Development of a Novel High Throughput Directional Cloning Platform

Ron Hernan and Leanne Snyder

Sigma-Aldrich Biotechnology, St. Louis, MO

Abstract

The cloning and expression of a large number of genes has been an enormous challenge for gene function studies in the post-genomics era. One of the major limitations has been the lack of a pair of universal restriction enzymes for generating the cohesive ends required for directional cloning. Traditionally, different enzymes are used for directional cloning of each gene, which restricts throughput and increases workload. To overcome this problem, thionucleotides dATPαS and dGTPαS are incorporated during PCR amplification with restriction enzyme site sequences engineered into the primers. Digestion of the amplified product with Exonuclease III produces 5' ends that are restriction enzyme site compatible and capable of ligation into vectors digested with corresponding restriction enzymes. With this approach, cloning efficiencies are greater than 90 % (transformants with positive inserts in the correct orientation). Because this method of cloning is universal, it is highly adaptable to robotic applications. Using this technology we have developed a system to both systematically and directionally clone open reading frames in a universal fashion without consideration of internal restriction enzyme sites. We have cloned 384 unique *E. coli* ORFs using this technology with a 74 % success rate. In addition, we applied this technology to successfully clone selected ORFs from mammalian systems into expression vectors with comparable success.

Introduction

Functional analysis of open reading frames (ORF) requires the amplification and directional cloning of target sequences into an expression vector. The use of two restriction enzymes traditionally has been used to facilitate directional cloning with restriction enzymes recognizing a six base pair site being preferred. When cloning multiple ORFs the choice of restriction enzymes is limited to the sequence of a particular ORF. Probability of a six base pair site in a random sequence is 1:4096 while the probability of either restriction enzyme site being present is 1:2048. The need for a directional cloning platform that is sequence independent and is amenable to automation is apparent considering the enormous number of gene products that are yet uncharacterized.

This method uses Exonuclease III, a double-stranded specific exonuclease that catalyzes the stepwise removal of nucleotides from the 3' hydroxyl termini of double-stranded DNA, to produce cohesive ends. To control the extent of nucleotide removal we have incorporated thiopurines dATPαS and dGTPαS into double-stranded DNA during amplification. Because Exonuclease III cleaves phosphorothioates extremely slow, 5' termini are produced which are compatible with cohesive ends generated with common restriction enzymes in digested vectors.

Results

Thio-Purine Effect on PCR

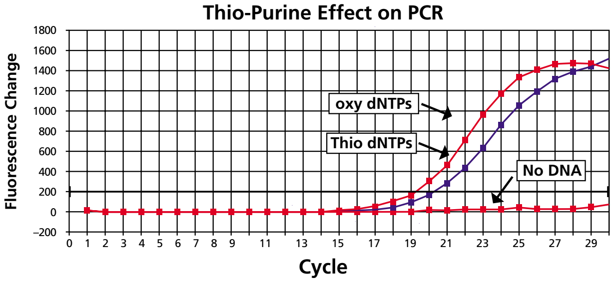

Figure 1. Quantitative PCR amplification of the *E. coli pho A* gene using thionucleotides. PCR was carried out on an ABI 7700 thermocycler using sequence detection software 1.7 and the fluorescent dye SYBER green. The addition of thionucleotides has minimal effect on PCR. C_T's for a normal nucleotide mix gave a value of 19 and with a thionucleotide mix a C_T of 20 was observed. We are able to achieve maximum incorporation with minimal effect on PCR. Under these conditions a mean 5' overhang length of 20 nucleotides was calculated assuming a completely random sequence.

Overview of Automated Cloning Platform

Overview of Automated Cloning

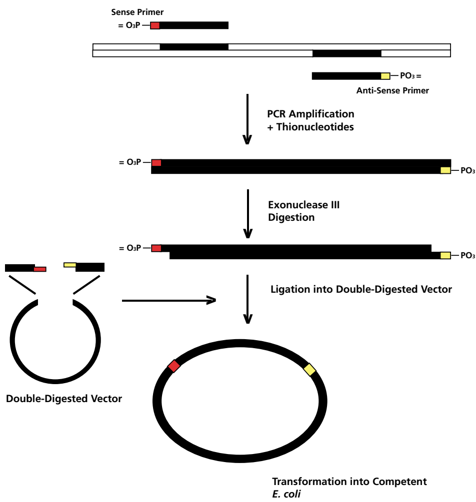

Figure 2. Thionucleotides are incorporated during amplification. Primers are designed to contain a restriction enzyme, allowing for controlled digestion with Exonuclease III to produce a 5' overhang. Amplified products are ligated into an appropriate double digested vector and then transformed into *E. coli.*

PCR Amplification of Bacterial and Mammalian ORFs

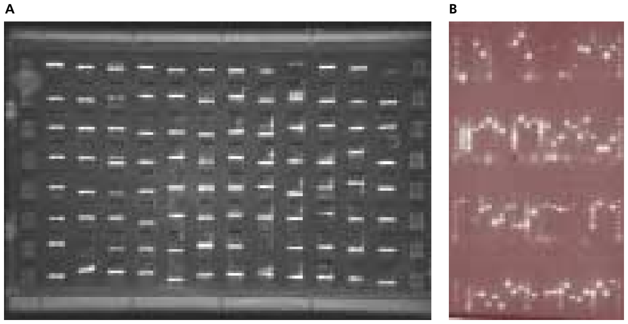

Figure 3. Amplification of A) *E. coli* and B) Mammalian ORFs. *E. coli* ORFs were amplified from strain W_{3110} using genomic DNA as template. Mammalian ORFs were amplified from prepared HeLa cDNA as template. All forward primers had the sequence AGCTT added to the 5' end to allow cloning into a *Hin*d III site while the reverse primer had a GATCT to allow cloning into a *Bgl* II site. Amplifications were performed in a 96-well format under identical conditions.

Expression Vectors

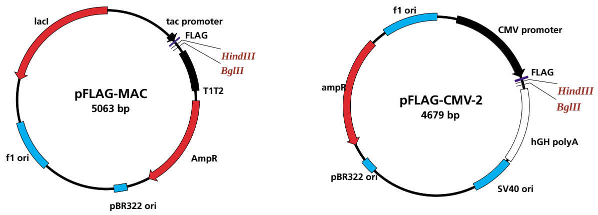

Figure 4. pFLAG-MAC was double digested with *Hin*d III and *Bgl* II and used for cloning of *E. coli* ORFs. pFLAG-CMV2 was double digested with *Hin*d III and *Bgl* II and used for cloning of mammalian ORFs.

E. coli ORFs

abc	yaeD	yafB	yafC	yafD	yafE	dniR	b0212	b0213	rnhA	dnaQ	b0217	b0218	b0219	b0220	b0221	gmhA	yafJ	yafK	yafQ	dinJ	yafL	yafM	fhiA
mbhA	dinP	yafN	yafO	yafP	b0235	prfH	pepD	gpt	yafA	crl	phoE	proB	proA	b0245	yafW	b0247	yafX	b0249	ykfB	yafY	yafZ	ykfA	perR
b0255	tra8_1	b0257	ykfC	yi52_1	ykfD	yagD	yagC	b0263	insB_2	insA_2	yagB	yagA	yagE	yagF	yagG	b0271	yagI	argF	insB_3	insA_3	yagJ	yagK	yagL
b0279	b0280	b0281	yagP	yagQ	yagR	yagS	yagT	yagU	b0288	yagV	yagW	yagX	yagY	yagZ	b0294	b0295	b0296	eaeH	b0298	b0299	ykgA	ykgB	b0302
b0303	ykgC	ykgD	ykgE	ykgF	ykgG	b0309	ykgH	betA	betB	betI	betT	yahA	b0316	b0317	b0318	b0319	b0320	b0321	b0322	b0323	b0324	b0325	b0326
b0327	b0328	b0329	b0330	b0331	b0332	b0333	b0334	b0335	codB	codA	cynR	cynT	cynS	cynX	lacA	lacY	lacZ	lacI	b0346	b0347	mhpB	b0349	b0350
mhpE	b0352	b0353	b0354	b0355	adhC	b0357	b0358	b0359	yi21_1	yi22_1	b0362	b0363	b0364	b0365	b0366	b0367	b0368	hemB	b0370	b0371	tra5_1	b0373	b0374
b0375	yaiH	sbmA	b0378	b0379	b0380	ddlA	yaiB	phoA	psiF	yaiC	proC	yaiI	aroL	yaiA	aroM	yaiE	b0392	yaiD	yajF	b0395	araJ	sbcC	sbcD
phoB	phoR	brnQ	b0402	malZ	yajB	queA	tgt	yajC	secD	secF	yajD	tsx	yajI	ybaD	ribD	ribH	nusB	b0417	b0418	b0419	b0420	ispA	xseB
b0423	thiJ	apbA	b0426	b0427	cyoE	cyoD	cyoC	cyoB	cyoA	ampG	yajG	bolA	tig	clpP	clpX	lon	hupB	b0441	b0442	b0443	b0444	ybaE	cof
b0447	mdlA	mdlB	glnK	amtB	tesB	b0453	b0454	ybaA	b0457	b0458	b0459	hha	ybaJ	acrB	acrA	acrR	b0465	ybaM	priC	ybaN	apt	dnaX	ybaB
recR	htpG	adk	hemH	ybaC	gsk	ybaL	fsr	ushA	b0481	b0482	b0483	b0484	b0485	b0486	b0487	b0488	b0489	b0490	b0491	b0492	b0493	tesA	ybbA
b0496	rhsD	ybbC	b0499	ybbD	b0501	b0502	ybbB	b0504	b0505	b0506	gcl	gip	b0509	b0510	b0511	b0512	b0513	b0514	b0515	b0516	b0517	b0518	b0519
b0520	ybcF	purK	purE	ybbF	ppiB	cysS	ybcI	ybcJ	folD	b0530	b0531	b0532	b0533	ybcG	fimZ	intD	b0538	b0539	b0540	tra5_2	b0542	b0543	b0544
b0545	b0546	b0547	b0548	b0549	b0550	b0551	yi52_2	nmpC	b0554	b0555	b0556	b0557	b0558	b0559	nohB	b0561	b0562	b0563	appY	ompT	envY	ybcH	nfrA
nfrB	b0570	b0571	b0572	b0573	b0574	ybdE	pheP	ybdG	nfnB	ybdF	b0580	b0581	yi81_2	entD	fepA	fes	entF	fepE	fepC	fepG	fepD	ybdA	fepB

Figure 5. A register of 384 *E. coli* ORFs, which were amplified and cloned using thionucleotide incorporation. Primers were purchased from Sigma Genosys (Woodlands TX). Sense primers were designed to be cloned into a *Hin*d III site, while antisense primers were designed to be inserted into a *Bgl* II site of the expression vector pFLAG-MAC double digested with *Hin*d III and *Bgl* II. Following this approach, we are able to automate the PCR, exonuclease digestion and ligation for cloning into *E. coli* expression vectors using a BIOMEK 2000 liquid handling robot (Beckman Coulter Instruments) to increase throughput.

Mammalian ORFs

tumor necrosis factor type 2 receptor associated protein (TRAP3)	HuCds1 kinase	ZIP kinase
CDK tyrosine 15-kinase (WEE1Hu)	PLTAIRE-1 for serine/threonine protein kinase	KKIALRE for serine/threonine protein kinase
44kDa protein kinase related to rat ERK1	41kDa protein kinase related to rat ERK2	40kDa protein kinase related to rat ERK2
protein kinase PITSLRE alpha 1	NIMA-like protein kinase 1 (NLK1)	protein phosphatase 1, regulatory (inhibitor) subunit
protein inhibitor of activated STAT-1 (PIAS1)	goodpasture antigen binding protein (COL4A3BP)	inhibitor of p53-induced apoptosis-alpha (IPIA-apha)
CDC-2 related protein kinase (PISSLRE)	ATF-alpha transcription factor	nuclear factor I/C (NFIC)
pre-B cell leukemia transcription factor 2 (PBX2)	pre-B cell leukemia transcription factor 1 (PBX1)	general transcription factor IIH, polypeptide 3
sequence-specific DNA binding protein (AP-2)	transcription factor IID	transcription factor, E4TF1-47
activator of S phase kinase	ribosomal protein kinase, S6	protein kinase inhibitor, p58
serine/threonine protein kinase PRP4h	adenosine kinase	63kDa protein kinase related to rat ERK3
protein kinase (JNK2)	protein kinase HSTPK13	phosphorylase kinase (PSK-C3)
PRK-1 associated protein AWP1	TNF receptor-associated factor 2 (TRAF-2)	Ser/Thr protein phosphatase type 2C beta 2 isoform
p35, cyclin-like CAK1-associated protein	cyclin A/CDK2-associated p45 (Skp2)	cyclin A/CDK2-associated p19 (Skp1)
transcription factor (NRF)	general transcription factor IIA	activating transcription factor 7 (ATF7)
general transcription factor IIB (GTF2B)	heat shock transcription factor 1	HCF-binding transcription factor Zhangfei (ZF)
BTF2p44 for transcription factor, 44kDa subunit	basic transcription factor 62kDa subunit (BTF2)	CCAAT-binding protein (HAP2)
Ste20-like kinase (MST2)	IkB kinase, gamma subunit	Integrin-Linked kinase (ILK)
PLSTIRE for serine/threonine protein kinase	PSSALRE for serine/threonine protein kinase	PCTAIRE-2 for serine/threonine protein kinase
clone PSK-J3 cyclin-dependent protein kinase	cdc-2 related protein kinase	thymidylate kinase (CDC8)
cyclin-dependent kinase 2 interacting protein (CINP)	protein phosphatase 2, regulatory subunit A (PR65) alpha isoform	cyclin-dependent kinase inhibitor 3 (CDKN3)
Snk interacting protein 2-28	transcriptional enhancer factor (TEF1)	MHC class I mic-B antigen
POU domain, class 4, transcription factor 1 (POU4F1)	pre-B cell leukemia transcription factor 3 (PBX3)	interferon-stimulated transcription factor 3, gamma
c-myc transcription factor (puf)	Ku protein subunit	ionizing radiation resistance conferring protein
BCL2/adenovirus E1B 19kD-interacting protein 1 (BNIP1)	replication factor C, 37kDa subunit	replication factor C, 40kDa subunit
cdc25c associated protein kinase (C-TAK1)	cyclin-dependent kinase interactor 1 (CDI1)	protein kinase PAK1 (Alpha-PAK)
cAMP-dependent protein kinase, catalytic subunit	phosvitin/casein kinase type II beta subunit	CDK-activating kinase
serine kinase	protein kinase PITSLRE beta 1	protein kinase PITSLRE alpha 2-4
CDC28 protein kinase 2 (CKS2)	CDC28 protein kinase 1 (CKS1)	ras homolog gene family member G (rho G)
protein kinase C inhibitor-I	growth arrest specific protein (gas)	human cyclin H
general transcription factor IIH polypeptide 1	E2F transcription factor 2 (E2F2)	transcription factor AP-2 alpha (TFAP2A)
ribonuclease/angiogenin inhibitor	IFN-responsive transcription factor subunit	transcription elongation factor (SII)
Ezrin	replication factor C, 36kDa subunit	HEB helix-loop-helix protein

Figure 6. A list of ORFs amplifed and cloned from HeLa cDNA. Primers were purchased from Sigma Genosys (Woodlands TX). Sense primers were all designed to be inserted into a *Hin*d III site and the antisense primers were designed to be cloned into the *Bgl* II site of pFLAG-CMV-2. Cloning was carried out in an automated fashion using a BIOMEK 2000 liquid handling robot.

Automated Cloning Performance

	384 *E. coli* ORFs	96 Mammalian ORFs
PCR amplified	97 %	88 %
Ligated and Transformed	92 %	84 %
Sequence verified	74 %	70 %

Table 1. 384 individual *E. coli* ORFs and 96 unique mammalian ORFs were attempted for amplification and subsequent cloning. Percentages were determined from the original target number. Sequenced verified clones were recovered with the same success rate for both bacterial and mammalian sources. Size of targets ranged between 125 to 2900 base pairs with GC composition ranging between 30 % to 65 %.

Conclusions

- Multiple ORFs from both bacterial and mammalian sources were directionally inserted into plasmid vectors by incorporating thiopurines into amplified product followed by digestion with Exo III.
- Thionucleotide incorporation allows for a universal directional cloning platform without the addition of extra sequences.
- The system allows one to design primers for amplification without consideration of internal sequence composition.
- One may systematically clone unique ORFs without having to search for unique restriction enzyme sites that are compatible within multiple ORFs or having to potentially perform sequential digests.
- Construction of clones with this platform greatly increases throughput via simplified primer design for hundreds of ORFs allowing one person to produce 400-800 expression ready clones per week.
- This platform is commercially available (Director Universal Cloning System, Sigma Catalog Number RDC-$_1$ RDC-LIG$_1$, and RDC-L$_3$).

Acknowledgements

We would like to acknowledge Keming Song, Chaun Li, Kristen Opper and Stephanie Uder for their help in this work.

References

1. Putney, S.D., Benkovic, S.J., Schimmel, P.R., A DNA fragment with an α-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by Exonuclease III and can be replicated *in vivo. Proc Natl Acad Sci U S A* **78,** 7350-4 (1981).
2. Hsiao K., Exonuclease III induced ligase-free directional subcloning of PCR products. *Nucleic Acids Res.* **21,** 5528-9(1993).

Corresponding Author

Ron Hernan
Life Science and Technology Center
Sigma-Aldrich Biotechnology
PO BOX 14508
St. Louis, MO 63103
E-Mail: rhernan@sial.com
Phone: (314) 771-5765
Fax: (314) 286-7645