International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 4 Issue 6, June 2017

ISSN: 2394-4404

Primer Design From Conserved Sequences Of The Artemisia Annua Amorpha-4,11-Diene Synthase And Artemisia Annua Amorpha-4,11-Diene Mono Oxygenase Synthase Gene

Ocholi Jonathan Adamu

Department of Pharmaceutical Microbiology and Biotechnology, Faculty of Pharmaceutical Sciences, Gombe State University, Gombe Nigeria

Adrian Slater

School of Allied Health Science, Faculty of Health and Life Sciences, De Montfort University, The Gateway, Leicester, LE1 9BH

Abstract: ***Understanding the artemisinin biosynthetic pathway made up of genes important for artemisinin production is vital in successfully finding a breakthrough in improved commercial yields of artemisinin. Conducting polymerase chain reactions (PCR) on isolated artemisinin biosynthetic genes gives quantifiable amounts of genes that can be transformed into various expression systems for alternative production platforms for artemisinin. The study was aimed at developing appropriate primers for Artemisia annua amorpha 4, 11 diene synthase (ADS) and Artemisia annua amorpha- 4, 11- diene mono oxygenase (CYP71AV1) from detected conserved sequences from their mRNA complete coding sequences. The primers developed will lead to amplifying regions of the gene that are important for translation into active proteins.***

The study reveal that suitable primers satisfying the guidelines for primer design can be designed from conserved sequences of genes for both ADS and CYP71AV1. The primers developed were used to obtain amplified cDNA of the genes of interest. The method used was adequate in designing primers that will ensure conserved sequences are amplified during PCR. This method would ensure amplicons obtained contain the necessary sequence areas for ideal gene activity.

Keywords: Primers, Artemisinin, Reverse transcriptase Polymerase Chain Reaction, Conserved sequences, Artemisia annua

I. INTRODUCTION

Artemisinin is a sesquiterpene lactone peroxide isolated from the aerial parts of *Artemisia annua* plants (Schramek et al., 2010; Weathers, Elkholy & Wobbe, 2006). It contains an endoperoxide bridge and a tri-oxane ring which are responsible for its therapeutic activity (Arsenault, Wobbe & Weathers, 2008). *Artemisia annua* is currently the only direct source of Artemisinin. The production of artemisinin from this source is insufficient to cope with the global demand for the drug which has increased over the years as a result of the prevalence of malaria in south East Asia and sub Saharan Africa (Rathore, McCutchan, Sullivan & Kumar, 2005). The low yield and short fall in supply of artemisinin has led to Scientists looking at various means of increasing the yield of artemisinin in *Artemisia annua* as well as alternative means of production of artemisinin. Strategies implemented to increase the production of artemisinin include; conventional strategies such as improved agronomy and processing, non transgenic efforts to improve *Artemisia annua*, transgenic *Artemisia annua* and genetic manipulation of heterologous organisms.

Understanding the biosynthetic pathway of artemisinin obtained from *Artemisia annua* is vital in obtaining a breakthrough in increased yield of artemisinin from its source. Certain genes involved in artemisinin production have been identified from the biosynthetic pathway of artemisinin (Weathers et al., 2006). The enzyme amorphadiene synthase (ADS) a sesquiterpene synthase which cyclizes farnesyl diphosphate obtained from the various reactions of two units of isopentyl pyrophosphate (IPP) with one unit of dimethyl

allyl pyrophosphate (DMAP) forming amorpha- 4, 11- diene is the first step in artemisinin production (Schramek et al., 2010; Weathers et al., 2006). The amorpha- 4,11- diene mono oxygenase (CYP71AV1) enzyme appears to catalyze the next three steps in the pathway for artemisinin biosynthesis i.e. steps from amorpha- 4, 11- diene to artemisinic alcohol, then artemisinic aldehyde and then to artemisinic acid (Teoh *et al*., 2006; Weathers et al., 2006). It has been suggested that the final step of conversion of artemisinic acid to artemisinin may involve non enzymatic spontaneous photo-oxidative or oxygen dependent reactions (Arsenault et al., 2008; Teoh et al., 2006).

The Polymerase chain reaction (PCR) is used for the *in vitro* amplification of DNA at the logarithmic scale (Singh & Kumar, 2001). For PCR to be specific, the oligonucleotide primers are the most important component of the PCR mix. The efficacy and sensitivity of PCR largely depends on primer efficiency (He, Marjamaki , Soini, Mertsola & Vijanen, 1994). The primer sequence determines several things such as the length of the product, its melting temperature and ultimately the yield (Abd-Elsalam, 2003). There are certain criteria primers unique for target DNA sequences to be amplified should fulfil, they include; primer length, GC content, annealing and melting temperature, 5' end stability, 3' end specificity etc (Dieffenbach, Lowe & Dveksler, 1993).

The primer length is a critical and hard core factor since both specificity and the temperature and time of annealing are partly due to it (Wu, Ugozzoli, Pal, Qian, & Wallace, 1991). Primers to be used should have a minimum length that will ensure a denaturation temperature between 55 and 56^{o}c. this will greatly enhance its specificity and efficiency (Singh and Kumar, 2001). In general, a primer length of between 18 and 30 nucleotides are the best for PCR amplification. Primers should be at least 18 nucleotides in length to minimize the chances of encountering problems with a secondary hybridization site on the vector or insert (Abd-Elsalam, 2003).

The GC content, melting temperature and annealing temperature are strictly dependent on one another (Rychlik, Spencer, & Rhoads, 1990). A GC content of 50-60% is recommended. The optimal melting temperatures for primers in the range of 52-58 ^{o} c generally produce better results than primers with lower melting temperatures (Abd-Elsalam, 2003). The GC content is an important characteristic of DNA and provides information about the strength of annealing (Abd-Elsalam, 2003). Primers should have a GC content of between 45 and 60 % (Dieffenbach, Lowe & Dveksler 1995).

The primer design was done using the program Primer3plus. This program ensures that primers that are complementary to the template DNA are obtained. It ensures that the simple guidelines of primer design mentioned above are being followed. It also ensures that equal Melting temperature of primer pairs and optimal melting temperatures between 55^{0} c and 60^{0} c are obtained. (Newton & Graham, 1997).

II. MATERIALS AND METHODS

A. MATERIALS

RNeasy mini kit (cat no. 74104) size: 25 reactions from QIAGEN

Superscript One-Step RT-PCR with Platinum *Taq* (cat.nos. 10928-034) size: 25 reactions from Invitrogen

Liquid nitrogen

Rnase- Free Dnase set (cat. Nos. 79254) from QIAGEN

Tissue lyser machine from QIAGEN

Bench top Centrifuge machine

PCR machine (Peltier thermal cycle) from MJ Research

Electrophoresis machine

Vortex machines

Microwave MT96T from Matsui

Artemisia annua plant (accession number P7382)

B. BIOINFORMATIC SEARCH FOR MRNA COMPLETE CODING SEQUENCES OF THE GENES

The mRNA, complete coding sequence of the two genes of interest involved in the artemisinin biosynthetic pathway was accessed from GenBank at the NCBI database.

- ✓ *Artemisia annua* amorpha- 4,11-diene synthase (ADS) mRNA
- ✓ *Artemisia annua* amorpha- 4,11-diene –mono oxygenase (CYP71AV1) mRNA.

C. NUCLEOTIDE BLAST FOR SIMILAR SEQUENCES (ORTHOLOGUES AND PARALOGUES) FOR THE TWO GENES OF INTEREST

A restricted BLAST for similar sequences (paralogues) within the *Artemisia annua* plant was done by placing *Artemisia annua* into the Entrez Query window and using the following settings.

Data base used: nucleotide collection (nr/nt)

Optimization was done for highly similar sequences.

All other default settings were used.

Similarly, another BLAST for similar sequences (orthologues) to the artemisinin biosynthetic genes from other species was done by removing the restrictions to *Artemisia annua* and using the same settings as above.

D. OBTAINING PROTEIN FAMILIES FROM GENES OF INTEREST AND SIMILAR SEQUENCES FROM BLAST RESULT

The sequences that gave similar sequences to the two genes of interest from the BLAST were translated to their protein sequences and a search for protein families was done using the Pfam database.

E. MULTIPLE SEQUENCE ALIGNMENT OF GENES OF INTEREST AND SIMILAR SEQUENCES USING CLUSTAL W

The CLUSTAL W program was used to perform a multiple sequence alignment of the two genes of interest with their corresponding similar sequences to obtain regions of conserved sequences.

F. DESIGN OF PRIMERS USING PRIMER3PLUS

Primers were designed from the nucleotides within the conserved sequences obtained from the multiple sequence alignments for the Artemisinin biosynthetic genes. Regions of conserved sequences were included using the include region button.

G. REVERSE TRANSCRIPTASE POLYMERASE CHAIN REACTION (RT-PCR) WITH DESIGNED PRIMERS

The Artemisia annua plant was collected and identified at the University of Leicester herbarium by Mr. Mike Smith and deposited at the herbarium under accession number P7382.

RNA extracts from leaves of the *Artemisia annua* plant were obtained and subjected to a Superscript one-step RT-PCR reaction with Platinum *Taq* mix. The primers developed for the ADS and CYP71AV1 genes were used to amplify the extracted RNA to obtain cDNA amplicons. Controls were set up which include a known RNA sample as a positive control, a known DNA sample with *taq* polymerase as well as RNA samples extracted from *Artemisia annua* with *Taq* polymerase as negative controls. All these controls were necessary to ensure that RNA was amplified and not DNA. The following program was followed for cDNA synthesis and then amplification of the cDNA obtained;

cDNA synthesis and pre-denaturaton: 1 cycle of 55^0c for 30 min and then 94^0c for 2 min.

PCR amplification: Denaturation at 94^0c for 30 sec, annealing at 60^0c for 30 sec and extension at 72^0 c for 1 min all for 40 cycles.

Final extension: 1 cycle of 72^0c for 10 min.

H. GEL ELECTROPHORESIS OF RT-PCR PRODUCTS

The amplicons obtained from the RT-PCR were subject to an agarose gel electrophoresis run alongside a 1 kilobase pair molecular weight ladder. The following settings were used;

A 1% gel prepared in Tris borate EDTA (TBE)
0.5µg/ml SYBR green
5µl loading dye
Voltage of 90V for 45 minutes

I. SEQUENCING OF CDNA FOR THE TWO (2) GENES OF INTEREST

The amplicons obtained from the reverse transcriptase polymerase chain reaction (RT-PCR) of the extracts from *Artemisia annua* with the designed primers were sequenced.

J. BLAST OF SEQUENCED cDNA

The sequences obtained were subjected to a nucleotide basic local alignment search (BLAST) against the National Centre for Biotechnology Information (NCBI) database.

III. RESULTS AND DISCUSSIONS

Gene of interest	Length of linear sequence (bp)	Accession number
ADS	1643	EF197888.1
CYP71AV1	1488	EU684540.1

Table 1: mRNA, Complete coding Sequence of Genes of Interest

The GenBank at the National Centre for Biotechnology Information (NCBI) database allows for access to biomedical and genomic information which have been gathered and stored over the years by submission made by research scientists over various fields of genetic research. This database along with the DNA database of Japan (DDBJ) and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database at the European Bioinformatics Institute (EBI) are publicly accessible and are available to extract any required information pertaining biological data (Pevsner, 2009). The two genes of interest were accessed from GenBank at the NCBI database and their mRNA, complete coding sequence were obtained as seen in table 1.Obtaining complete coding sequences and not partial sequences for the two genes of interest was vital.

Gene of interest	Similar sequences	Accession number	E value	Score	% identity
ADS	*Tanacetum parthenium E*-beta carophylene synthase (Cars) mRNA, complete cds	JF819849.1	1e – 110	411(222)	73%
	Artemisia absinthium sesquiterpene synthase gene, partial cds	AY860851.1	3e – 91	346 bits (187)	94%
CYP71AV1	*Lactuca sativa* germacrene A oxidase (GAO1) mRNA, complete cds	GU198171.1	0.0	1286 bits (696)	83%
	Saussurea costus germacrene A oxidase mRNA, complete cds	GU256645.1	0.0	1273 bits (689)	83%

Table 2: Nucleotide BLAST for similar sequences for the two genes of interest

Results from the BLAST reveal no similar sequences obtained when the search was limited to *Artemisia annua* for both the ADS and CYP71AV1 gene. Two similar sequences were obtained from species different from *Artemisia annua* with ADS and CYP71AV1 respectively as seen in table 2. The parameters (E value, score and % identity) obtained from the BLAST result, appear to show that the similar sequences obtained from other species for both ADS and CYP71AV1 gene are significantly related to our genes of interest.

Gene of interest	Similar sequences	Accession number	Protein family shared	
ADS	*Tanacetum parthenium E*-beta carophylene synthase (Cars) mRNA,	JF819849.1	terpene synthase N Terminal domain family	terpene Synthase Family, Metal binding domain

	complete cds		
	Artemisia absinthium sesquiterpene synthase gene, partial cds	AY860851.1	terpene Synthase Family, Metal binding domain
CYP71AV1	*Lactuca sativa* germacrene A oxidase (GAO1) mRNA, complete cds	GU198171.1	Cytochrome P450 family
	Saussurea costus germacrene A oxidase mRNA, complete cds	GU256645.1	Cytochrome P450 family

Table 3: Protein families from genes of interest and similar sequences from BLAST result

The result obtained from the protein family search further confirms the significance of the relationship between our genes of interest and the similar sequences obtained from the BLAST. Table 3 indicates that these genes produce proteins that belong to the same family. Proteins belonging to the same family exhibit the same function.

Gene of interest	Regions of conserved domains (bp)
ADS	**401-1414**
CYP71AV1	**30-1488**

Table 4: Regions of conserved sequences obtained from multiple sequence alignments of gene of interest with similar sequences

Table 4 gives regions of conserved sequences within the mRNA coding sequence of our genes of interest. The regions were obtained after performing a multiple sequence alignment with the similar sequences obtained from the BLAST result using CLUSTAL W. The regions outlined in the table show 100% conservation which was indicated by asterisks spread across the sequences aligned over different ranges. These conserved regions are important for gene function as they have been maintained from one specie to another during evolution. Thus these regions are important for gene expression.

Gene of interest	primer	sequence	Length (bp)	Expected Length of amplicon (bp)	Time °C	GC %
ADS	Forward primer	acaacgggcactaaagcaac	20	188	60.2	50
	Reverse primer	aagctttccaccatttgcac	20		60.1	45
CYP71AV1	Forward primer	caggcacagacacttcctca	20	218	60.0	55
	Reverse primer	ctctcttggcagaaccaagg	20		60.0	55

Table 5: Primers designed for the genes of interest using primer3plus

Table 5 gives the pairs of primers for both ADS and CYP71AV1 developed from their conserved sequences. The Primer3plus program gives a choice of 5 primer pairs to choose from for each pair of primers designed. These 5 pairs can be carefully scrutinized and an ideal primer pair for each gene picked based on their GC content, melting temperature, and primer length. These primers developed adhered to the requirements for appropriate PCR primers for amplifying the genes of interest.

The expected amplicon sizes will enable us determine to an extent whether the amplicons obtained from PCR are the expected amplicons and not non specific products. It also helped us to determine whether the amplicon size that will be obtained will not be too large to be seen on the agarose gel.

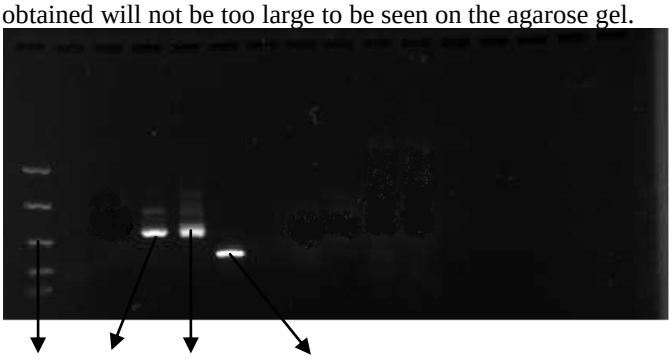

Figure 1: Gel electrophoresis of RT-PCR products

L= molecular weight ladder

+C= Positive control

CYP71AV1 = *Artemisia annua* amorpha 4,11-diene – mono oxygenase cDNA

ADS = *Artemisia annua* amorpha- 4,11 – diene synthase cDNA

Figure 1 shows the results of gel electrophoresis of amplicons obtained after a RT-PCR with the developed primers. Amplicons obtained for both ADS and CYP71AV1 appeared sharp, indicating the presence of good quantity of cDNA obtained. The amplicon sizes also comply with the expected amplicon sizes when viewed against the molecular weight ladder.

S/NO	cDNA	Length			GC%
		Normal	QV≥16	QV≥20	
1	Art ann ADS	880	618	286	36.0
2	Art ann CYP71AV1	897	580	390	38.0

Table 6: Sequencing of genes of interest from amplicons obtained

Table 6 shows the results of the sequenced cDNA of both ADS and CYP71AV1 after a RT-PCR with developed primers. The key to understanding results obtained is from the quality value, the highest quality value being 20 and the second highest a 16. Thus, the more the number of nucleotides from the normal count having quality values ≥16 and then ≥20 the better the quality of the sequence obtained. The two genes of interest appear to yield good quality values. Figure 2 and 3 gives the nature of sequences obtained for the ADS and CYP71AV1 cDNA respectively.

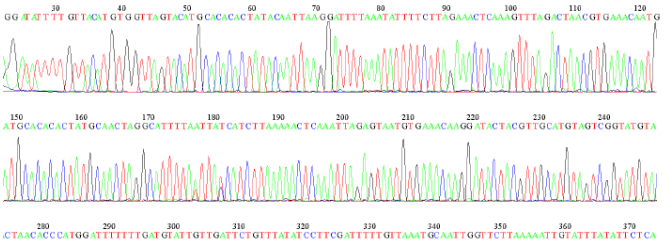

Figure 2: Sequenced Artemisia annua Amorpha 4, 11 diene synthase cDNA

ISSN: 2394-4404

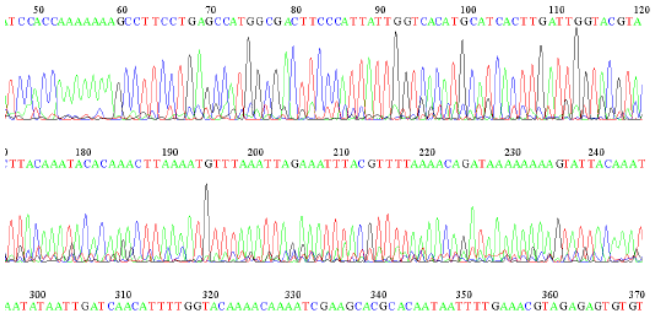

Figure 3: Sequenced Artemisia annua amorpha 4, 11 diene mono oxygenase (CYP71AV1) cDNA

SEQUENCED cDNA	SIMILAR SEQUENCE	ACCESSION NO	E value	Score	% IDENTITY
***Artemisia annua* Amorpha 4, 11 diene synthase (ADS) cDNA**	, *Artemisia annua* amorpha 4, 11 diene synthase (ADS) gene, complete coding sequence	FJ432667.1.	4e-28	119 bits (64)	88
***Artemisia annua* amorpha 4, 11 diene mono oxygenase (CYP71AV1) cDNA**	*Artemisia annua* P450 mono oxygenase (CP71) gene, complete cds	DQ667170.1	0.0	926	**99**

Table 7: BLAST for the sequenced genes

The results for the BLAST of sequenced cDNA against the NCBI database are seen in table 7. Nucleotides from the best regions of each sequence were obtained (30-350 for ADS and 60-500 for CYP71AV1) and a BLAST search was done restricting the search for similar sequences to *Artemisia annua*. The results show an 88% identity for the ADS and a 99% identity for the CYP71AV1 gene. This result confirms obtaining these genes from primers developed.

IV. CONCLUSION

Designing appropriate primers from conserved sequences for the ADS and CYP71AV1 genes from the artemisinin biosynthetic pathway was achieved. These primers designed satisfy the guidelines for primer design. . The primers developed were used to obtain amplified cDNA of the genes of interest. The result of the BLAST confirmed the identity of both ADS and CYP71AV1.

The method used was adequate in designing primers that will ensure conserved sequences are amplified during PCR. This method would ensure amplicons obtained contain the necessary sequence areas for ideal gene activity.

REFERENCES

[1] Abd-Elsalam K.A. (2003). Bioinformatic tools and guidelines for PCR primer design. African Journal of Biotechnology, 2(5), 91 – 95.
[2] Arsenault P, Wobbe KK, Weathers PJ. (2008). Recent advances in artemisinin production through heterologous expression. Current medicinal chemistry, 15 (27): 2886-2896.
[3] Dieffenbach C.W, Lowe T.M, and Dveksler G.S. (1993). General concepts for PCR Primer Design. Genome Research, 3: S30 – S37.
[4] Dieffenbach C.W, Lowe T.M.J, Dveksler G.S. (1995). General concepts for PCR Primer Design. In: PCR Primer, A Laboratory Manual. New York. Cold Spring Harbor Laboratory Press. 133 – 155.
[5] He Q, Marjamaki M, Soini H, Mertsola J, Vijanen MK. (1994). Primers are decisive for sensitivity of PCR. Biotechniques, 17: 82 - 87
[6] Newton and Graham. (1997). PCR. Oxford: BIOS Scientific publications. p23.
[7] Pevsner J. (2009) Bioinformatics and functional genomics. New Jersey. John Wiley and sons' Inc. 101-127, 179-190.
[8] Rathore D, McCutchan TF, Sullivan M, Kumar S (2005). Antimalarial drugs: current status and new developments. Expert opinion on investigational drugs, 14 (7): 871-883.
[9] Rychlik W, Spencer W.J, and Rhoads R.E (1990). Optimization of annealing temperature for DNA amplification in vitro. Nucleic Acid Research, 18: 6409 - 6412
[10] Schramek N, Wang H, Romisch-Margl W, Keil B, Radykewicz T, Winzenhorlein B, Beerhues L, Bacher A, Rohdich F, Gershenzon J, Liu Benye, Eisenreich W (2010). Artemisinin biosynthesis in growing plants of Artemisia annua. A 13CO2 study. Phytochemistry, 71(2): 179-187.
[11] Singh V.K and Kumar A. (2001). PCR Primer Design. Molecular Biology Today, 2(2): 27-32.
[12] Teoh K.H, Polichuk D.R, Reed D.W, Nowak G et al. (2006). Artemisia annua L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. FEBS letters, 580 (5): 1411-16.
[13] Weathers PJ, Elkholy S, Wobbe K (2006). Artemisinin: the biosynthetic pathway and its regulation in Artemisia annua, a terpenoid-rich species. In vitro cell Development. Biology-Plant, 42: 309-317.
[14] Wu D.Y, Ugozzoli W, Pal B.K, Qian J, and Wallace R.B. (1991). The effect of temperature and oligonucleotide primer length on specificity and efficiency of amplification by the polymerase chain reaction. DNA Cell Biology, 10: 223-238.