CHAPTER IV

RESULTS AND DISCUSSION

4.1 cDNA cloning of oxidosqualene cyclase gene from *Alangium lamarckii* leaves

Full-length AlOSCs were amplified from *Alangium lamarckii* leaves cDNA by using primers, forward primer ALB_f and reverse primer ALB_r. The expected PCR product band at 2.3 kb was extracted and purified to ligate to pGEM-T vector and transformed into *E. coli* DH5α for propagation and sequencing. White colonies were individually picked and grown for plasmid extraction. The extracted plasmid was cut by *EcoR*I restriction enzyme for insertion checking. Six samples that contained expected insertion at 2.3 kb were sent for sequencing. The sequences were named as the followings: *AlOSC1*, *AlOSC2*, *AlOSC3*, *AlOSC4*, *AlOSC5* and *AlOSC6*.

The sequencing results with 2292 bp in length were translated into a 763 amino acid sequence by using BioEdit version 7.1.9 (Tom Hall, Ibis Biosciences, An Abbott company, USA) and Clone Manager (Scientific and Education Software, USA). The amino acid sequence identities between AlOSCs were compared by using Clone Manager (Scientific and Education Software, USA) (Table II-1).

Five of the six sequences were found to contain conserved amino acid regions, QW motif, which presumably has a role to stabilize the protein structure, and DCTAE motif, which is responsible for substrate protonation. These motifs are conserved in the OSC family [14]. Moreover, the genes were also found to have MWCYCR motif which is specifically found in the sequence producing β-amyrin (Appendix G) [15]. The results suggested that the genes from *A. larmarckii* leaves are triterpene synthase genes.

The DNA sequence identities between AlOSCs were compared by using Clone Manager (Scientific and Education Software, USA) (Table IV-2). *AlOSC4* showed to be the most identity to ALB, which is a putative triterpene synthase from *A. lamarckii* [45], both in DNA and amino acid sequence. The sequencing results were blasted to NCBI data base. The 5 sequences were identity to the reported genes in OSC family (Table IV-3). *AlOSC1, AlOSC3, AlOSC5 and AlOSC6* showed identity to each other and to amyrin synthase from *Catharanthus roseus*

[37], while *ALOSC4* showed identity to mixed amyrin synthase from *Olea europaea* [34].

Table IV-1 Amino acid sequence identity between AlOSCs

	AlOSC3	*AlOSC4*	*AlOSC5*	*AlOSC6*	*ALB*
AlOSC1	99	86	98	99	88
AlOSC3		86	98	99	88
AlOSC4			86	86	97
AlOSC5				98	88
AlOSC6					88

Table IV-2 DNA sequence identity between *AlOSC*s

	AlOSC3	*AlOSC4*	*AlOSC5*	*AlOSC6*	*ALB*
AlOSC1	99	90	99	99	91
AlOSC3		90	99	99	91
AlOSC4			90	90	97
AlOSC5				99	91
AlOSC6					91

Table IV-3 DNA sequence identity between *AlOSC*s and reported *OSC*s

Putative *AlOSC* gene	Reported genes	%identity
AlOSC1	Mixed amyrin synthase (*Catharanthus roseus*) AEX99665.1	80%
AlOSC3	Mixed amyrin synthase (*Catharanthus roseus*) AEX99665.1	80%
AlOSC4	Mixed amyrin (*Olea europaea*) BAF63702.1	79%
AlOSC5	Mixed amyrin synthase (*Catharanthus roseus*) AEX99665.1	80%
AlOSC6	Mixed amyrin synthase (*Catharanthus roseus*) AEX99665.1	80%

4.2 Functional expression of cDNA in yeast GIL77

Each *AlOSC* gene was prepared by introducing yeast consensus sequence (HAMAMA) up-stream of the sequence and was changed nucleotide codons of

the second amino acid from TGG to TCC for increasing expression in yeast. There were 40 samples from the diversity of introduced yeast consensus sequence and changing the amino acid codon. In this study, the samples were then classified into 2 groups, (1) TGG group, a group which yeast consensus sequence was introduced up-stream to *AlOSCs* and nucleotide codon at amino acid position 2 were TGG and (2) TCC group, a group which yeast consensus sequence was introduced up stream to *AlOSCs* and nucleotide codon at amino acid position 2 were TCC. The genes were then constructed into yeast expression pYES2 vector under the control of *GAL1* promoter. Forty constructs of pYES2-AlOSC were heterologously expressed in a yeast mutant, *Saccharomyces cerevisiae* GIL77 (*gal2 hem3-6 erg7 ura3-167*) which lacks lanosterol synthase that accumulated the triterpene synthase substrate, 2,3-oxidosqualene, and needed ergosterol supplement for growing the yeast. The empty vector pYES2 was transformed into the mutant yeast serving as the negative control in parallel to the experiment. The yeast samples were extracted by hexane and screened on a TLC plate. Interesting samples showed a pink spot corresponding to the standard β-amyrin and the positive control OEA [34] (Figure IV-1 and Figure IV-2).

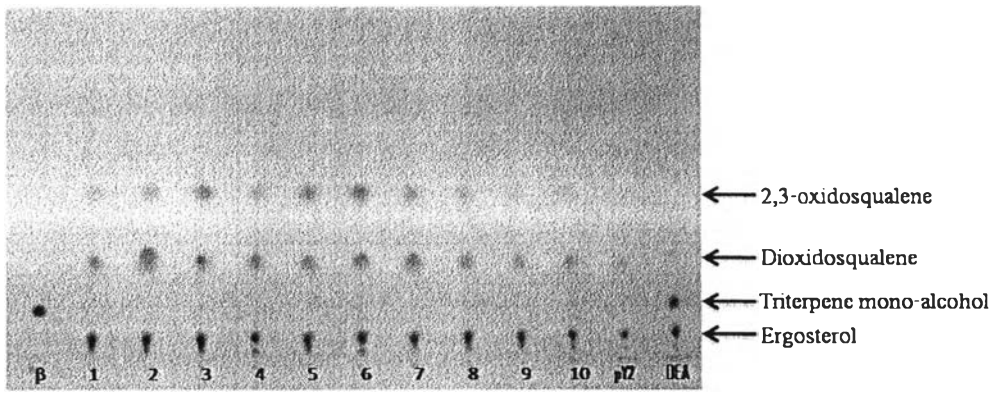

Figure IV-1 AlOSC product from yeast no.1-no.10 on a TLC plate sprayed with anisaldehyde-sulphuric acid reagent under visual light. Spot positions were labeled following the arrows. β, β-amyrin standard; pY2, pYES2 (negative control), OEA, multifunctional triterpene synthase from *Olea europaea* (positive control).

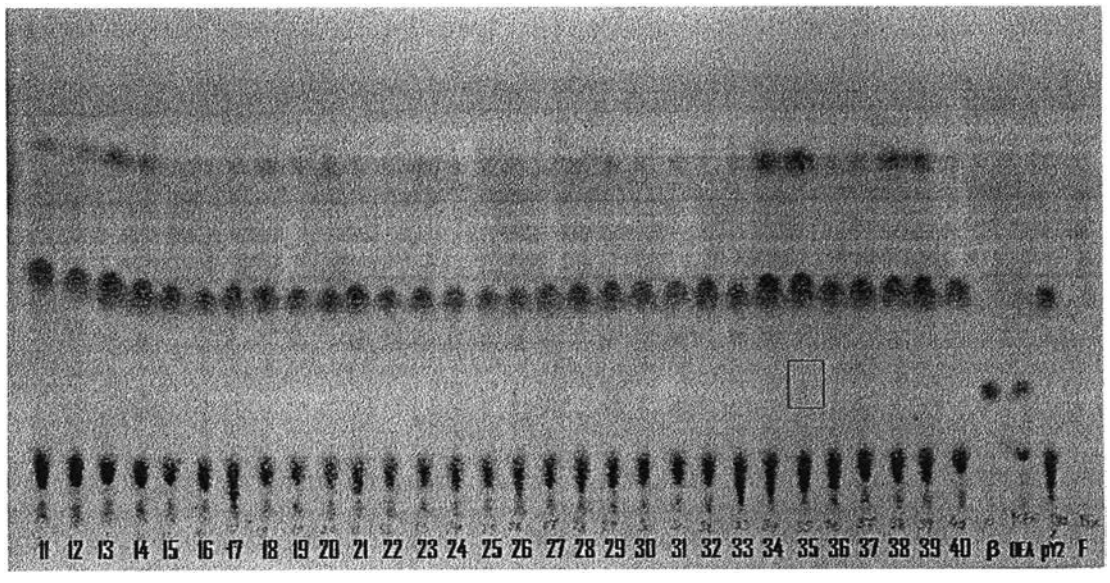

Figure IV-2 AlOSC product from yeast no.11-no.40 on a TLC plate sprayed with anisaldehyde-sulphuric acid reagent under visual light. β, β-amyrin standard; pY2, pYES2 (negative control), OEA, multifunctional triterpene synthase from *Olea europaea* (positive control).

The sample no. 35 showed to have an apparent pink spot corresponding to the standard ß-amyrin and positive control OEA than the others (square on Figure IV-2). Therefore, the sample no. 35 was scaled up to a large volume to increase the amount of product which has the same Rf to the standard. The extract was purified by TLC and analysed by HPLC (SHIMADZU). Only the obtained HPLC chromatogram (Figure IV-3) of the sample showed a peak at the retention time of 15 min compared to the standards and negative control. Compared to the standards, this peak is presumably the triterpene mono-alcohol taraxasterol or ß -amyrin.

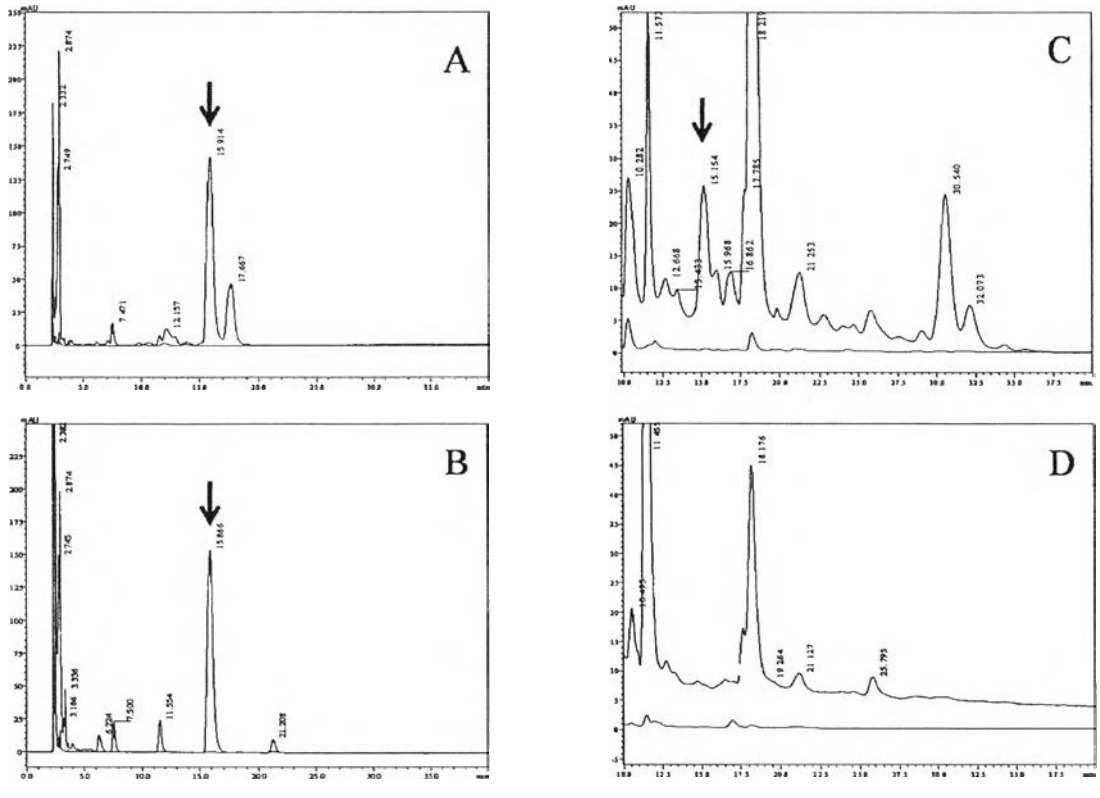

Figure IV-3 HPLC chromatograms of (A) standard taraxasterol (15.86 min), (B) standard ß -amyrin (15.91 min), (C) sample no. 35 and (D) negative control pYES2.

The peak was further confirmed by using the technique of liquid chromatography mass spectrometry (LC-MS) (Agilent 1200 series, Bruker MicrOTOF). Purified of putative triterpene mono-alcohol (35F1) and triterpene ketone (35F2) of sample no.35, which have the Rf corresponding to ß-amyrin and friedelin respectively, were analysed by using retention time and LC- APCIMS EIC compared to the standards and negative control (pYES2). Retention time and LC-

APCIMS EIC at *m/z* = 409-411 were used to compare with the triterpene mono-alcohol standards (taraxsterol, β-amyrin, Ψ-taraxasterol). Retention time and LC-APCIMS EIC at *m/z* = 425-430 were used to compare with triterpene ketone or friedelin. The peaks *m/z* = 409 of 35F1 were shown at retention times correspond to the peaks of triterpene mono-alcohol standards, taraxsterol (19.9 min), β-amyrin (20.4 min) and Ψ-taraxasterol (21.8 min) (Figure IV-4, Figure IV-5 and Figure IV-6).

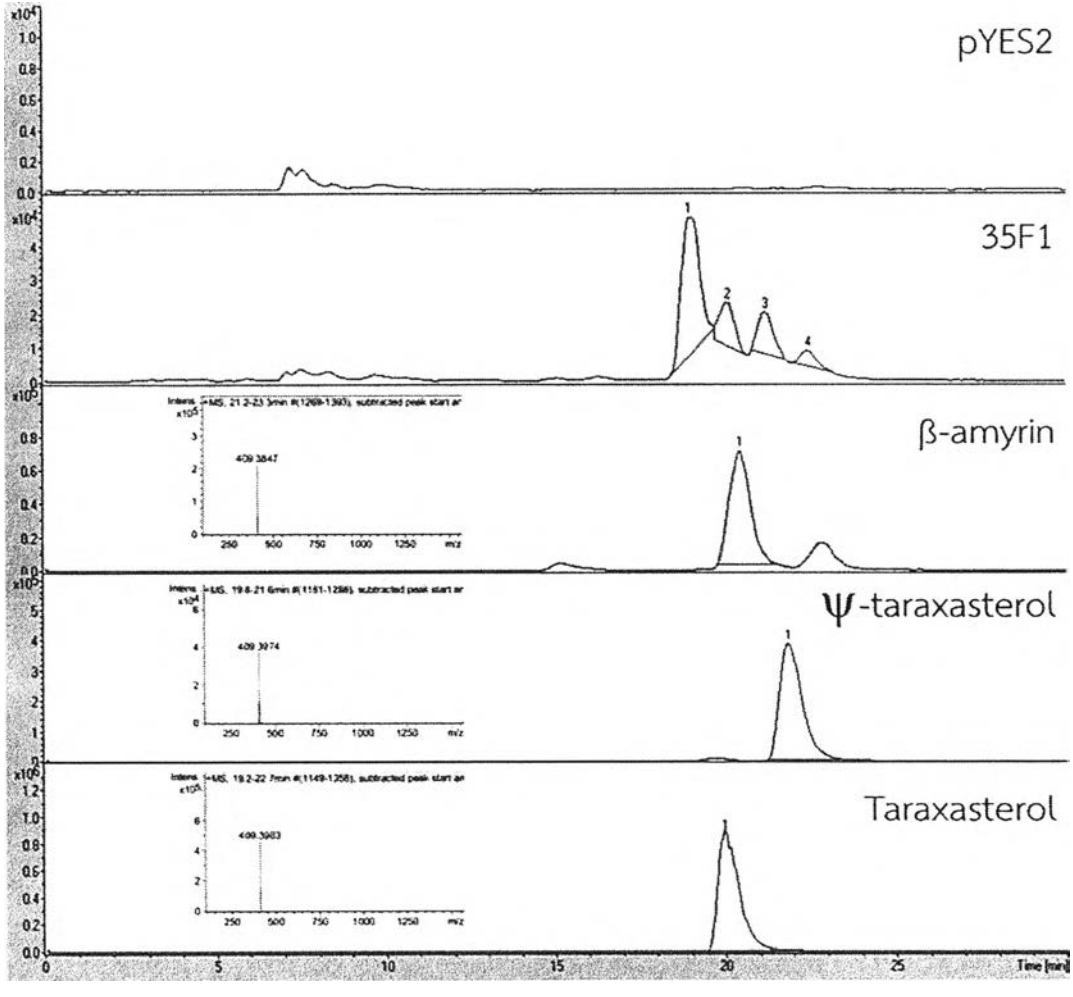

Figure IV-4 The extracted ion chromatogram of fraction 35F1, pYES2 and the standards. 35F1 showed the *m/z*=409 peaks correspond to the standards and produced taraxasterol as major product (1), β-amyrin (2), Ψ-taraxasterol (3) and unknown triterpene mono-alcohol (4)

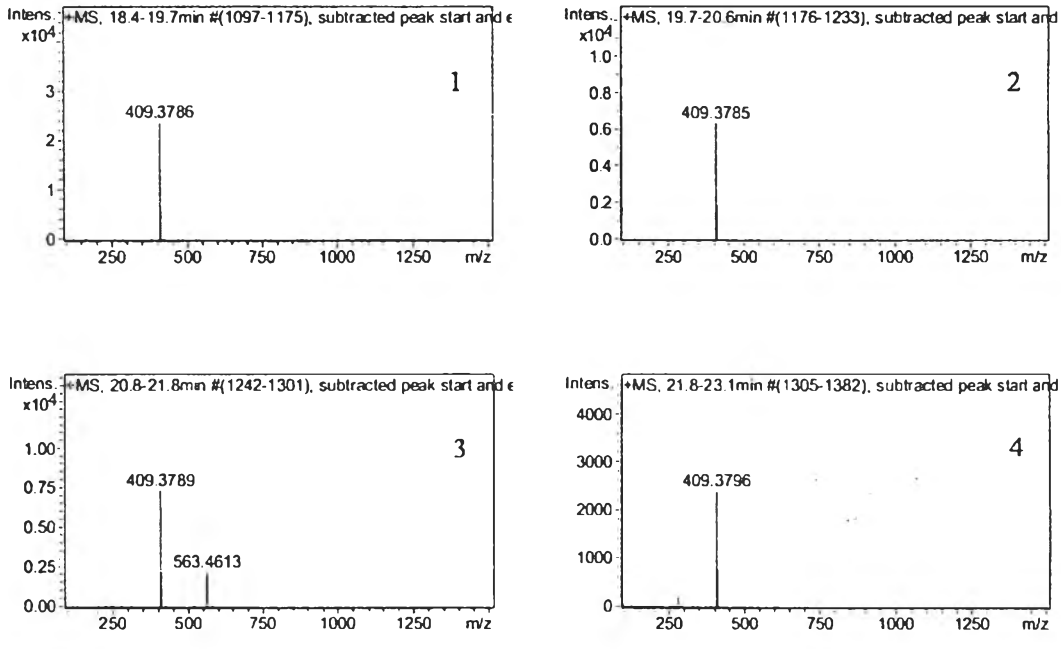

Figure IV-5 Mass spectrum of the peaks in the extracted ion chromatograms of fraction 35F1. 1) at retention time 18.9 min, 2) at retention time 20.1 min, 3) at retention time 21.2 min and 4) at retention time 22.5 min.

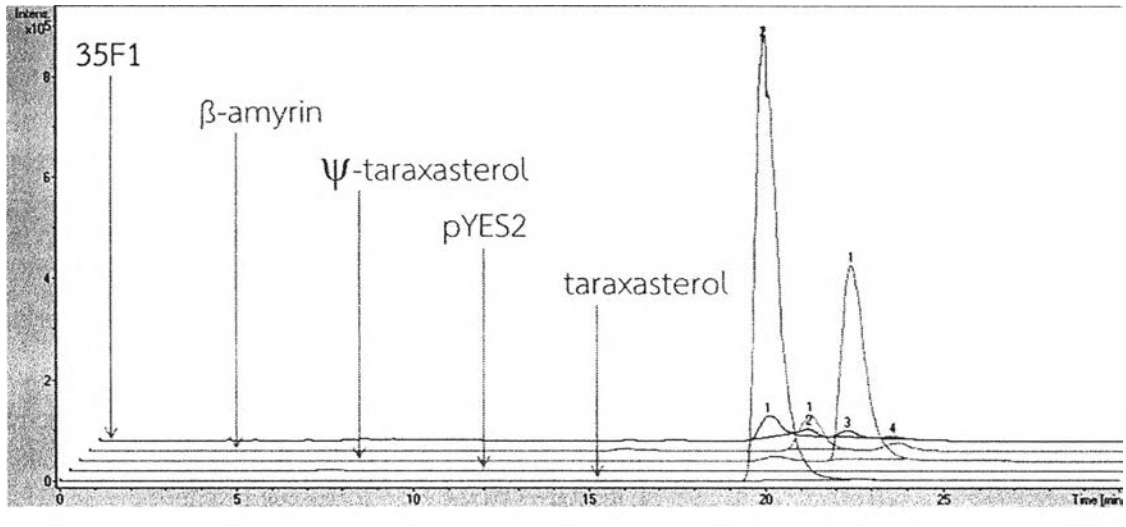

Figure IV-6 The stacked plots of extracted ion chromatogram of 35F1 and standards.

The sample no. 35, which is one of the 4 genes, showed 80% identity to amyrin synthase from *C. roseus*, appeared to produce taraxasterol as a major product, β-amyrin, Ψ-taraxasterol and an unkown triterpene mono-alcohol, while there was no detection of triterpene ketone of 35F2 compare to *m/z* = 427 of friedelin (Figure IV-7).

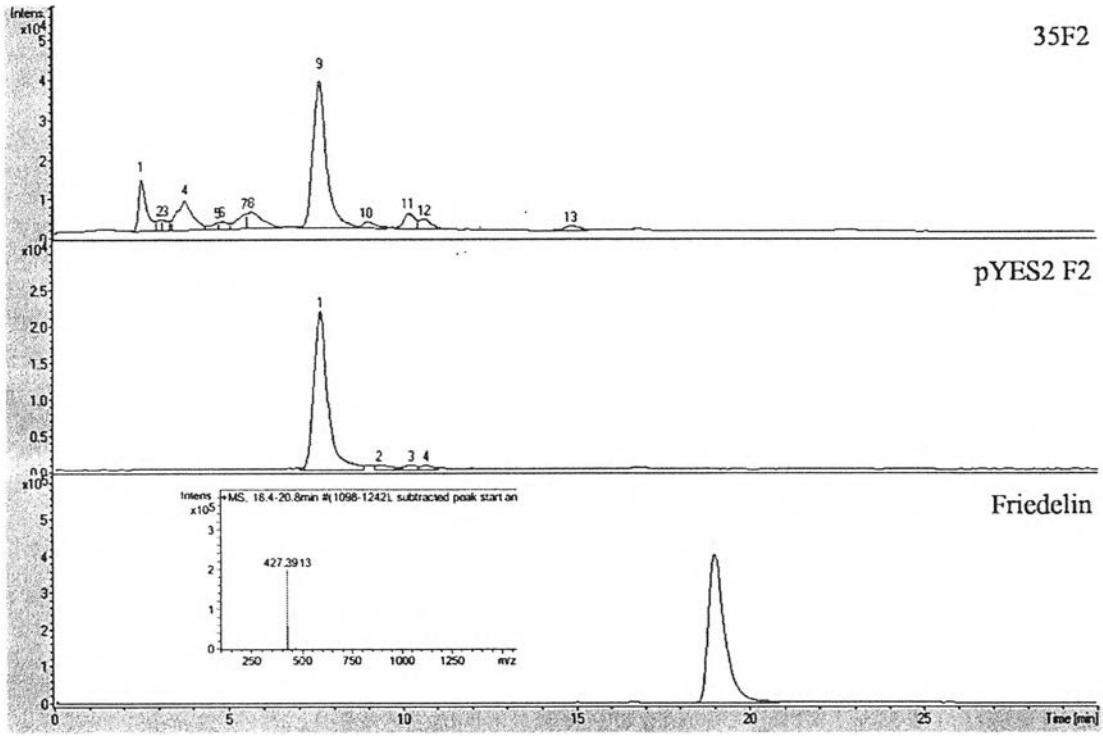

Figure IV-7 The extracted ion chromatogram of fraction 35F2, pYES2 and the standard friedelin. 35F2 produced none of *m/z*=427 peak that correspond to the retention time of the standard friedelin.

According to the literature review, many of OSC genes that produce taraxasterol were reported as multifunctional triterpene synthases [18], [8], [29] and [34]. The biosynthesis pathway of taraxasterol and Ψ-taraxasterol has to pass lupenyl cation before the rearrangement of a methyl and hydrides to be oleanyl cation and taraxasteryl cation respectively, and deprotonated at C20. In addition, the cyclisation step of β-amyrin also has to pass oleanyl cation and deprotonated at C12 (Figure II-2). From the biosynthetic pathway, it can be proposed that the 3 triterpenes share the same pathway that they have to pass the intermediate oleanyl cation before the rearrangement to their preferable products. According to the mentioned pathway, there is an interesting point on the relationship between the pathway and the amino acid sequence of the

MWCYCR motif which is normally present in the gene producing β-amyrin including *AlOSC6* found in this study.

Previous study [32] has shown that tyrosine (Y) is responsible for E-ring expansion to be a pentacyclic triterpene, and tryptophan (W) plays an important role to stabilize oleanyl secondary cation with cation-¶ interaction. From the pathway between to the taraxastanes (taraxasterol and Ψ-taraxasterol) and to oleanane (β-amyrin), the possible way of methyl and hydride shift can occur to get both structure types. For the taraxastanes, there would present of a basic amino acid near the E-ring and deprotonate at C20 but preferable to form taraxasterol as a major product.

4.3 Yeast consensus sequence and OSC protein expression

AlOSCs have high % identities to ALB as previously described. Previous study showed that ALB had low expression of triterpene production and product detection was unclear. Therefore, this investigation introduced yeast consensus following protocol suggestion. From the original sequence, the second codon after ATG of AlOSCs is TGG which encode for tryptophan. To introduce the yeast consensus sequence, TGG was changed to be TCS (C or G) which encode for serine. The sample no.35 which is in TGG group showed to express for triterpene mono-alcohol production. Therefore, all AlOSC6 were sent for sequencing of the yeast consensus sequence. After sequencing, the second codon sequence of sample no.40 which supposed to be in the TGG group were TCC, it might be the error from primer synthesis. While the second codon sequence of sample no.5 and no.20 were TCG which also encodes for serine, so the TCC sequences were chose for scaling up. Therefore, the samples were divided to 2 groups, TCC and TGG (Table IV-4). The yeast samples no.15 and 25 for TCC and the yeast samples no.10 and 30 for TGG were cultured at 800 ml of each sample.

Table IV-4 Sequencing results of primer introduced yeast consensus sequence

Group	Sample no.	introduced yeast consensus sequence result
TCC	5	GTACTG**AAGCTT**CACAAATGTCG
	15	GTACTG**AAGCTT**CAAACAATGTCC
	20	GTACTG**AAGCTT**CAAAAAATGTCG
	25	GTACTG**AAGCTT**TAAAAAATGTCC
TGG	10	GTACTG**AAGCTT**AAAAAAATGTGG
	30	GTACTG**AAGCTT**CAAACAATGTGG
	35	GTACTG**AAGCTT**CACACATGTGG
	40	GTACTG**AAGCTT**TAAAAAATGTCC

Each isolated product F1, which has the Rf related to authentic triterpene mono-alcohol, from yeast extract was applied to LC-APCIMS using the same condition to sample no.35. The results showed that the sample from TGG group expressed for triterpene production corresponding to sample no.35 and the standards (Figure IV-8, Figure IV-9 and Figure IV-10), while the samples from TCC group showed to produce none of interested triterpene.

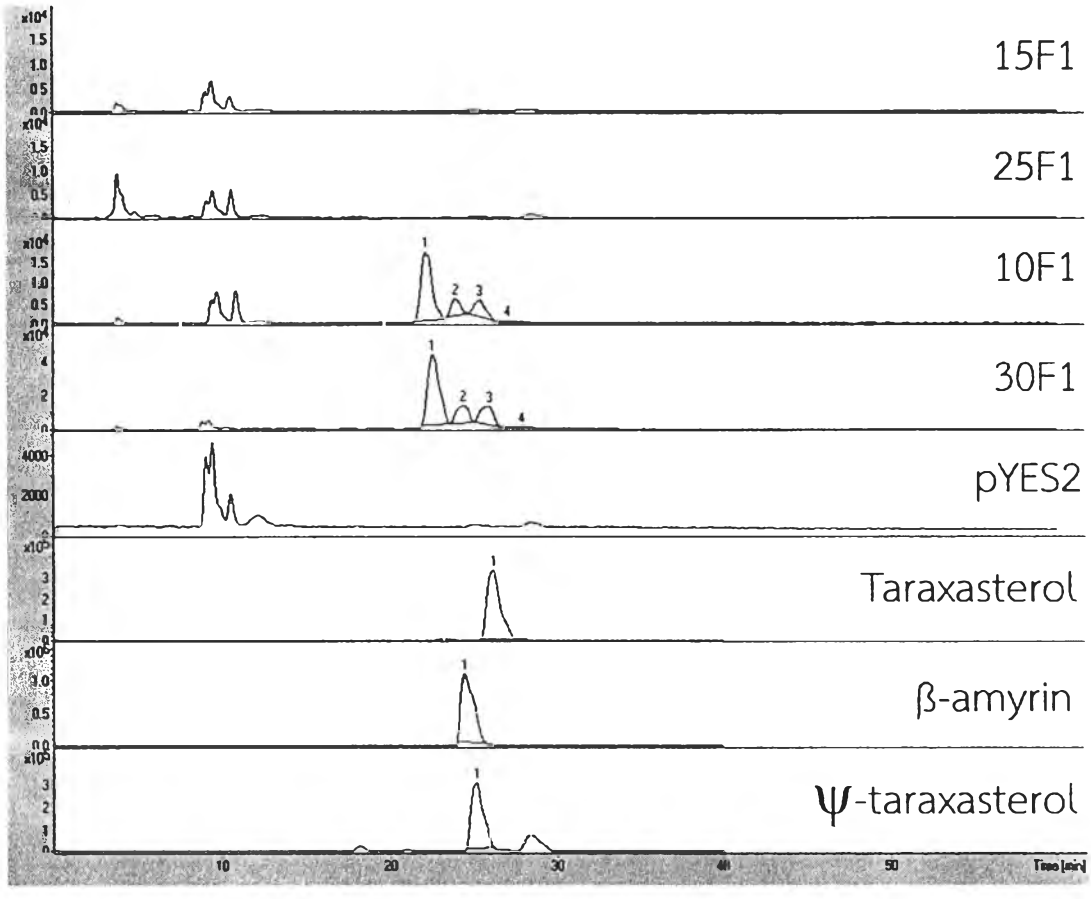

Figure IV-8 The extracted ion chromatogram of fraction TCC group (15F1 and 25F1), TGG group (10F1 and 30F1), pYES2 and standards.

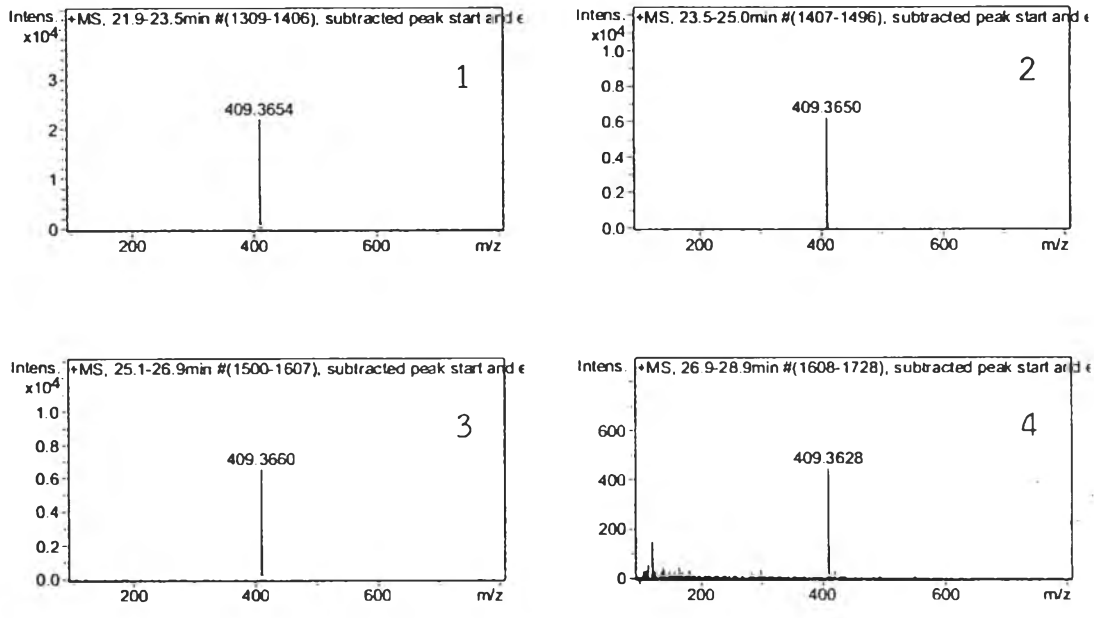

Figure IV-9 Mass spectrum of the peaks in the extracted ion chromatograms of fraction 10F1. 1) at retention time 22.5 min, 2) at retention time 24.5 min, 3) at retention time 25.9 min and 4) at retention time 27.9 min.

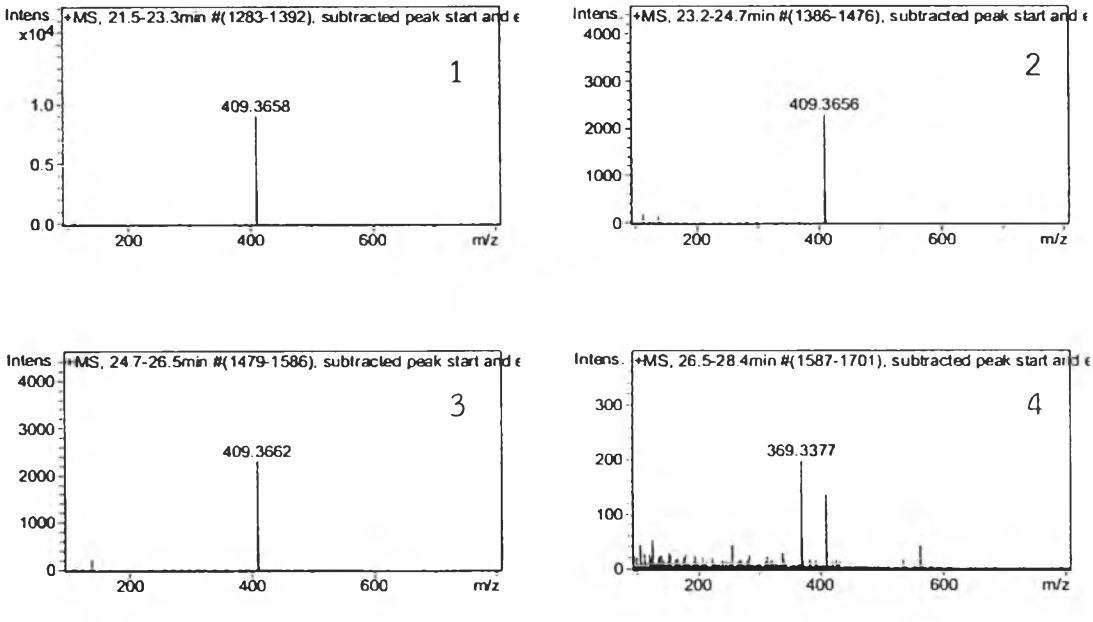

Figure IV-10 Mass spectrum of the peaks in the extracted ion chromatograms of fraction 30F1. 1) at retention time 22.1 min, 2) at retention time 23.9 min, 3) at retention time 25.4 min and 4) at retention time 27.1 min.

Due to the small amount of the isolated compounds of each sample, the interested peaks at expected retention times might be interrupted by the residue. Spike internal standard technique was used. Three standards (taraxasterol, β-amyrin and Ψ-taraxasterol) were used. Retention time and LC-APCIMS EIC at *m/z* = 409-411 were used for analysis (Figure IV-11 and Figure IV-12). Based on these present results, it can be concluded that TCS (C or G) codon which encode for serine had an effect to protein structure and resulted to none of triterpene production, and the yeast consensus sequence was assumed to have an effect on the efficiency of translation initiation in this investigation.

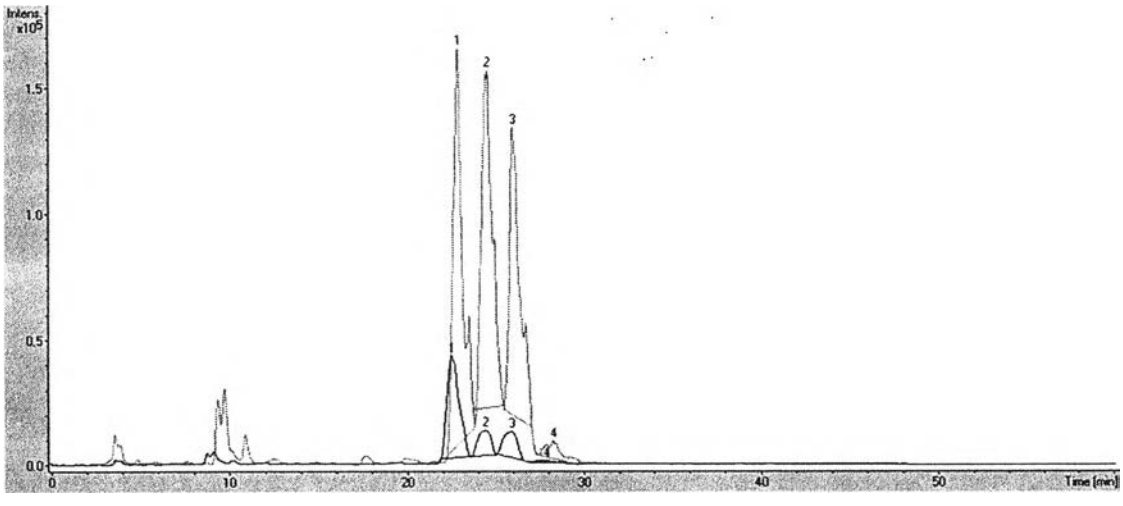

Figure IV-11 The stacked plots of extracted ion chromatogram of 10F1 and spiked the internal standards.

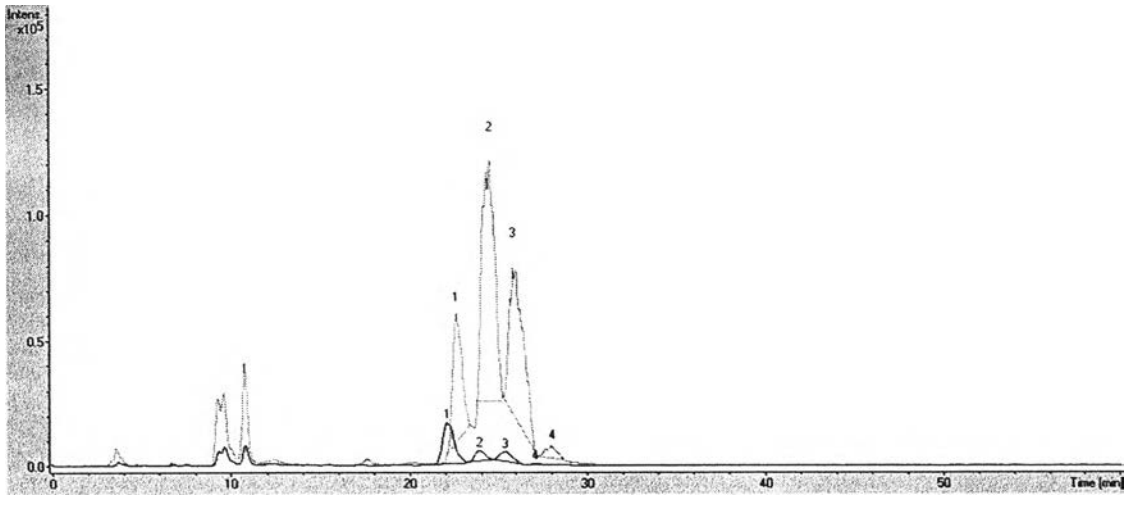

Figure IV-12 The stacked plots of extracted ion chromatogram of 30F1 and spiked the internal standards.

4.4 Triterpenoid analysis in *Alangium larmarckii* leaves

A. lamarckii is a small to medium tree belonging to the family Alangiaceae. Previous study from India has been shown that friedelin is present in *A. lamarckii* leaves [5]. Therefore, this plant is an ideal source for the study of this triterpene ketone compound. However, due to the geography and the environment problems that could affect the chemical constituents in the plant and also there has no report of the triterpenoid constituents for *A. lamarckii* in Thailand, triterpene analysis in this Thai plant is needed.

A. lamarckii leaves were collected freshly, dried at 50°C, extracted with MeOH and partition with hexane. The extract was concentrated and separated on a TLC plate developed with 5% acetone in hexane twice. After spraying with anisaldehyde-sulphuric acid reagent (AS) and heated at 100°C for 5 mins, fraction f1 showed a light red spot related to the standard ß-amyrin. Fraction f2 showed light yellow and pink spots overlapped at the Rf correspond to the standard friedelin which gives light yellow color for a short time after spraying with anisaldehyde-sulphuric acid reagent and only during heating at 100°C (Figure IV-13).

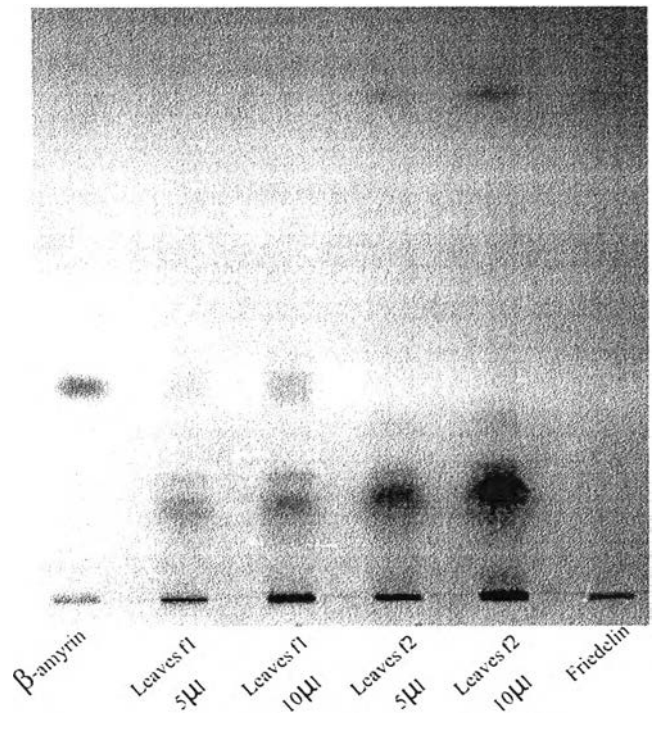

Figure IV-13 TLC chromatogram of the *A. lamarckii* leaves extracts.

The f1 and f2 eluates were then subject to LC-APCIMS. At APCI mode, EIC at *m/z*=409-411, the eluate f1 showed peaks at *m/z* = 409 $[M+H-H_2O]^+$ compared to the retention times of the standards, lupeol, taraxasterol, β-amyrin and Ψ-taraxasterol, relatively. There are four peaks that have retention time related to the standards. From the spiking of three internal standards, taraxasterol, β-amyrin and Ψ-taraxsterol, into the sample no. 10 and no.30 showed that the mixing of the standards and sample gave peaks shift earlier especially the taraxsterol peak. This can be explained that the peaks no. 2, 4, 5 and 6 of the leaves extract 1 are lupeol, taraxasterol, β-amyrin and Ψ-taraxasterol, relatively (Figure IV-14, Figure IV-15 and Figure IV-16), while the other peaks of the leaves extract 1 may be other triterpene mono-alcohols because the fragmentation gave all ionized mass at *m/z* = 409. However, the eluate f2 showed peak at *m/z* = 427 $[M+H]^+$ related to the retention time and *m/z* of the standard friedelin (Figure IV-17). These results can be concluded that *A. lamarckii* leaves contain of triterpene mono-alcohols and triterpene ketone.

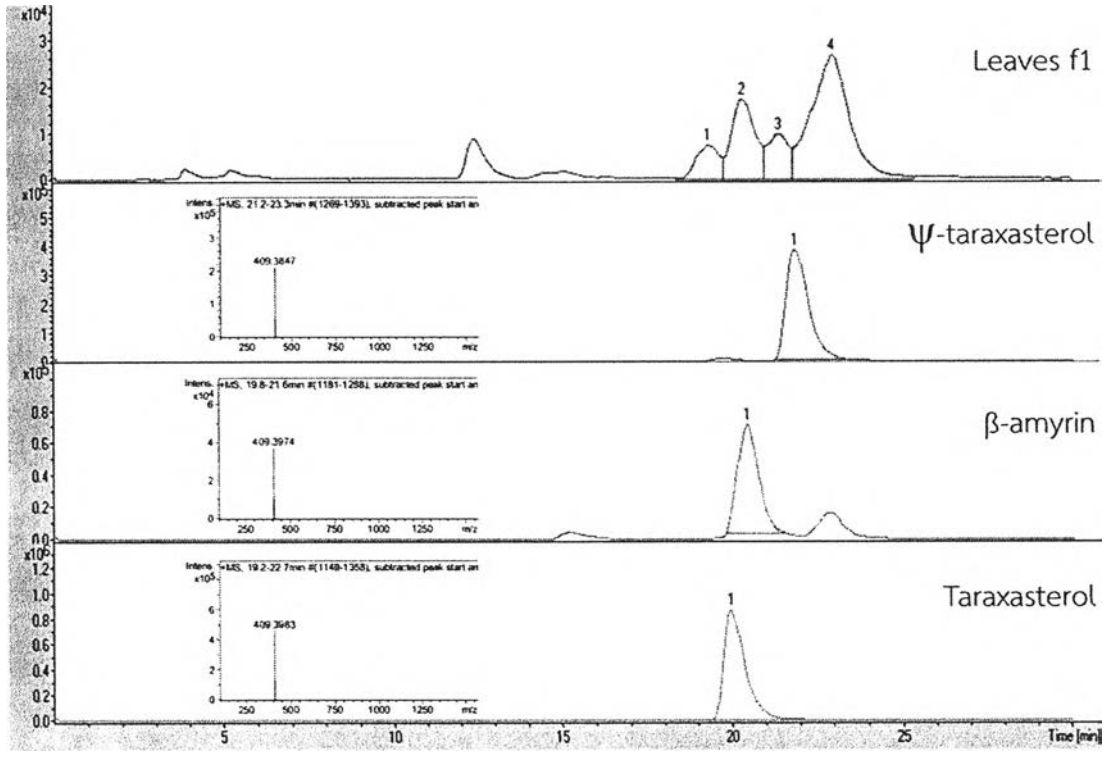

Figure IV-14 The extracted ion chromatogram of leaves extract fraction f1 and standards.

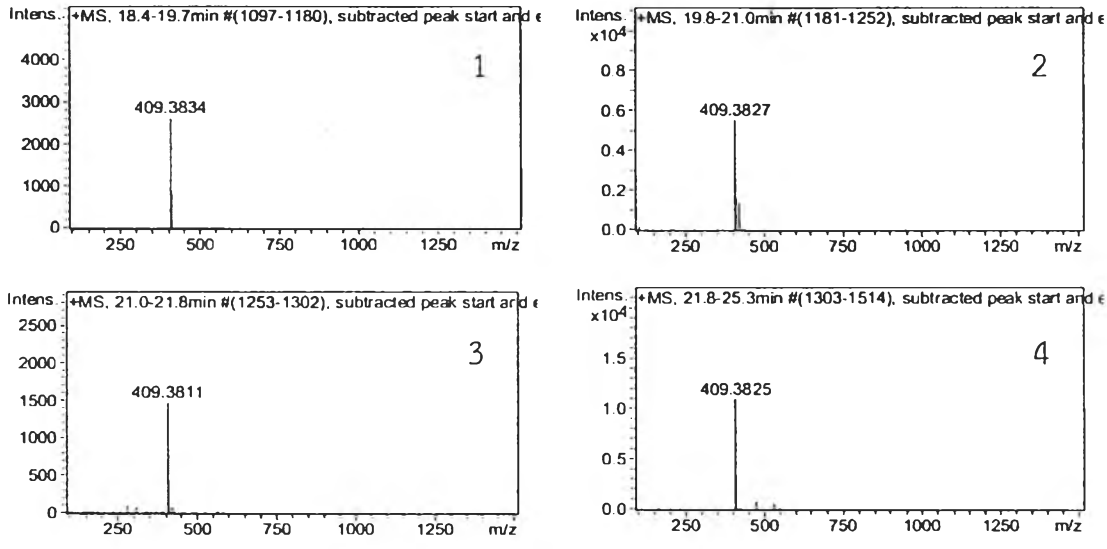

Figure IV-15 Mass spectrum of the peaks in the extracted ion chromatograms of fraction f1. 1) at retention time 19.3 min, 2) at retention time 20.3 min, 3) at retention time 21.4 min and 4) at retention time 23.0 min.

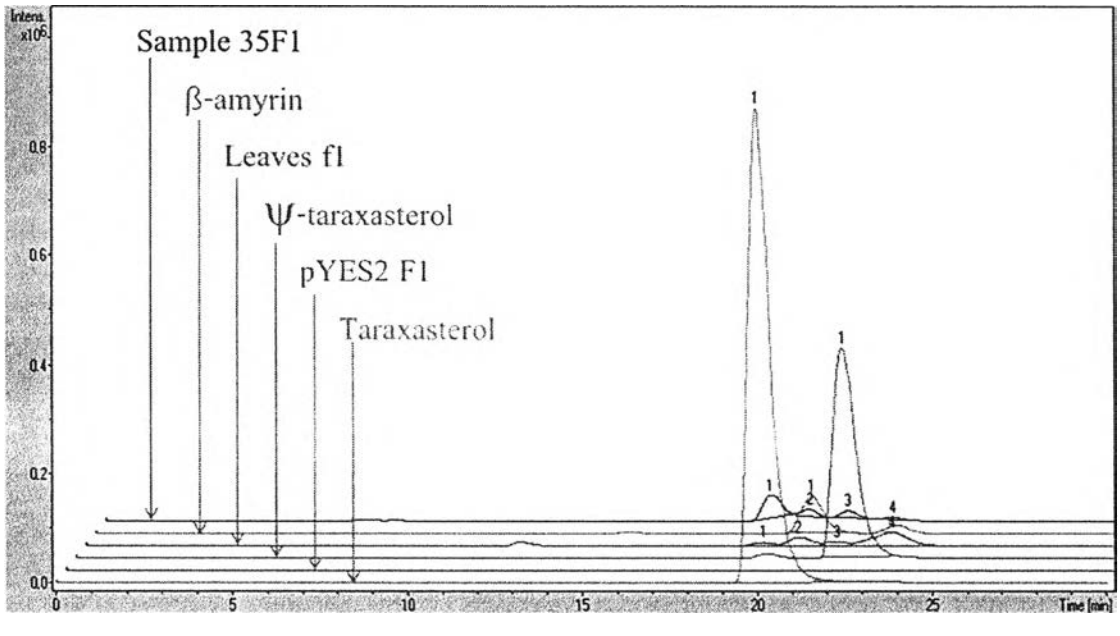

Figure IV-16 The stacked plots of extracted ion chromatogram of leaves f1, sample 35F1 and the standards.

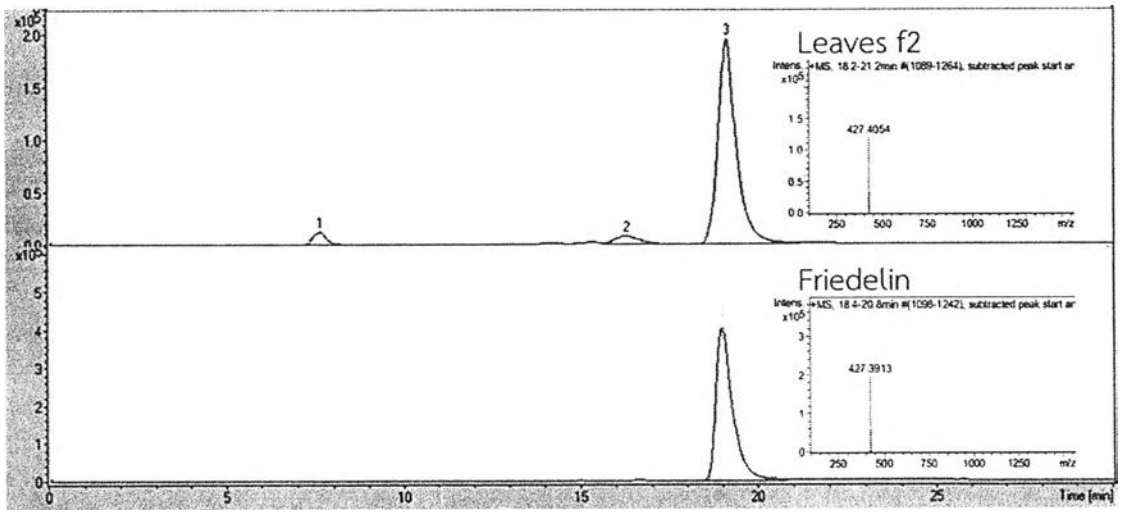

Figure IV-17 The extracted ion chromatogram of leaves extract fraction f2 and friedelin.

Phylogenetic tree analysis

A rooted phylogenetic tree was created by using MEGA version 6 program based on the deduced amino acid sequences of AlOSC1, AlOSC3, AlOSC4, AlOSC5, AlOSC6 and characterized OSCs from other plants with human lanosterol synthase as the outgroup [47] (Figure IV-18). The phylogenetic tree shows the clusters of lanosterol synthase and cycloatenol synthase separated from other OSCs. While the branches of OSCs produce specific β-amyrin show to be relate with OSCs produce multi-triterpenes, which are mostly reported to produce β-amyrin as one of their products. AlOSCs show related to CrAS and OEA which characterized as multifunctional enzymes. By the way, AlOSC1, AlOSC3 and AlOSC6 show very close relation, which would be possible that AlOSC1 and AlOSC3 would be predicted to express the same as AlOSC6.

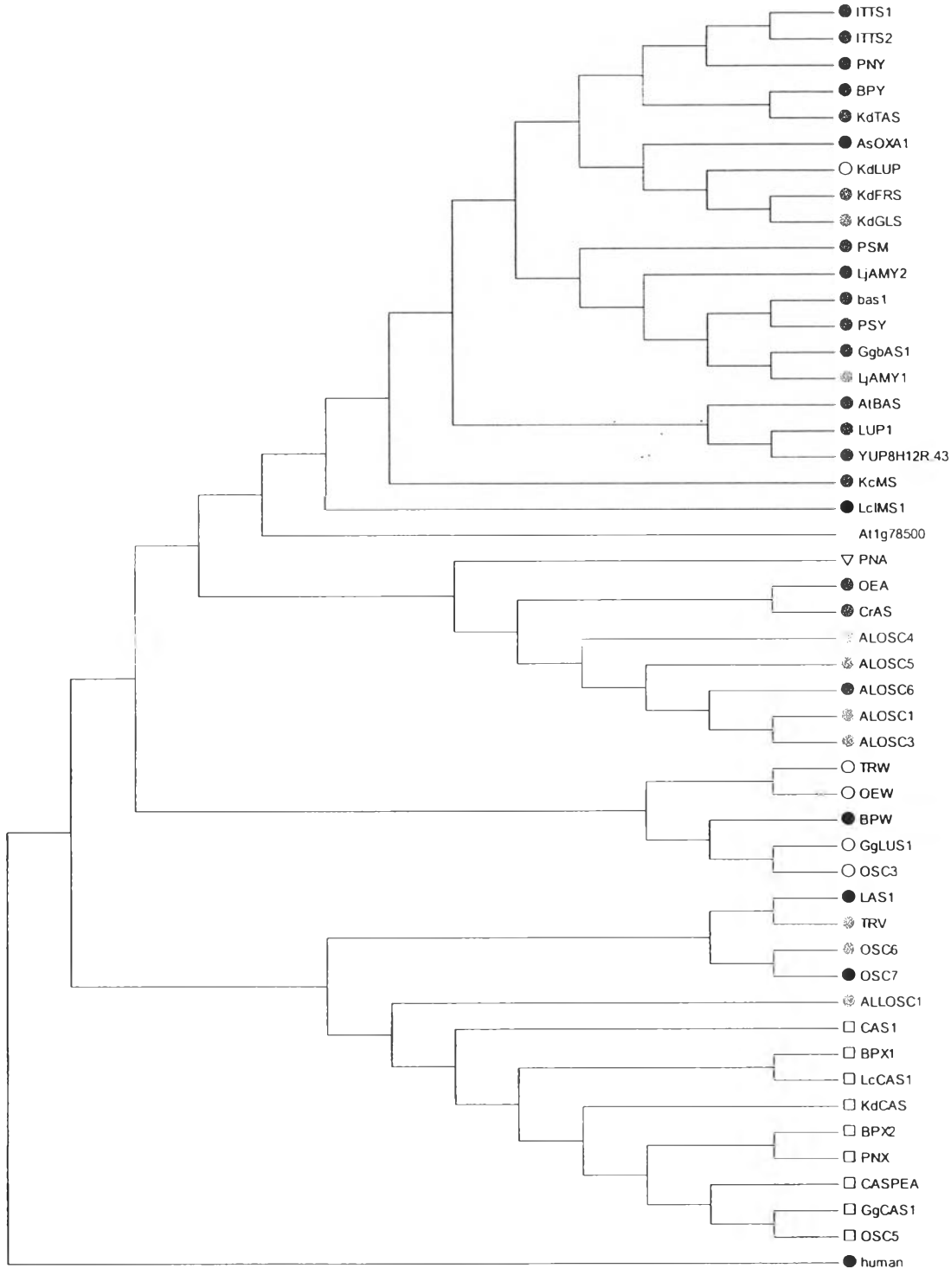
ITTS1
ITTS2
PNY
BPY
KdTAS
AsOXA1
KdLUP
KdFRS
KdGLS
PSM
LjAMY2
bas1
PSY
GgbAS1
LjAMY1
AtBAS
LUP1
YUP8H12R.43
KcMS
LcIMS1
At1g78500
PNA
OEA
CrAS
ALOSC4
ALOSC5
ALOSC6
ALOSC1
ALOSC3
TRW
OEW
BPW
GgLUS1
OSC3
LAS1
TRV
OSC6
OSC7
ALLOSC1
CAS1
BPX1
LcCAS1
KdCAS
BPX2
PNX
CASPEA
GgCAS1
OSC5
human

Figure IV-18 A rooted phylogenetic tree analysis comparing AlOSCs amino acid sequence and reported OSCs from plants (major products and accession numbers were given in Table II-1). *A. macrostemon*, ALLOSC1; *A. sedifolius*, AsOXA1; *A. thaliana*, At1g78500, LUP1, YUP8H12R.43, CAS1, LAS, LSS and At1g78950; *B. platyphylla*, BPX1, BPX2, BPW and BPY; *C. roseus*, CrAS; *G. glabra*, GgLUS1, GgCAS1 and GgbAS1; *K. candel*, KcMS; *K. daigremontiana*, KdFRS, KdCAS, KdLUP, KdGLS and KdTAS; *L. japonicas*, LjAMY1, LjAMY2, OSC3, OSC5, OSC6 and OSC7; *L. lindrica*, LcIMS1 and LcCAS1; *M. truncatula*, bAS; *O. europaea*, OEA, OEX and OEW; *P. ginseng*, PNA, PNX and PNY; *P. sativum*, PSY, PSM and PSX; *S. lycopersicum*, ITTS1 and TTS2; *T. officinale*, TRV and TRW.