker to be inserted in either "direction", i.e. without polarity, into recipient DNA.

Linkers may be synthesized in different lengths to maintain or to disrupt the downstream codon reading-frame of recipient DNA. Furthermore, the length of a linker may be varied to introduce different numbers of amino acids into a structural gene. The location of such insertions can be established by conventional restriction analysis and/or DNA sequencing.

The use of synthetic DNA linkers to achieve mutations in vitro has been previously described for the Tn3 transposon (1). Recently linkers have been used to discern separate functional regions within a single gene (2). The use of a nonpalindromic oligonucleotide sequence lacking a restriction endo-

nuclease site to introduce the TGA stop codon in three reading frames has also been described (3). For its insertion, restriction sites were ligated to the ends of the oligonucleotide to polarize the direction of oligonucleotide insertion into recipient DNA cleaved with the same two restriction enzymes. These linkers were not intended to, nor do they allow blunt-end shotgun or blunt-end targeted insertion of codons into recipient DNA. Directed insertion of a stop or start codon has also been described by Narang and Wu (4). Their oligonucleotide adapter molecules contain a restriction enzyme site and either a stop codon located upstream of the site or a start codon located downstream of the site. The adaptors provide a means of inserting a start or stop signal into DNA in a single reading frame. Recently, single stranded hexameric adaptor oligonucleotides have been described (5). These have been designed to insert two codons into a pre-existing cohesive-end restriction site thereby creating a new restriction site.

The present work describes a method which can be used to insert a sense or nonsense codon into DNA in any reading frame. Palindromic linkers consisting of a centrally located six base recognition restriction site and a sense or nonsense codon repeated in the linker in the three protein translation reading frames have been constructed. The linkers are termed multi-reading frame insertion linkers (abbreviated MURFI linkers).

MATERIALS AND METHODS

Plasmid constructions

Bacterial transformation, plasmid preparation, and restriction mapping procedures were all as described by Maniatis *et al.* (6). The *E. coli* lac deletion host strain LG90 (F^- Δ *lac pro* XIII) has been previoulsy described (7). The plasmid pMR200, a constitutive *lacZ*$^+$ derivative of pMR100, used as a recipient for linker insertion has been described previously (8). Non-phosphorylated *XbaI* amber terminator Murfi linkers (described in Results) were synthesized in the laboratory of Dr. K. Bostian. The linkers were ligated with pMR200 cleaved within the *lacZ* gene at either the *SacI* or the *SspI* site as follows: pMR200 was cleaved at the unique *SacI* site (generating 3' protruding ends, at position 1948 in the *lacZ* sequence) and was nuclease-treated (using T4 DNA polymerase in the presence of dGTP) to produce blunt ends for linker ligation (6). In contrast, producing the *SspI* (blunt-end cleavage) recipient site in *lacZ* required partial *SspI* digestion of pMR200 due to a second *SspI* site found in the plasmid. An appropriate

digestion time with SspI yielded predominantly full length linear (single cut) pMR200 (approximately 1 μg plasmid DNA, 10 units SspI in 50 μl volume incubated for 1 min at 37°C). Samples of SspI and SacI cleaved pMR200 DNA (≈0.5μg) were mixed with ≈0.5μg XbaI Murfi linker DNA and ligated using conventional reaction conditions (6). Excess unligated linker was eliminated by selective precipitation of plasmid DNA (6% PEG 6000, 0.6M NaCl, final concentration). Linear plasmid DNA was converted to the nicked circular form by producing single-stranded (cohesive) linker "tails" and self-annealing. This was accomplished by heating to 65°C in 0.15M NaCl for 10 min and cooling to 4°C for 1 hr. Prior to transformation the "linker at SacI" plasmid DNA preparation was treated with SacI to linearize any non-recombinant parental plasmid DNA. The "linker at SspI" plasmid DNA construction could not be similarly SspI treated due to the second SspI site to be preserved in the vector. Bacterial transformant colonies which were white (lac^-) on MacKonkey-lactose-ampicillin agar were screened for containing the XbaI site contributed by the Murfi linker. Restriction mapping of the XbaI Murfi linker confirmed the SacI and SspI locations in lacZ.

For demonstrating conditional gene disruption, in-frame insertion of the same linker at the SalI site in the tet^r gene in pBR322 was carried out as follows: pBR322 DNA, which had been cleaved with SalI enzyme, was incubated with DNA polymerase I Klenow fragment and the deoxynucleotide triphosphates to obtain "filled-in" blunt ends (4 nucleotide addition) as previously described (6). Following ligation with the 14-mer XbaI linker (see above), and prior to bacterial transformation, the recombinant DNA preparation was treated with SalI enzyme to linearize non-recombinant parental plasmid DNA. The E. coli strain JM83 (ara, Δlac-pro, strA, thi, φ80dlacZ ΔM15) obtained from J. Messing was transformed and ampicillin resistant colonies selected. These colonies were subsequently replica-plated onto 10 and 25 μg/ml tetracycline-containing LB-agar plates. Plasmid DNA was isolated from ten colonies exhibiting tetracycline sensitive phenotypes. Cleavage with XbaI and restriction mapping were utilized to show linker insertion at the former SalI site in the tet^r gene for all ten isolates. The ability to suppress the linker nonsense mutation in tet was tested by transforming an E. coli suppressor strain, DB6660 [($r^-m^-trp^{am}$, lac^{am}, pyrF::Tn5,Su3) obtained from D. Botstein] and the normal strain JM83 (negative control). The tetracycline resistance phenotypes of ampicillin resistant transformant colonies were examined by replica-plating (200-500 colonies per plate) to LB agar containing 5, 10, and 25 μg/ml tetracycline. Growth was scored after

overnight incubation at 37°C. Tetracycline sensitivity of the host cells was tested by replica-plating colonies from simple LB-agar.

Protein analysis

E. coli LG90 transformant cells carrying pMR200 and pMR200 modified by the Murfi linker insertions described above were cultured in LB medium plus 50 μg/ml ampicillin. Both mid-log phase and stationary phase cells were harvested for lacZ protein analysis. The β-galactosidase proteins encoded by pMR200 and pMR200 derivatives containing the XbaI Murfi terminator linker were analyzed in 10% polyacrylamide-SDS gels as described elsewhere (9). Crude cell extracts derived from 0.1ml of stationary phase transformed cells or an equivalent number of log-phase cells were prepared for electrophoresis as previously described (8). After the bromphenol blue tracking dye in the cell extracts had electrophoresed to the end of the gel, the gel was either stained with Coomassie blue or alternatively electroblotted onto nitrocellulose as previously described. β-galactosidase derived proteins were detected by incubating the filter successively with polyclonal anti-β-galactosidase antiserum (kindly provided by Linda Gritz) and 0.5μCi ^{125}I-labeled protein-A (NEN Products) as previously described (8).

RESULTS AND DISCUSSION

Truncation of β-galactosidase with XbaI amber terminator Murfi linkers

Murfi linkers carrying a nonsense codon in three reading frames may be inserted into a gene to disrupt and therby identify the polypeptide encoded by that gene. The use of a nonsense linker for structural gene disruption was demonstrated with the lacZ gene in E. coli. For truncating the encoded β-galactosidase polypeptides, a 14 nucleotide Murfi linker of sequence 5'CTAGTCTAGACTAG3' was designed. The linker which carries a centrally located XbaI site (TCTAGA) was inserted at two locations (SspI site at position 1240 and SacI at 1948) in lacZ as described in Materials and Methods. Subsequently the β-galactosidase proteins synthesized in cells of E. coli strain LG90, carrying the lacZ-pMR200 linker-modified plasmids were analyzed on polyacrylamide gels. Simple Coomassie blue-stained gels displaying total cell protein from log phase cells revealed β-galactosidase protein (117k daltons) encoded by the pMR200 native lacZ gene (Figure 1 lane 1). Shorter truncation products were not apparent at the positions on the gel predicted by the locations of the terminator linkers (75k daltons for SacI protein and 47k daltons for SspI). However when similar but unstained gels were electroblotted and probed with antibody as described in Materials

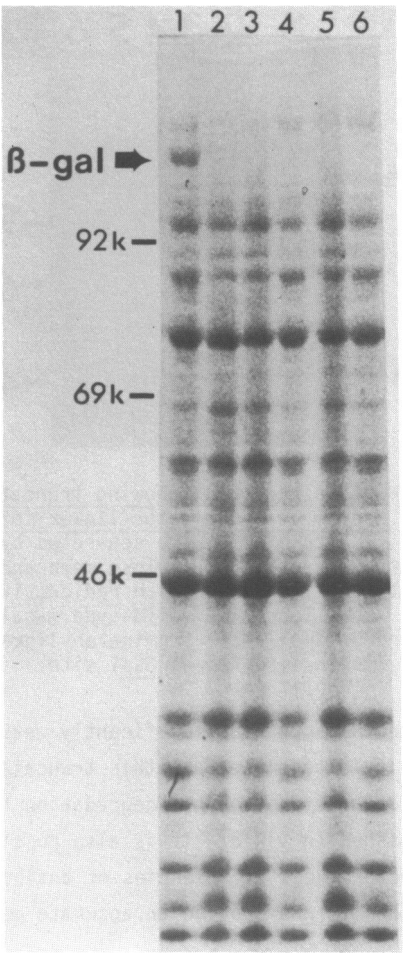


Figure 1. Coomassie blue stained polyacrylamide gel showing separation of E. coli proteins. Crude extracts from log phase E. coli cells carrying pMR200 derivative plasmids were analyzed by gel electrophoresis as described in Materials and Methods. Lane 1, pMR200; lanes 2 and 3, XbaI amber terminator linker inserted at SacI site in lacZ; lanes 4, 5 and 6, XbaI amber terminator linker inserted at SspI site in lacZ. Different lanes contain extracts from independent E. coli transformants. Phosphorylase B (92k), bovine serum albumin (69k), and ovalbumin (46k) were electrophoresed as protein molecular weight standards.

and Methods, the β-galactosidase truncation products resulting from linker insertion were detected at the predicted locations (Figure 2). Similar results were obtained for stationary phase cell extracts (data not shown). The radioactive β-galactosidase band intensities for the SacI linker

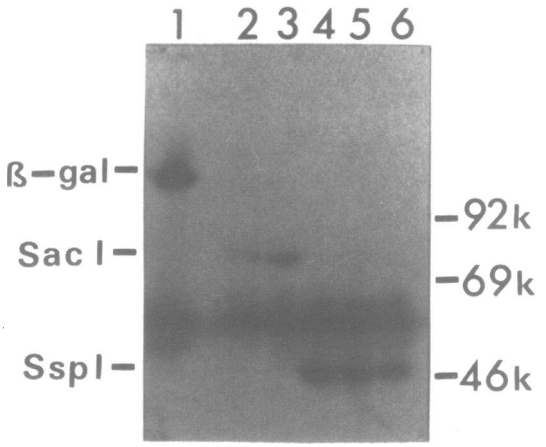


Figure 2. Autoradiogram of western blot showing truncated lacZ polypeptides created by insertion of XbaI amber terminator linker into the lacZ gene of plasmid pMR200. E. coli cell proteins were separated by gel electrophoresis as in Figure 1, transferred to a nitrocellulose membrane, incubated with anti-β-galactosidase antibody and probed with radioactive protein A (see Materials and Methods). Lane 1, pMR200 (wild-type β-galactosidase polypeptide); lanes 2 and 3, XbaI amber terminator linker inserted at SacI site; lanes 4, 5 and 6, linker inserted at SspI site.

construction (lanes 2 and 3) appeared significantly weaker than for the native gene (lane 1). It is possible that this truncated polypeptide was more susceptible to endogenous proteolytic degradation than the SspI fragment and was therefore diminished in yield. It is also possible that the lesser bands intensities reflect reduced efficiencies of antibody binding to this truncated polypeptide and may not provide an accurate measure of the polypeptide yields.

The ability to suppress the linker nonsense mutations at the SacI and SspI sites could not be tested because insertion of the 14-mer linker at each restriction site produced downstream sequences which were out-of-frame (addition of 14 nucleotides at the blunt-end SspI cleavage site and net addition of 10 nucleotides at the exonuclease-treated, four base 3'overhang SacI site).

Suppression of XbaI linker nonsense mutation in the E. coli tet^r gene.

If the same XbaI linker were inserted within a gene so that the DNA sequence downstream from the linker was positioned in-frame with the sequence upstream, protein "read-through" (supression) should be possible in nonsense suppressor host cells. To test this expectation, a pBR322 plasmid derivative was constructed containing the above XbaI amber nonsense linker (14-mer)

Table 1. Suppression of XbaI linker nonsense mutation in tet^r gene of pBR322

	Tetracycline resistant growth		
	5	10	25[a]
JM83 host cells	-	-	-[b]
with pBR322	+	+	+
with linker	-	-	-
DB6660 host cells	-	-	-
with pBR322	+	+	+/-
with linker	+	+	-

[a] Growth of the ampicillin sensitive host cells JM83 and DB6660 and ampicillin resistant plasmid transformants was monitored on LB-agar containing 5, 10, and 25 μg/ml tetracycline as described in Materials and Methods. The transformants carried either pBR322 or pBR322 with an XbaI Murfi linker inserted at the SalI site in the tet^r gene.

[b] Normal overnight growth of all replica-plated colonies (+); diminished growth with only 10-25% of replica-plated colonies showing normal colony size (+/-); absence of growth (-).

inserted into a "filled-in" SalI site near the middle of the tet^r gene (see Materials and Methods). Expression of the in-frame modified tet gene (containing an overall insertion of 18 nucleotides) was monitored in normal (JM83) and in amber nonsense suppressor (DB6660) E. coli host cells. Tetracycline resistance levels were also compared with those expressed by the pBR322 parental plasmid in both strains (Table 1). Insertion of the linker in tet^r abolished tetracycline resistance in the JM83 host cells but produced only a small reduction in resistance relative to pBR322 resistance in the DB6660 suppressor cells. Only at 25 μg/ml tetracycline could a difference be detected between the wild-type resistance level expressed by pBR322 and that expressed by the linker insertion derivative in the DB6660 cells. This slightly reduced resistance may represent incomplete nonsense suppression, a reduction in the activity of the tet^r gene product due to the six amino acid linker insertion, or both. We believe that incomplete nonsense suppression is the probable explanation because Barany recently demonstrated that several different in-frame amino acid coding mutations at the SalI site in tet^r did not reduce tetracycline resistance (5).

Generalized structure and use of Murfi linkers

Murfi linkers containing sense codons in multiple reading frames can be used to incorporate specific amino acids into a protein or polypeptide. A Murfi linker containing the desired codon(s) is inserted at specific, at semi-random (e.g. four-base recognition sites), or at random cleavage points in a DNA sequence. The effects of these alterations on the structure and function of the protein can be examined. On the other hand, Murfi linkers which contain nonsense codons (TAG, TAA or TGA) in multiple reading frame can be used to effect termination of protein translation in various contexts. These linkers can be inserted between two structural genes or structural gene segments to allow conditional fusion of the translation products of the genes or gene segments. Thus, in nonsense-suppressor host cells, the fused gene product is expressed. However, in wild type or temperature sensitive nonsense-suppressor bacterial strains, protein synthesis will be terminated by the linker insert yielding the unfused product of the structural gene situated upstream of the linker. Thus, for example, the pure upstream product can be used for immunization and antibody production without the complication of producing antibody to the downstream gene product (eg. lacZ or other indicator genes) as well. In addition, the metabolism or the cellular localization of the upstream gene product can be characterized without the complication of association with the downstream product. For example, a downstream lacZ protein has been shown to complicate and inhibit the secretion of upstream periplasmic proteins in E. coli (10,11). The linkers thus simplify gene construction in that one gene fusion allows either production of fused proteins or the single upstream protein. Another use of nonsense Murfi linkers is the disruption of translation within a single gene as demonstrated above. Inserted within a gene at preselected or at random points the linkers will produce truncated gene products in wild- type host cells and thereby identify a given gene with its expressed products. Murfi linkers contain single recognition sites for one or more Type II restriction endonucleases. The recognition site provides a means for mapping the location of an inserted linker in a DNA molecule and it can serve as a site for DNA recombination. The recognition site may consist of a four base recognition sequence but inclusion of less common six base sequences facilitates screening of clones for successful insertion of the linker. Typically the recognition site is centrally located within the linker. This permits palindrome linkers which are inserted in tandem to be cleaved with the cognate restriction enzyme and religated so that a single, complete copy

of the linker remains in the recipient DNA.

In all Murfi linkers, the sense or nonsense codon selected for insertion appears at least three times, once in each of the three protein translation reading frames. In most cases, the codon lies upstream of, within and downstream of the centrally located enzyme recognition site. In order to minimize the length of the linkers, restriction enzyme recognition sites which have an appropriate codon within the site can be used. For example, the PstI recognition site contains the TGC triplet which codes for cysteine and consequently may be used in Murfi linker designed for cysteine insertion.

In their most versatile form the double-stranded Murfi linkers have two-strand rotational (commonly denoted palindromic) symmetry. This symmetry ensures that the linker, when inserted into DNA, gives an equivalent structure regardless of the orientation of the linker. The deoxyribonucleotide sequence of most Murfi linkers can be represented by the following oligonucleotide formula:

$$5' \; X \; I \; N \; II \; N' \; III \; Y \; 3'$$

The three sense or nonsense codons are represented by the Roman numerals I, II and III. In a proline linker, for instance, I, II and III are any of the codons CCT, CCC, CCA and CCG. In an amber nonsense linker, codons I, II and III are TAG. The codons are separated by nucleotide(s) N and N' such that each codon appears in a different reading frame. N and N' cn be an equal number of nucleotides, preferably one or two, or alternatively N and N' may differ in number by three nucleotides. For example, if N is one nucleotide, then N' may be four nucleotides or if N is 2 nucleotides N' may be five. However, neither N nor N' can be a number of nucleotides which is a multiple of three. X or Y may be zero, one, or more nucleotides provided that X and Y are not equal in number.

As a double stranded palindromic structure the segment XIN of a linker pairs with YIII N'. Y can include nucleotides which are added to satisfy base pairing requirements of one end of the oligonucleotide with the other. X and Y may vary in length depending upon the desired overall length of the linker. In Murfi linkers containing sense codons, no nucleotide triplet in any reading frame should constitute a stop codon. Some Murfi linkers do not fit the formula given above, such as linkers which have an overlapping (nested) arrangement of codons I, II and III (see subsequent examples).

Murfi linkers may be ligated to blunt-end recipient DNA in phosphorylated or nonphosphorylated form. Nonphosphorylated linkers form a

phosphodiester bond with only one recipient strand of a DNA duplex. The unligated linker strands can be removed by heat denaturation to leave single strand linkers (in trans configuration at the cleavage junction). When these strands are hybridized to each other a single linker insert remains.

The design of additional Murfi linkers

Murfi linkers should prove to be a useful tool in either site-directed or shotgun protein structural alteration. The linkers can be used to insert any designated amino acid or nonsense codon into a polypeptide sequence. The effect of such insertion on the biological functioning of the protein can then be studied. Regions of proteins that are necessary for biological or biochemical activity or physical structure can be functionally altered with these linkers and mapped. Such regions may for example include enzyme catalytic sites, hormone receptor binding sites and other functional domains. The linkers may also be utilized to extend or alternatively disrupt structural domains including α-helix and β-sheet regions. Amino acid codons which are particularly interesting candidates for use in these protein engineering linkers include those encoding amino acids such as cysteine or amino acids which disrupt protein secondary structure such as proline. In addition, methionine codon-bearing Murfi linkers will be valuable in constructing proteins having a displaced amino terminus. Cysteine and methionine linkers may also be added to genes between or beyond their functional domains to enhance ^{35}S-isotope labeling of their protein products. Similar cysteine addition may also facilitate disulfide bridge formation between monomer polypeptides if desired or permit immobilization on thiol group-containing matrix material. Linkers encoding multiple basic or acidic amino acids may also be useful genetic engineering elements. For example the isoelectric point of a protein segment could be altered by adding such a linker.

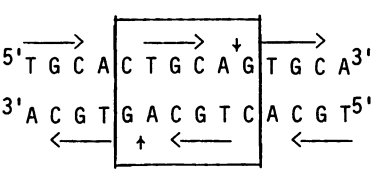
Typically a total of 4-6 amino acids are encoded by each Murfi linker. Consequently 3-5 amino acids besides the designated amino acid are translated into the protein product of the gene. By taking the entire inserted sequence into account, Murfi linkers can be carefully designed to alter protein structure similarly in all three reading frames (see cationic and β-turn linkers below).

For the purpose of shotgun insertion of Murfi linkers into genes, digesting the recipient DNA with blunt-end cutting restriction enzymes, such as AluI, HaeIII or RsaI is convenient. Such cleavage will permit insertion of linkers which are even multiples of 3 nucleotides, e.g. 12 or 18

nucleotides. The inserted segments will maintain the original reading frame of the recipient DNA. Alternatively if the recipient DNA is cleaved to produce single strand protruding ends which are then converted to blunt ends, linkers of different lengths must be used to restore the original downstream reading frame. For example a 14 nucleotide linker will restore the reading frame at a 4 nucleotide overhang site which has been "filled in" to form blunt ends.

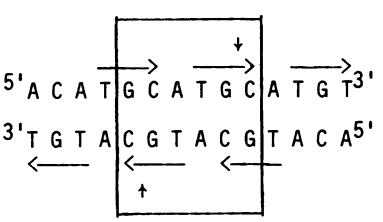
The following examples illustrate various Murfi linker structures which may be useful. The horizontal arrows indicate codons intended for insertion and the boxed sequence represents the recognition site for the designated restriction endonuclease. Vertical arrows indicate the cleavage points of the enzyme. In each case one or two nucleotides may be added to the ends of these linker structures if required to restore the downstream reading frame, e.g. CA added to the 5' end would be matched by TG added to the 3' end of the linker.

Example 1: Cysteine Murfi linker with PstI site.



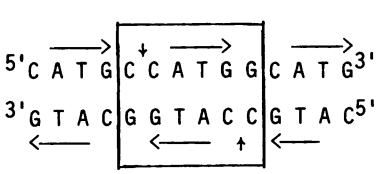
14-mer.

Example 2: Cysteine Murfi linker with SphI site.



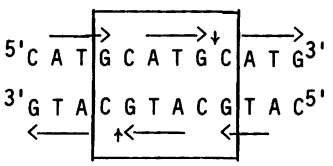
14-mer.

Example 3: Methionine Murfi linker with NcoI site.



14-mer.

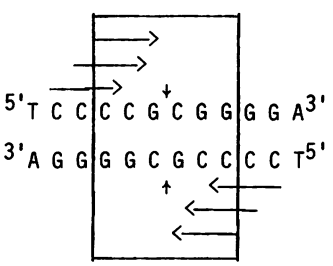
Example 4: Methionine Murfi linker with SphI site.



12-mer.

Note: This linker is similar to the cysteine linker but it is two nucleotides shorter in length.

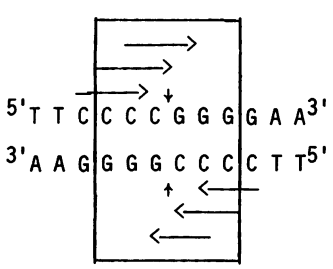
Example 5: Proline Murfi linker with SacII site.



12-mer.

Note: The proline codons in this linker are "nested" unlike the previous linker codons which are in series.

Example 6: Proline Murfi linker with SmaI linker.



12-mer.

Example 7: Beta-turn Murfi linkers.

The usefulness of Murfi linkers in altering protein secondary structure must take into consideration all of the amino acids introduced via the linker. For example, the probability of introducing a disruptive Beta-turn into a protein is increased by combining the proline with other amino acids

enumerated in the Chou and Fasman protein secondary structure tables (12). For instance, sequences such as ser-pro-gly; pro-pro-gly; cys-pro-gly-gly; asn-pro-asn-gly, and others, highly favor the β-turn structure. The Murfi linkers below encode amino acid sequences which include both proline and glycine regardless of the reading frame. The 12-mer in-frame palindromes (for insertion into blunt-end cleaved DNA sequences) are illustrated below. Only one of the two identical DNA strands is shown.

SmaI Beta-turn Murfi linkers

a. 5'T C C C C C G G G G G A 3'

Reading Frame: 1 2 3 1 2 3 1 2 3 1 2 3

Amino acid Sequence

Reading Frame:
1 ser pro gly gly
2 pro pro gly ---
3 pro arg+ gly ---

b. 5'C C C C C C G G G G G G 3'

Amino acid Sequence
same as above except pro substitutes for ser

Example 8. Cationic Murfi linkers

It may be valuable to insert codons for positively charged amino acids within a structural gene for the purpose of altering biochemical properties such as the isoelectric point of the encoded protein. The linkers below are designed for insertion of positively charged amino acids within any reading frame. The 18-mer length is appropriate for conserving reading frame when inserted into blunt-end cleaved DNA.

a. BssHII (+) charged Murfi linker

5'G C G C A C G C G C G C G T G C G C 3' 18-mer

Reading Frame: 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3

Amino Acid Sequence

Reading Frame:						
1	ala	his+	ala	arg+	val	arg+
2	arg+	thr	arg+	ala	cys	ala
3	ala	arg+	ala	arg+	ala	

b. StuI (+) charged Murfi linker

Reading Frame: 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3

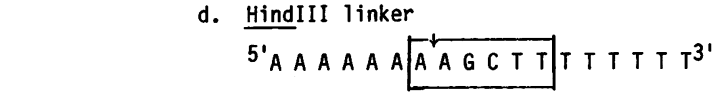
		Amino Acid Sequence				
Reading Frame:	1	lys^+	lys^+	arg^+	pro	phe phe
	2	lys^+	lys^+	gly	leu	phe
	3	lys^+	lys^+	ala	phe	phe

c. ScaI linker

Reading Frame: 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3

		Amino Acid Sequence				
Reading Frame:	1	lys^+	lys^+	ser	thr	phe phe
	2	lys^+	lys^+	val	leu	phe
	3	lys^+	lys^+	tyr	phe	phe

d. HindIII linker

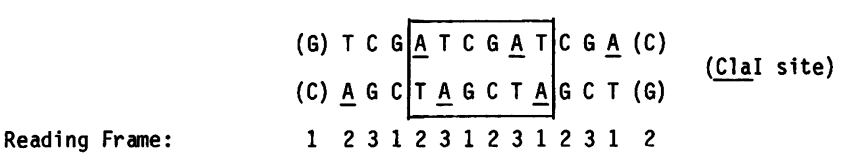


Reading Frame: 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3

		Amino Acid Sequence				
Reading Frame:	1	lys^+	lys^+	lys^+	leu	phe phe
	2	lys^+	lys^+	ser	phe	phe
	3	lys^+	lys^+	ala	phe	phe

Example 9. The tri-tetranucleotide Murfi linker

One class of Murfi linker contains an amino acid codon repeated three times, each time separated by a single and constant nucleotide. For example:



This class of Murfi linkers has the particularly desirable property that with the serial three-fold repetition of a tetranucleotide sequence, more than one amino acid may be encoded in all three reading frames by the single Murfi linker. In the above example three amino acids, asp^-, arg^+ and ser are repeated in all reading frames:

Reading Frame:	1	val	asp$^-$	arg$^+$	ser
	2	ser	ile	asp$^-$	arg$^+$
	3	arg$^+$	ser	ile	asp$^-$

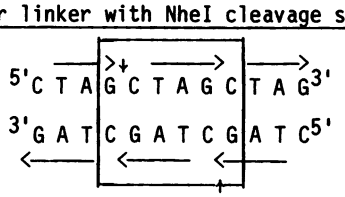
Given the constraints of a palindrome sequence there are 16 hexanucleotide recognition sequences possible for the tri-tetranucleotide linkers:

1.	ATATAT	5.	TAATTA	9.	GCATGC	13.	CGATCG
2.	ATTAAT	6.	TATATA	10.	GCTAGC	14.	CGTACG
3.	ATGCAT	7.	TAGCTA	11.	GCGCGC	15.	CGGCCG
4.	ATCGAT	8.	TACGTA	12.	GCCGGC	16.	CGCGCG

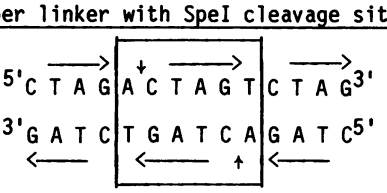
Sequence 3 is recognized by the enzyme NsiI. The Murfi linker encodes met, his$^+$ and ala. Sequences 4 and 13 are recognized by enzymes ClaI and PvuI respectively. The amino acids encoded by the tri-tetranucleotide Murfi linkers are asp$^-$, arg$^+$ and ser (in all three reading frames). Sequence 8 is recognized by the enzyme SnaBI and the Murfi linker encodes tyr, val and arg$^+$. Sequence 9 has already been described (SphI, cysteine and methionine Murfi linkers). The Murfi linkers encode cys, met, his$^+$ and ala in all three reading frames. Sequence 10 is recognized by NheI and contains the amber nonsense codon TAG. Sequence 11 has also been described (BssHII, cationic Murfi linker). Sequences 12 and 15 are recognized by enzymes NaeI and XmaIII respectively. The Murfi linkers encode pro, arg$^+$ and gly in both cases. The remaining sequences above are not recognized by restriction enzymes presently available commercially. (New England Biolabs 1985-86 commercial catalog.)

Example 10: Additional Murfi terminator linkers

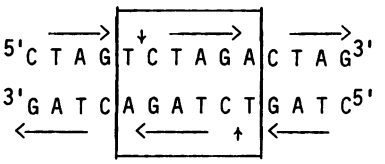
a. Amber linker with NheI cleavage site, 12-mer



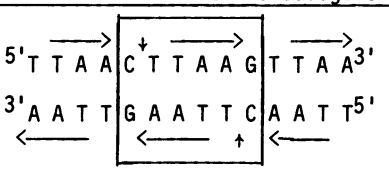
b. Amber linker with SpeI cleavage site, 14-mer



c. Amber linker with XbaI cleavage site, 14-mer

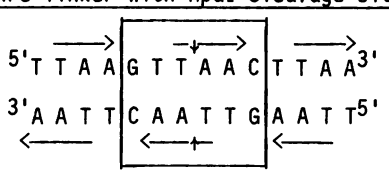


d. Ochre linker with AflII cleavage site, 14-mer



Example 6. A linker having the same sequence as the linker of Example 5 but with appropriate nucleotide additions as shown in Example 4 above produces 16- and 18-mer linkers. With suppression at the first TAA, the linker adds a suppressor amino acid plus leu, lys, and leu. Suppression at the second TAA adds leu, thr, suppressor amino acid and Val. Suppression at the third TAA adds asn, leu, ser, and suppressor amino acid.

e. Ochre linker with HpaI cleavage site



ACKNOWLEDGMENTS

We wish to acknowledge helpful discussions with Dr. H. Baum, the technical assistance of P. Raney, and the assistance of Dr. K. Bostian in synthesizing the first Murfi linker. We also wish to thank New England Biolabs and Pharmacia P-L Biochemicals for providing restriction endonucleases and for initiating production of a variety of Murfi linkers. This work was supported by a grant from the Alfred Jurzykowski Foundation.

REFERENCES

1. Heffron, F., Magdalene, S., and McCarthy, B.J. (1978) Proc. Natl. Acad. Sci. USA. 75, 6012-6016.
2. Stone, J.C., Atkinson, T., Smith, M., and Pawson, T. (1984) Cell 37, 549-558.
3. Pettersson, R.F., Lundstrom, K., Chattopadhyaya, J.B., Josephson, S., Philipson, L., Kaariainen, L. and Palva, I. (1983) Gene 24, 15-27.

4. Narang, S.A., and Wu, R.J. (1981) European Patent Application, No. 81301304.2.
5. Barany, F. (1985) Proc. Natl. Acad. Sci. USA. 82, 4202-4206.
6. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
7. Guarante, L. and Ptashne, M. (1981) Proc. Natl. Acad. Sci. USA. 78, 2199-2203.
8. Gray, M., Colot, H.V., Guarante, L., and Rosbash, M. (1982) Proc. Natl. Acad. Sci. USA. 79, 6598-6602.
9. Perlman, D. and Halvorson, H.O. (1981) Cell 25, 525-536.
10. Bassford, P.J., Silhavy, T.J. and Beckwith, J.R. (1979) J. Bacteriol. 139, 19-31.
11. Moreno, F., Fowler, A.V., Hall, M., Silhavy, T.J., Zabin, I. and Schwartz, M. (1980) Nature 286, 356-359.
12. Chou, P.Y., and Fasman, G.D. (1978) Adv. Enzymol. 47, 45-148.