

UNIVERSIDADE ESTADUAL PAULISTA

“JÚLIO DE MESQUITA FILHO”

Instituto de Química

Campus de Araraquara

ESTUDO DOS RESÍDUOS DE AMINOÁCIDOS DA ENZIMA FRIEDELINA SINTASE
DE *Maytenus ilicifolia* ENVOLVIDOS COM SUA ESPECIFICIDADE BIOSINTÉTICA

MELISSA REMLINGER

Orientador: Prof. Dr. Cleslei Fernando Zanelli

Co-orientadora: Dra. Tatiana Maria de Souza Moreira

Araraquara - SP

2017

UNIVERSIDADE ESTADUAL PAULISTA

“JÚLIO DE MESQUITA FILHO”

Instituto de Química

Campus de Araraquara

ESTUDO DOS RESÍDUOS DE AMINOÁCIDOS DA ENZIMA FRIEDELINA SINTASE
DE *Maytenus ilicifolia* ENVOLVIDOS COM SUA ESPECIFICIDADE BIOSINTÉTICA

MELISSA REMLINGER

Dissertação apresentada ao Programa de Pós
Graduação em Biotecnologia, do Instituto de
Química da Universidade Estadual Paulista
“Júlio de Mesquita Filho”, como parte dos
requisitos para obtenção do título de Mestre
em Biotecnologia.

Orientador: Prof. Dr. Cleslei Fernando Zanelli

Co-orientadora: Dra. Tatiana Maria de Souza Moreira

Araraquara - SP

2017

FICHA CATALOGRÁFICA

R384e Remlinger, Melissa
Estudo dos resíduos de aminoácidos de friedelina sintase de *Maytenus ilicifolia* envolvidos com sua especificidade biossintética / Melissa Remlinger. – Araraquara : [s.n.], 2017
78 f. : il.

Dissertação (mestrado) – Universidade Estadual Paulista, Instituto de Química
Orientador: Cleslei Fernando Zanelli
Coorientador: Tatiana Maria de Souza Moreira

1. Espinheira santa. 2. Terpenos. 3. Proteínas-Análise. 4. Mutagênese. 5. *Saccharomyces cerevisiae*. I. Título.

MELISSA REMLINGER

Dissertação apresentada ao Instituto de Química, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Mestra em Biotecnologia.

Araraquara, 24 de abril de 2017.

BANCA EXAMINADORA

Tatiana maria de Souza Moreira

Dr^a Tatiana Maria de Souza Moreira (Co-orientadora)
Faculdade de Ciências Farmacêuticas – UNESP, Araraquara - SP

Cíntia Milagre

Prof^a Dr^a Cíntia Duarte de Freitas Milagre
Instituto de Química – UNESP, Araraquara - SP

Rafael VC Guido

Prof. Dr. Rafael Victorio Carvalho Guido
Instituto de Física – USP, São Carlos - SP

RESUMO

Triterpenos são produtos naturais de plantas estruturalmente complexos com numerosas aplicações medicinais. Como exemplo, friedelina é um triterpeno pentacíclico com atividade gastroprotetora, enquanto que seus derivados quinonametídicos maitenina e pristimerina apresentam promissoras atividades anti-inflamatória, antitumoral, antimicrobiana, antimalárica, espermicida e antioxidante. O único triterpeno pentacíclico cetônico formado diretamente na ciclização do oxidoesqualeno é a friedelina. Sua produção se dá a partir da enzima friedelina sintase, uma oxidoesqualeno ciclase que se difere das demais oxidoesqualeno ciclases por estabilizar a ciclização promovida pelo carbocátion formado no sítio catalítico produzindo uma cetona. As oxidoesqualeno ciclases têm o mesmo substrato, porém se diferenciam pela especificidade de seus produtos formados e, por isto, a análise da estrutura primária da friedelina sintase permite estudar sua especificidade pela produção de friedelina. Desta forma, o presente trabalho teve como objetivo realizar o estudo da especificidade da friedelina sintase clonada de *Maytenus ilicifolia* por meio de mutantes desta enzima. Por meio de análises de *docking* da molécula de friedelina no sítio ativo da enzima foram observados resíduos de aminoácidos cuja interação poderia se relacionar à produção singular de friedelina. Os mutantes destes resíduos foram então gerados por mutagênese sítio-dirigida para avaliar tais interações de acordo com o produto formado heterologamente pelos mutantes expressos em *Saccharomyces cerevisiae*. Os resíduos estudados e gerados por mutação sítio-dirigida foram: F183L, C369A, W417H, D484E e W612F. Os produtos assim diferencialmente produzidos foram avaliados por cromatografia gasosa acoplada a espectrometria de massas (CG-EM). Assim, foi possível observar que os resíduos D484 e W417 são essenciais para a atividade da enzima, enquanto que os resíduos C369, W612 e F183 são importantes para a especificidade de produto, uma vez que a troca levou a produção de outros compostos terpênicos e/ou também a friedelina. Tais resultados permitiram avaliar resíduos importantes para a biossíntese de friedelina, contribuindo para o entendimento desta singular oxidoesqualeno ciclase.

Palavras-chave: *Maytenus ilicifolia*. Triterpenos. Friedelina sintase. Análise *in silico*. Sistema heterólogo

ABSTRACT

Triterpenes are natural plant products structurally complex with numerous medicinal applications. As an example, friedelin is a pentacyclic triterpene with gastroprotective activity whereas its quinone methide derivatives maytenin and pristimerin present promising anti-inflammatory, antitumor, antimicrobial, antimalarial, spermicidal and antioxidant activities. Friedelin is the only pentacyclic ketone triterpene obtained directly from cyclization of oxidosqualene. Its production occurs from the enzyme friedelin synthase, an oxidosqualene cyclase, which differs from the other oxidosqualene cyclases by stabilizing the cyclization promoted by the carbocation formed at the catalytic site forming a ketone. The oxidosqualene cyclases have the same substrate, but are differentiated by the specificity of their products formed and, therefore, the analysis of the primary structure of friedelin synthase allows to study its specificity by the production of friedelin. Thus, the present work aimed to study the specificity of the friedelin synthase cloned from *Maytenus ilicifolia* using mutants of the enzyme. Through the analysis of docking of the molecule friedelin in the active site of the enzyme, it was observed amino acid residues whose interaction could be related to the unique production of friedelin. Mutants of these residues were generated by site-directed mutagenesis to evaluate such interactions according to the product formed heterologously by the mutants expressed in *Saccharomyces cerevisiae*. Residues studied and generated by site-directed mutation were: F183L, C369A, W417H, D484E and W612F. The products thus differentially produced were evaluated by gas chromatography coupled to mass spectrometry (GC-MS) from the expression and production of the mutants using the heterologous *S. cerevisiae* system. Thus, the residues D484 and W417 are essential for enzyme activity, whereas residues C369, W612 and F183 are important for product specificity, since the exchange led to the production of other terpene compounds and / or also friedelin. These results allowed the evaluation of important residues for the biosynthesis of friedelin, contributing to the understanding of this unique oxidosqualene cyclase.

Keywords: *Maytenus ilicifolia*. Triterpenoids. Friedelin synthase. *In silico* analysis. Heterologous system

LISTA DE FIGURAS

- Figura 1** - Cátions intermediários e produtos da ciclização de 2,3-oxidosqualeno por triterpeno-ciclases. **10**
- Figura 2** - Via biossintética de diferentes triterpenos a partir da molécula de 2,3-oxidosqualeno com os diferentes cátions intermediários e seus produtos. **12**
- Figura 3** - Sequência de aminoácidos de friedelina sintase de *Maytenus ilicifolia*. **14**
- Figura 4** - Efeitos de mutações na estrutura e função da proteína β -amirina sintase de *Avena strigosa* (*sad1*). **16**
- Figura 5** - Análise do *docking* de lanosterol e friedelina no sítio ativo da enzima friedelina sintase e os aminoácidos considerados flexíveis. **20**
- Figura 6** - Resultados da predição da estrutura molecular de friedelina sintase. **29**
- Figura 7** - Definição do sítio de ligação para o *docking*. **30**
- Figura 8** - *Docking* de friedelina no sítio ativo da enzima friedelina sintase predita evidenciando os resíduos de aminoácidos a serem mutados. **31**
- Figura 9** - Alinhamento global múltiplo entre a sequência selvagem de MiFRS e as sequências mutadas obtidas. **32**
- Figura 10** - Cromatograma da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência de MiFRS **33**
- Figura 11** - Metabólitos gerados pela produção heteróloga do selvagem e do plasmídeo sem o gene. **34**
- Figura 12** - Cromatogramas da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência selvagem de MiFRS e os mutantes que perderam a função de produção de friedelina: W417H e D484E. **35**
- Figura 13** - Análise *in silico* por modelagem molecular de friedelina sintase e seus mutantes com perda de função: W417H e D484E. **37**
- Figura 14** - Cromatogramas da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência selvagem de MiFRS e os mutantes W612F e C369A. **38**
- Figura 15** - Análise *in silico* por modelagem molecular de friedelina sintase e seus mutantes que favorecem a manutenção da atividade de friedelina sintase. **39**
- Figura 16** - Cromatograma da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência de MiFRS e mutantes com ganho de função. **41**
- Figura 17** - Análise *in silico* por modelagem molecular de friedelina sintase e o mutante com manutenção da atividade de friedelina sintase e o ganho de função: F183L. **43**

LISTA DE TABELAS

- Tabela 1** - Lista com os *primers forward* e *reverse* utilizados na PCR de inserção da cauda de histidina **21**
- Tabela 2** - Lista com os *primers forward* e *reverse* para cada reação de mutação sítio-dirigida. **24**
- Tabela 3** - Lista com os *primers forward* usados no sequenciamento para cada reação de mutação sítio-dirigida. **25**
- Tabela 4** - Resíduos de aminoácidos do sítio ativo da friedelina sintase selecionados para mutagênese sítio dirigida **31**

ABREVIATURAS

BLASTp: *Basic Local Alignment Search Tool of Proteins*

BLOSUM: *Blocks Substitution Matrix*

DCTAE: motivo conservado e catalítico da enzima, Asp-Cys-Thr-Ala-Glu

DNA: ácido desoxirribonucléico

DOPE: *Discrete Optimized Protein Energy*

EDTA: *Ethylenediamine tetraacetic acid*

LB: *Luria Bertani*

OSC (s): oxidoesqualeno ciclases

PBS: tampão fosfato salino

PCR: reação em cadeia da polimerase

PDB: *Protein Data Bank*

PEG: polietilenoglicol

Ta: temperatura de *anelamento*

YNB: *yeast nitrogen base without amino acids*

YPD: *yeast extract peptone dextrose*

SUMÁRIO

1 INTRODUÇÃO	9
1.1 Triterpenos	9
1.2 Estudo mutagênico da biossíntese dos triterpenos	14
2 OBJETIVO	18
3 MATERIAIS E MÉTODOS	19
3.1 Determinação dos resíduos a serem mutados	19
3.2 Adição de cauda de histidina no gene da friedelina sintase	20
3.3 Transformação em <i>Escherichia coli</i>	21
3.4 Sequenciamento e confirmação da inserção de 6x-His	22
3.5 Reações de mutagênese sítio-dirigida	23
3.6 Transformação dos mutantes para levedura	25
3.7 Expressão heteróloga dos mutantes de friedelina sintase em <i>Saccharomyces cerevisiae</i>	26
3.8 Extração dos produtos gerados heterologamente em <i>S. cerevisiae</i>	26
3.9 Análise dos produtos gerados	27
4 RESULTADOS E DISCUSSÕES	28
4.1 Determinação dos resíduos a serem mutados	28
4.2 Obtenção dos mutantes do sítio ativo de <i>MiFRS</i>	32
4.3 Análise dos produtos triterpênicos gerados heterologamente	32
<u>4.3.1 Mutantes com perda de função</u>	34
<u>4.3.2 Mutantes com manutenção da atividade da friedelina sintase</u>	38
<u>4.3.3 Mutante com ganho de função</u>	40
5 CONCLUSÕES	45
REFERÊNCIAS	46
APÊNDICE A	50
APÊNDICE B	51
APÊNDICE C	66
ANEXO A	78

1 INTRODUÇÃO

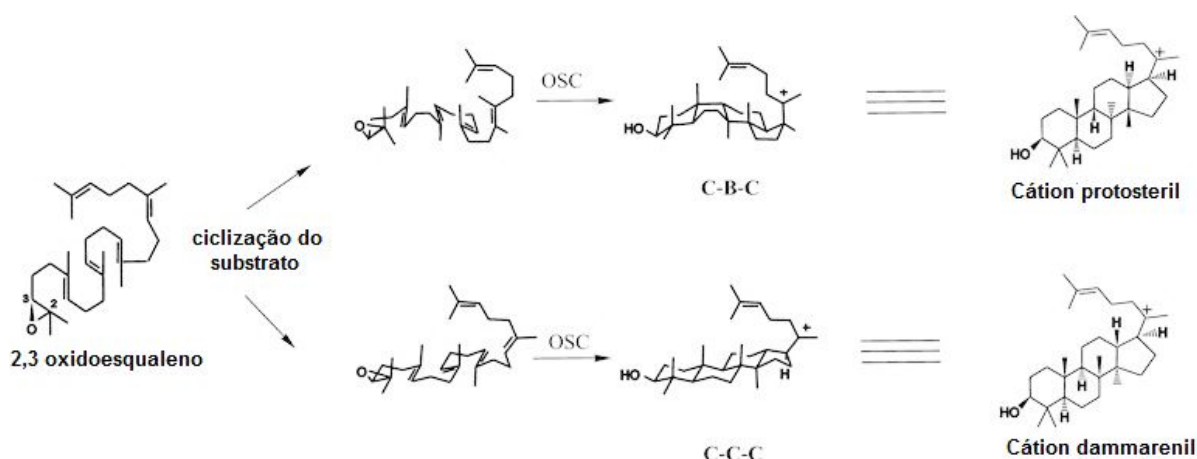
1.1 Triterpenos

Os triterpenos formam um dos grupos mais numerosos e diversificados de produtos naturais de plantas, sendo também produzidos em outros organismos como bactérias, fungos e mamíferos. Os triterpenos simples e conjugados têm uma ampla gama de aplicações pelo homem, sendo as principais no setor de alimentos, de saúde e da indústria biotecnológica (THIMMAPPA et al., 2014). Entre as atividades biológicas dos triterpenos, podem-se destacar anti-inflamatória (ANDRE et al., 2012), anticâncer (SALVADOR et al., 2012) e antiplasmodial (BERO; FREDERICH, QUETIN-LECLERCQ, 2009).

A via biossintética dos triterpenoides é iniciada a partir da união de seis unidades de isopreno para formar o esqualeno. Em procariotos, o esqualeno é diretamente ciclizado a triterpenos hopanoides, enquanto que em eucariotos, ele é primeiramente convertido a 2,3-oxidoesqualeno e, em seguida, ciclizado. As ciclizações são altamente regio-estereo-específicas e realizadas pelas enzimas oxidoesqualeno ciclases (OSCs) (WANG et al., 2010).

O passo inicial de dobramento do substrato é um passo crítico, pois o tipo de conformação do substrato leva à formação de vias de diferentes produtos. Por exemplo, a conformação cadeira-barco-cadeira leva a um cátion protosteril intermediário, o qual dá origem a triterpenos esteroidais (GASPASCUAL et al., 2014). Já a conformação do tipo cadeira-cadeira-cadeira leva a ciclização ao carbocátion damarenil, dando origem aos diferentes esqueletos triterpênicos (XUE et al., 2012), como mostrado na Figura 1.

Figura 1 - Cátions intermediários e produtos da ciclização de 2,3-oxidosqualeno por oxidosqualeno ciclases.



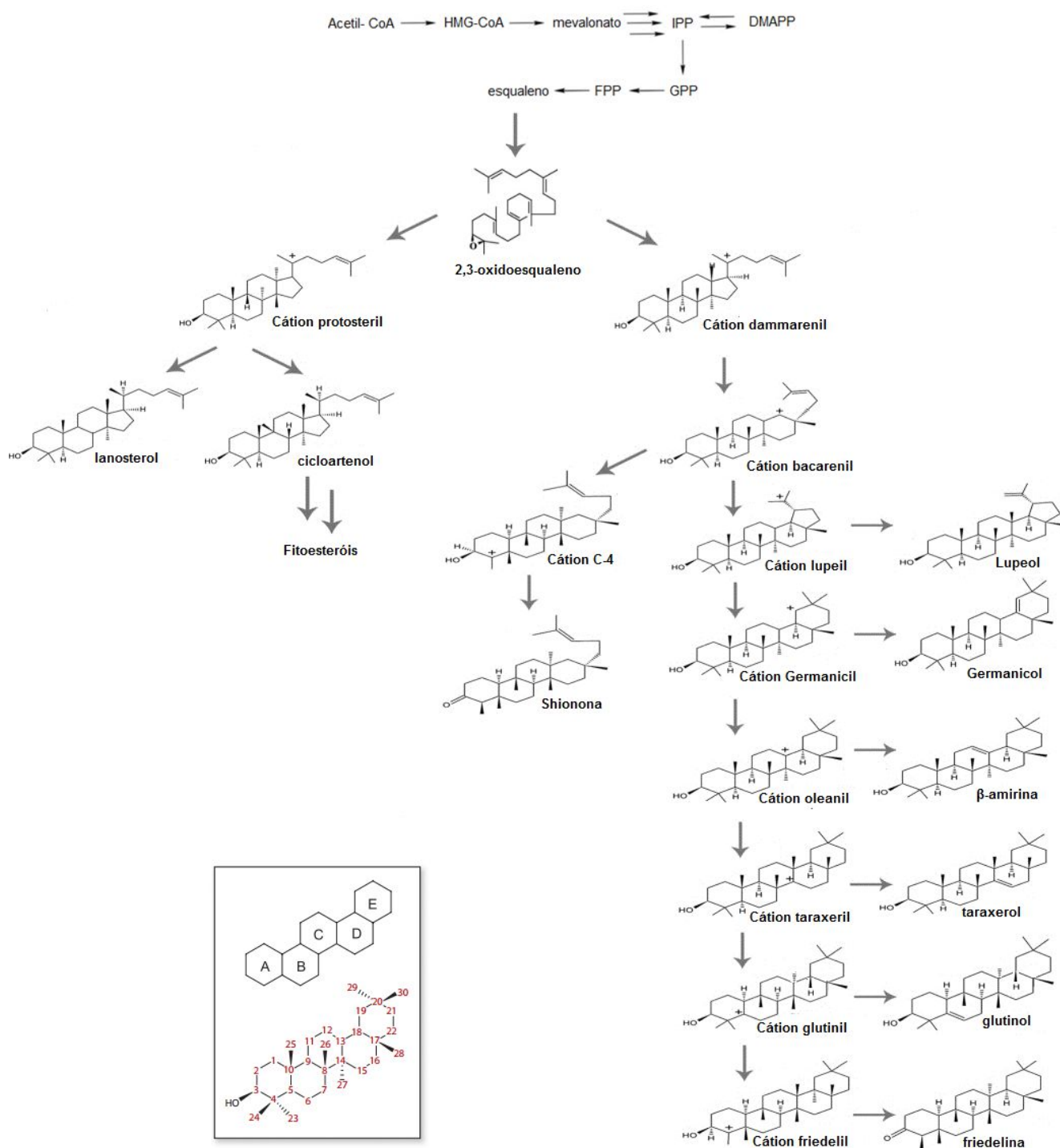
O dobramento do substrato dirige a ciclização do 2,3-oxidosqualeno. O substrato 2,3-oxidosqualeno adota padrões de dobramento distintos que, quando diretamente ciclizados por enzimas oxidosqualeno-ciclase (OSC), produzem produtos de cátions estereoquimicamente distintos (XUE et al., 2012).

Desta forma, o início da reação de ciclização das enzimas oxidosqualeno ciclase, ocorre pela protonação do epóxido de 2,3 oxidosqualeno. Em seguida ocorre o dobramento do substrato podendo este, assumir a conformação cadeia-barco-cadeira levando a formação do cátion protosteril, ou cadeira-cadeira-cadeira levando a formação do cátion damarenil, como mostrado na figura 1. Após a formação dos cátions, ocorre uma série de ataques eletrofilicos por ligações duplas próximas, resultando numa cascata de reações para formações de anel C-C. Seguindo-se a via do cátion protosteril produz-se os esteróis, que ocorre através da formação de diferentes cátions intermediários (C2->C6->C10->C14->C20), resultando no cátion intermediário C-20. Assim ocorre uma série adicional de migrações distintas de metilas e hidretos, conduzindo à formação de lanosterol (fungos e animais) pela lanosterol sintase ou cicloartenol (plantas) pela cicloartenol sintase (figura 2). Porém, a via do cátion protosteril pode também produzir cucurbitadienol, através da cucurbitadienol sintase ou de outros metabólitos considerados especializados porem não ligados a síntese de esteróis primários.

A via do cátion damarenil é responsável pela mais diversa produção de triterpenos pentacíclicos, através da ciclização inicial do cátion damarenil C-20, em seqüência rearranjos de diferentes cátions intermediários (C13->C14->C8->C9->C10->C5->C4), expansão do anel e migração de metilas e hidretos. As enzimas

oxidosequaleno ciclases que seguem na via do cátion damarenil produzem os diferentes triterpenos pentacíclicos, por exemplo: a produção de lupeol pela lupeol sintase antes da expansão do anel E; a produção de β -amirina pela β -amirina sintase contendo no anel A uma hidroxila; e, a produção de friedelina pela friedelina sintase contendo no anel A uma cetona. Importante notar-se que nesta via ocorre também a produção da shionona, único triterpeno tetracíclico com uma cetona no anel A, produzido pela shionona sintase, através do cátion damarenil seguido dos cátion intermediários cátion bacarenil (C18) e cátion C4 (C4) (figura 2).

Figura 2 - Via biossintética de diferentes triterpenos a partir da molécula de 2,3-oxidoesqualeno com os diferentes cátions intermediários e seus produtos.



Acetil-CoA: acetil coenzima A; HMG-CoA = 3-hidroxi-3-metil-glutaril-coenzima A redutase; IPP = isopentenila; DMAPP = pirofosfato de dimetilalila; GPP = pirofosfato de geranila; FPP = pirofosfato de farnesila. A numeração dos átomos de carbono nos triterpenos pentacíclicos e o sistema de

referência para os anéis (A-E) são mostrados na parte inferior esquerda da figura. Adaptado de SOUZA-MOREIRA et al. (2016).

A friedelina sintase tem capacidade de produzir a friedelina através da realização do número máximo de rearranjos, com formação de uma cetona pela desprotonação do intermediário hidroxilado pelo resíduo de aspartato que iniciou a ciclização, sem auxílio de uma enzima oxidorreductase (ABE; PRESTWICH, 1995; WANG et al., 2010). A carga positiva da posição C-20 para a posição C-2 envolve o número máximo possível de troca 1,2 (10 no total). Quando o cátion chega a posição C-2 é atacado pelo grupo 3β -OH para formar uma cetona em C-3, formando assim o triterpeno pentacíclico friedelina. Friedelina também é precursora dos quinonametídeos triterpênicos maitenina e pristimerina (CORSINO et al., 2000), para os quais foram descritas diversas atividades biológicas como anti-inflamatória, anticâncer, antimicrobiana, antimalárica, espermicida e antioxidante (DEEB et al., 2015; HE et al., 2016; MOORE; RUBEN; ROSEN, 1993; SANTOS, V et al., 2010).

Assim, a enzima friedelina sintase clonada de *Maytenus ilicifolia* (Genbank KX147270), também denominada como (3S)-2,3-epoxi-2,3-dihidroesqualeno mutase (EC 5.4.99.50) contém em sua seqüência os domínios conservados domínio da esqualeno ciclase subgrupo 1 (SQCY_1) e domínio da redutase de isopreno-C2 (ISOPREN_C2), os quais estão presentes em triterpeno sintase de classe II, que inclui as OSC's. Há também a presença do *motif* Asp-Cys-Thr-Ala-Glu (DCTAE), que contém o resíduo de ácido aspártico catalítico e os quatro *motif* QW (motivos ricos em aminoácidos aromáticos iniciando com Q-Gln e terminando em W-Trp), importantes na manutenção da estrutura enzimática. A presença dos resíduos característicos de Ser-Phe (SF) e o *motif* Met-X-Cys-Arg (MXCYCR) conservado em triterpenos sintase de triterpeno pentacíclicos nas posições previstas na figura 3, estão de acordo com a sua classificação funcional (SOUZA-MOREIRA et al., 2016).

Figura 3 - Sequência de aminoácidos de friedelina sintase de *Maytenus ilicifolia*.

MiFRS	1	MWKI KI ADRGN- CPYNEYLYTTNDFVGRQLWEFDPNSTGTPEELAELEEARRKFTENRYEV
MiCAS1		MWRLKF GAETI GDDGGSWLRSMNGHVGRVWEFCPELGSPEDESLVENARRGFTINTHLEK
MiFRS	60	KPASDLLWMQFLRKNNFK--QTIPPLRI GEKECVTYEDVTTALRRASSFFSALQASDGH
MiCAS1	61	KHSADLLMRTQFAKENPY--FVNLPQIKVKASEDVTDEVVTTTLRRAINFYSTIQADDGH
MiFRS	118	WPAENAGVSEFLPPFIFCLYITGHLNSITPEHRKELLRFIYNHNE DGGWGLHI EGHST
MiCAS1	119	WPGDYGGPWFLLPGLITTLHI T GALNAVLSREHOREMCRYLYNHNS DGGWGLHI EGPST
MiFRS	178	VFATAF T YVCMRI LGVGP D--EDACARARKWLD RGGI TYMASWGKTF SVLGLFDWYGC
MiCAS1	179	MFGTALNYVTLRLLGEGTEGGEGAI ENGRKWL DHGGATATTSWGKMWLSVLGVYEWIGN
MiFRS	236	NPMPPEFWL LPSYLP IHPAKMVCYCRMVYMPMSYLYGKRFVAPI TPLI LQLREELHTOPY
MiCAS1	239	NPLPPEVWLC P YLLPMPHRMCHCRMVYLPMSYLYGKRFVGPITATVLSLRKELYTPY
	
MiFRS	296	HEIIEWRKMRHRC AEEDLYEPHSLIQNF L WDSLYVAS EPLLTRWPF S KIRERALEKAMEHI
MiCAS1	299	HEVDWANKARNT CAKEDLYPHPLVQDVLWS SLYYAYEPI FMRWPARRLREKALQTVMOHI
MiFRS	356	HYEDENSRYITIGCVEKALCIVLCCWVEDPNGEYFKKHLARIPDYLWVAEDGMKVVSF- GS
MiCAS1	359	HYEDENTRYIGI GPNVKVLE NMLCCWVEDPQSEAFRLHI PRVYDYLWLAEDGMKVOGYNGS
MiFRS	415	QLWDATFGFQALVASNLITDEVAPITLVKAYDFIKKCOVRDNPSCNFEKMFRIH S KGSWTF S
MiCAS1	419	QLWDSAF AVQAI IISTDLVEEYGLCLKRAHINYLKDSOVLEDCPGDLSFWYRIH S KGAWPF S
MiFRS	475	DQDHGWQLS DCTAEALKCCLLAATMPEELVGEKLD P QW FESVNIIL S LQEPK TGGIAGW
MiCAS1	479	TADHWPT S DCTAEGLKAALLLSTFPSEIVGEPLDAERVYDAVNI L S LQNA- DGGFATY
		▲
MiFRS	535	EPVRIAGOWMEV LNPMEFLENIVIEHTYIECTGSSITLAFITLKKLFPGHRTKDI DNFI VNA
MiCAS1	538	ELTRSYQWLEF L NPAETFGDI VI DYPYVECTSAI QALTLF KRFYPTHRE EVDNGLTKA
MiFRS	595	I RYLEDEECPYDGSWYGNWGLCFI YSITMFALGGLAAGRTYKNCQAVRRGVDFLLI NOSDD
MiCAS1	598	GEFIERIQEENGSWYGSWGVCFSS YGGWFGIRGLVAAGKTY SNGHSLRKACAYLLS KELLRT
MiFRS	655	GGWGESYI S C PRKKYTPLEGRRSNVVQTAWAMLGLLYAGQAE RDPTPLHRGAKLLI NYQM
MiCAS1	658	GGWGESYLS CONKVVYTNLKGDRPHIVNTGWAMLALIEAGQAE RDPTPLHRGAKLLVNSQM
MiFRS	715	E EGGYPQOEITGVFKMNCMLHYPI YRNAFPI WALGEYRKRVP LPSKGN S WAMKINS A
MiCAS1	718	ENGDFPQGEI MGVENKNCMLSYSA YRNIFPI WALGVYRCRVLQSL-----

Quatro *motif* QW e o *motif* DCTAE estão sublinhados, indicando os ácidos envolvidos na formação do carbocátion do substrato. Outros resíduos conservados em OSC's são: MXCYCR (sublinhado por uma linha tracejada) e SF (sublinhado por duas linhas). O alinhamento global ilustra as duas OSC's clonadas de *M.ilicifolia*, sendo *MiCAS1*, uma cicloartenol sintase. (SOUZA-MOREIRA et al., 2016).

Isolar OSC's de forma que permaneçam ativas e caracterizar suas propriedades enzimáticas *in vitro* tem sido de extrema dificuldade. Assim, análises funcionais envolvendo experimentos com mutantes tem sido conduzidos *in vivo*, porém as funções dos sítios ativos das OSC's ainda não foram claramente elucidadas (HOSHINO, 2017).

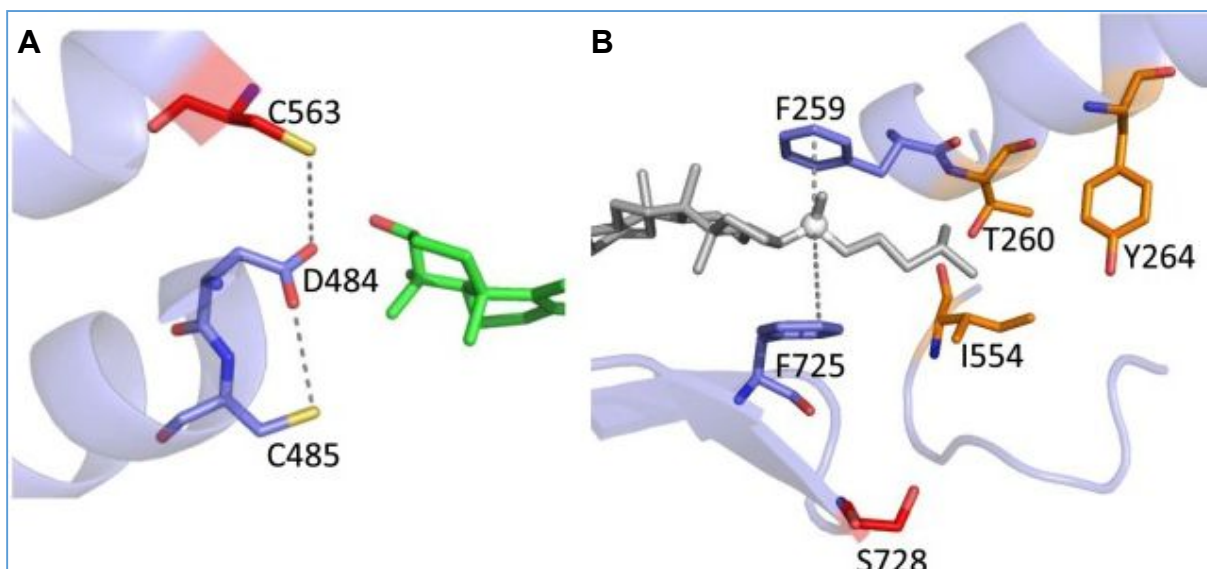
1.2 Estudo mutagênico da biossíntese dos triterpenos

O entendimento da especificidade das oxidoesqualeno ciclasas em formar os diferentes esteróides e triterpenos vem sendo realizado por estudos de mutagênese e com o uso de análogos do substrato, avaliando-se os rearranjos na ciclização que os resíduos de aminoácidos proporcionam, como demonstrado no estudo recente de

Hoshino (2017), em que foi realizado o mecanismo catalítico e o mecanismo de reconhecimento do substrato de β -amirina sintase, revelados por mutagenese sítio-dirigida e experimentos com substratos análogos, e que serão comentados na seção de Resultados e Discussão desta tese.

Um estudo recente (SALMON et al., 2016) realizado com mutantes de β -amirina sintase de *Avena strigosa* (SAD1) identificou resíduos de aminoácidos conservados envolvidos com a especificidade do produto e do substrato de triterpeno sintases de diversas plantas. Análise dos triterpenos produzidos pelo mutante Cys563Tyr de SAD1 na planta e utilizando o modelo de expressão heteróloga em *Saccharomyces cerevisiae* permitiu observar que houve bloqueio do início da ciclização (Figura 4A), sem formação de triterpenos e, que a mutação Ser728Phe de SAD1 favoreceu a produção dos triterpenos tetracíclicos epoxidamarano e damaranediol ao invés de pentacíclicos. Por meio do estudo de modelagem molecular de SAD1, os autores observaram que o resíduo de cisteína na posição 563 é importante por formar uma ligação de hidrogênio com o aspartato da posição 484 favorecendo a protonação do grupo epóxido do oxidoesqualeno, levando à perda de atividade pela enzima, enquanto que a troca de serina na posição 728 por fenilalanina promoveu a estabilização do carbocátion damarenil, produzindo os triterpenos truncados (Figura 4B). Esse estudo também fez a confirmação do envolvimento do resíduo de serina correspondente em lupeol sintase de *Arabidopsis thaliana* pela mutação Ser728Phe, além de ter demonstrado que estes mutantes têm preferência pelo substrato dioxidoesqualeno ao invés de 2,3-oxidoesqualeno, desenvolvendo assim uma nova oportunidade de síntese de novos triterpenos a partir de diferentes substratos (SALMON et al., 2016).

Figura 4 - Efeitos de mutações na estrutura e função da proteína β -amirina sintase de *Avena strigosa* (*sad1*).



A) O mutante *sad1* 358 tem uma mutação em Cys563 (vermelho), um resíduo que está ligado através do hidrogênio do aspartato catalítico (D484). Ligações de hidrogênio são mostradas por linhas tracejadas. O substrato é apresentado em verde.; B) mutantes de *sad1* 384 e 1023 têm uma mutação em Ser728 (vermelho), que está na proximidade de resíduos envolvidos no substrato (laranja) e Phe725 (roxo), envolvidos na estabilização do cátion intermediário tetracíclico em C-20. As interações de cátion- π são mostradas por linhas tracejadas em cinza. Adaptado de SALMON et al. (2016).

Recentemente (SOUZA-MOREIRA et al., 2016), nosso grupo avaliou mutantes de leucina na posição 482 de friedelina sintase de *Maytenus ilicifolia*, realizando a troca por um aminoácido de cada classe. Nas duas enzimas friedelina sintases descritas até o momento, (uma de *Kalanchoe daigremontiana*, *KdFRS* e uma de *M. ilicifolia*, *MiFRS*) o resíduo de leucina nesta posição é único em comparação com as demais OSCs e se encontra ao lado do motivo conservado e catalítico da enzima, Asp-Cys-Thr-Ala-Glu (DCTAE). Foi observado que a troca de Leu482 por Thr resultou na produção de β -amirina, enquanto que ao realizar a troca por Val, houve a produção tanto de friedelina quanto de β -amirina, sendo que a troca por Ile não interferiu na produção de friedelina. A partir dos produtos formados, modelagem molecular e *docking* do cátion oleanil, precursor comum de β -amirina e de friedelina foi possível descrever que o papel da leucina e a conformação que confere ao sítio ativo durante a ciclização são importantes para a continuidade dos rearranjos do carbocátion, possibilitando a formação do composto com mais rearranjos, a

friedelina, ao invés de estabilizar previamente o carbocátion na formação de β -amirina. Por outro lado, a troca do resíduo de leucina por valina, este conservado em β -amirina e lupeol sintases, além da produção de β -amirina, manteve a produção de friedelina, indicando que a troca apenas deste resíduo não foi o ponto divergente entre a especificidade das enzimas.

Estudos com outras oxidoesqualeno ciclases também já demonstraram a importância de diferentes resíduos envolvidos com sua especificidade biossintética. Desta forma, sabe-se, por exemplo, que a troca do resíduo Ile481 em cicloartenol sintase de *Arabidopsis thaliana* (AthCAS1) pelo resíduo conservado de valina em lanosterol sintase levou à produção de lanosterol, sendo complementar funcionalmente à linhagem nocaute de *ERG7* de *S. cerevisiae* (MATSUDA et al., 2000); enquanto que o *motif* Met-Trp-Cys-Tyr-Cys-Arg de β -amirina sintase de *Panax ginseng* (PNY) e o Met-Leu-Cys-Tyr-Cys-Arg de lupeol sintase de *Olea europaea* (OEW) estão envolvidos na especificidade de cada enzima, levando à interconversão do composto triterpênico majoritário formado e/ou formação de mais de um composto, ou mesmo à estabilização do carbocátion com um menor número de rearranjos, formando intermediários da síntese de triterpenos pentacíclicos quando trocados entre si (KUSHIRO; SHIBUYA; EBIZUKA, 2000).

Baseado nos estudos de mutagênese com outras OSCs, este trabalho teve como objetivo propor e avaliar os resíduos de aminoácidos de friedelina sintase de *M. ilicifolia* envolvidos com a sua especificidade desse triterpeno cetônico singular.

2 OBJETIVO

Este projeto teve como objetivo realizar o estudo da especificidade da enzima friedelina sintase de *M. ilicifolia* por meio da predição da estrutura tridimensional da enzima e dos consequentes mutantes gerados por meio de mutagênese sítio-dirigida.

Para tanto, o trabalho foi realizado de acordo com as seguintes etapas:

- Determinação de resíduos a serem mutados;
- Obtenção dos mutantes de friedelina sintase;
- Expressão e produção dos mutantes;
- Avaliação da especificidade da friedelina sintase.

3 MATERIAL E MÉTODOS

3.1 Determinação dos resíduos a serem mutados

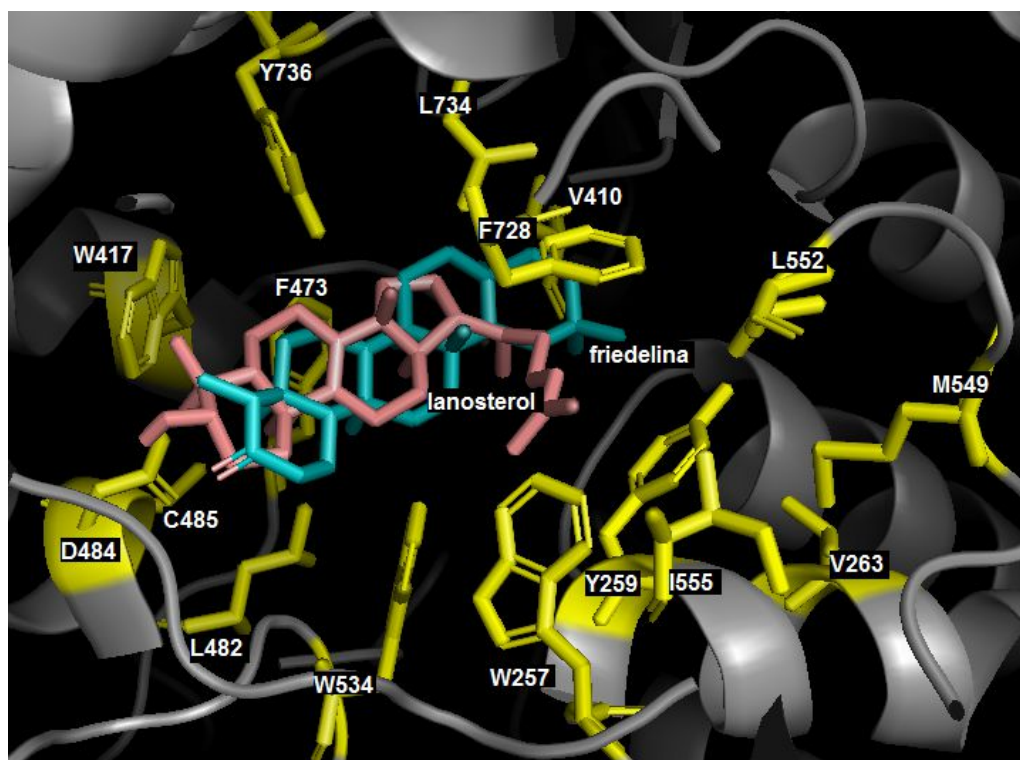
O estudo da especificidade da via biossintética da friedelina sintase de *M. ilicifolia* foi realizado a partir da análise dos resíduos de aminoácidos que constituem o sítio catalítico da enzima utilizando-se, para tanto, a estrutura primária, secundária e terciária desta enzima. Uma vez que ainda não foi obtido o cristal e a estrutura tridimensional desta enzima, a mesma foi obtida por modelagem molecular por homologia.

A predição da estrutura molecular da enzima friedelina sintase clonada das folhas de *M. ilicifolia* (número de acesso *Genbank* KX147270) foi realizada com base em modelagem molecular por homologia com a estrutura cristalográfica de lanosterol sintase de humano (*Protein Data Bank* ID 1W6K). O alinhamento das sequências de resíduos de aminoácidos foi feito a partir da ferramenta align2D do software Modeller. A construção das coordenadas do modelo estrutural foi realizada utilizando-se os algoritmos do software Modeller empregando a ferramenta ViTAMIn, com os seguintes parâmetros: 50 modelos; 500 ciclos de otimização de modelo; refinamento lento; e perturbação aleatória das coordenadas do modelo inicial durante os ciclos de refinamento. O melhor modelo foi selecionado de acordo com a análise dos resíduos de aminoácidos por gráficos de Ramachandran e potencial DOPE (*Discrete Optimized Protein Energy*). A última etapa foi a realização do docking das moléculas de friedelina e lanosterol no sítio ativo da enzima friedelina sintase.

O docking foi realizado usando-se o algoritmo genético lamarckiano (LGA) através do AutoDock 4.2 e o AutoDock Vina. Também foram definidos como flexíveis no centro do cubo a 4 Å de distância, os aminoácidos: Trp257, Tyr259, Val263, Val410, Trp417, Phe473, Leu482, Asp484, Cys485, Trp534, Met549, Leu552, Ile555, Phe278, Leu734 e Tyr736) (figura 5). As estruturas 3D dos cátions intermediários e da friedelina foram geradas usando parâmetros geométricos padrões disponíveis em MarvinSketch (Marvin 162.22), 2016, ChemAxon (<http://www.chemaxon.com>). A melhor conformação dos cátions ligantes foi energeticamente minimizada utilizando os parâmetros pré-definidos da MarvinSketch e analisados com auxílio do programa PyMOL™ Molecular Graphics System (versão 1.7.2.1). Esta etapa foi realizada em

colaboração com o Professor Dr. Rafael Guido e com o aluno de doutorado Gustavo M. A. de Lima, do Instituto de Física da USP de São Carlos.

Figura 5 - Análise do *docking* de lanosterol e friedelina no sítio ativo da enzima friedelina sintase e os aminoácidos considerados flexíveis.



Os resíduos considerados flexíveis com distância de 4 Å no sítio ativo da friedelina sintase (amarelo), conforme o *docking* dentro do sítio ativo definido através de um cubo de aresta de 12 Å. A molécula de lanosterol (rosa) e de friedelina (ciano) dockados no sítio ativo utilizando-se os parâmetros padrões de disponíveis em MarvinSketch e, visualizados através do PyMOL™ Molecular Graphics System (versão 1.7.2.1). (Autora)

3.2 Adição de cauda de histidina no gene da friedelina sintase

Para expressão constitutiva da sequência de friedelina sintase, foi utilizado o plasmídeo pSP-GM1 (PARTOW et al., 2010). Na posição C-terminal da sequência codificadora de friedelina sintase foi inserida uma cauda de seis histidinas (6x-His) por meio de reação em cadeia da polimerase (PCR) (LIU; NIUSMITH, 2008). Essa reação foi realizada com a enzima Phusion® High-Fidelity DNA Polymerase (New England Biolabs) utilizando os seguintes reagentes: 1,0 µL de DNA plasmidial (20 ng/ µL); 5,0 µL de tampão para Phusion High-Fidelity (5X; New England Biolabs); 0,2 µL MgCl₂ (25 mM); 0,6 µL dNTPs (10 mM); 0,8 µM de cada *primer* (Tabela 1); 0,5 µL da enzima Phusion® High-Fidelity DNA Polimerase (2 U/µL); e água q.s.p. volume

final de 25 µL. A reação foi conduzida nas seguintes condições: *hotstart* de 95 °C por 5 min, seguido de 18 ciclos de desnaturação a 95 °C por 1 min, anelamento a temperatura de 56 °C por 5 min e polimerização a 72 °C por 17 min e um ciclo final de polimerização a 45 °C por 1 min e 72 °C por 30 min.

Tabela 1 - Lista com os *primers forward* e *reverse* utilizados na PCR de inserção da cauda de histidina.

<i>Primer</i>	Sequência 5'-3'	T _m (°C)
MiFRSaddHisf (VZO2375)	GCTCATCACCATCACCATCACTGAGAGCTC TTAATTAACAATTCTTCGCCAGAGGTTTGGT CAAG	68,8
MiFRSaddHisr (VZO2376)	CTCTCAGTGATGGTGATGGTGATGAGCACT GTTTATCTTCATAGCCATACTGTTACCCTTA GAAGG	68

A região sublinhada indica a região codificadora das seis histidinas adicionadas na sequência codificadora de friedelina sintase. As letras f e r nos *primers* correspondem a *forward* e *reverse*, respectivamente. (REMLINGER M, 2016. Esta tabela foi elaborada pela autora, assim como todas as tabelas deste trabalho).

A reação de PCR foi submetida a digestão em uma reação contendo 1 µL da enzima *DpnI* (20 U/µL; New England Biolabs), 2,5 µL de tampão CutSmart (New England Biolabs) e água q.s.p volume final de 25 µL, a qual foi incubada *overnight* a 37 °C. Posteriormente, o produto reacional foi transformado em bactéria competente, de acordo com o protocolo descrito abaixo.

3.3 Transformação em *Escherichia coli*

Após a digestão, o produto plasmidial foi transformado em *E. coli* competente (DH5α) pelo método de choque térmico. Assim, preparou-se uma solução de 10 mL de Transfobuffer, ou seja, 1,0 mL de KCM 10X (1 M KCl; 0,3 M CaCl₂; 0,5 M MgCl₂), 1,5 mL de polietilenoglicol 3350 (PEG) 10% e 7,5 mL de água esterilizada. Desta solução, foram utilizados 80 µL e foram adicionados 20 µL do produto plasmidial em um tubo de microcentrifuga de 1,5 mL, ao qual foram homogeneizados 100 µL da suspensão de *E. coli* competente. A mistura foi incubada por 30 min em um banho de gelo. Em seguida, o tubo foi colocado em banho-maria a 42 °C por 2 min. Logo

depois, foi adicionado 1 mL de meio Luria-Bertani (LB) ao tubo, que foi incubado por 1 h a 37 °C com agitação de 200 rpm. Ao final da incubação, a transformação foi centrifugada a velocidade máxima por 1 min, o sobrenadante foi descartado, o pellet foi ressuspenso em 200 µL de água esterilizada e foi plaqueado em meio de cultura sólido LB contendo ampicilina (100 µg/mL), o qual foi e incubado a 37 °C por 12-16 h.

3.4 Sequenciamento e confirmação da inserção de 6x-His

Após o crescimento das colônias, cinco foram retiradas para análise da inserção da cauda à região C-terminal da sequência de friedelina sintase. Foram feitos inóculos de cada colônia em 3 mL de meio líquido LB contendo ampicilina (100 µg/mL) e estes foram inoculados *overnight* a 37 °C, sob agitação constante. O DNA plasmidial foi obtido utilizando o kit Five-Minute Plasmid Miniprep (Sigma-Aldrich), homogeneizando 40 µL do reagente de lise às culturas. A mistura foi deixada em descanso por 2 min enquanto a coluna de purificação foi preparada por lavagem com 500 µL da solução de preparo da coluna. À mistura contendo o inóculo foram adicionados 400 µL da solução de ligação e inverteu-se o tubo por 15 vezes. Todo o volume foi então colocado na coluna e centrifugado por 30 s, a velocidade máxima. O líquido que passou pela coluna foi descartado e foram adicionados 700 µL de solução de lavagem, centrifugando-se por 20 s, a velocidade máxima. Novamente, descartou-se o líquido que passou pela coluna e foram adicionados, 200 µL da solução de lavagem centrifugando-se a coluna por 30 s, a velocidade máxima. A coluna foi colocada em um tubo novo e foram adicionados e centrifugados 40 µL da solução de eluição. A quantificação do DNA plasmidial eluído foi realizada no espectrofotômetro Nanodrop 2000 (Thermo Scientific), observando-se a razão de qualidade entre os comprimentos de onda 260/280 entre 1,8 a 2,0.

O DNA plasmidial (400 ng) obtido das diferentes colônias foi então empregado na reação de PCR de sequenciamento utilizando o kit Big Dye[®] Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), juntamente com o *primer* VZO2374 (a 3,2 µM) 5'- GTTGGGTTTGTATACGCCG -3' e as condições de amplificação: *hot start* de 96° C por 1 min, seguida de 25 ciclos de 96° C por 10 s, 43° C por 5 s, 60° C por 4 min e extensão final de 60° C por 5 min. Em seguida à amplificação, as amostras foram purificadas com o kit Big Dye[®]X Terminator™ Purification Kit (Applied Biosystems) seguindo o protocolo do fabricante. O sequenciamento automático foi

realizado com 20 μL de cada amostra no sequenciador GeneticAnalyzer 3130 (Applied Biosystems) e a confirmação da inserção de 6xHis foi observada pela presença da sequência CATCACCATCACCATCAC após o último resíduo de friedelina sintase.

3.5 Reações de mutagênese sítio-dirigida

O plasmídeo contendo a sequência codificadora de friedelina sintase com adição de uma cauda de seis histidinas na posição C-terminal [P_{TEF1} -MIFRS-6xHis, P_{PGK1} -tHMG1], foi preparado com o kit Five-Minute Plasmid Miniprep, da mesma forma como descrito no item anterior. A PCR para troca dos resíduos de interesse foi realizada com a enzima Phusion[®] High-Fidelity DNA Polymerase, sendo conduzida, separadamente, uma reação de polimerização com o *primer forward* e outra com o *reverse* (Tabela 2), contendo os seguintes reagentes: 1,0 μL de DNA plasmidial (500 ng/ μL); 5,0 μL de tampão (5X); 0,25 μL MgCl_2 (25 mM); 0,75 μL dNTPs (10 mM); 0,8 μM de *primer*; 0,3 μL da enzima polimerase (2 U/ μL); e água q.s.p. volume final de 25 μL . A reação foi conduzida nas seguintes condições: *hotstart* de 98 °C por 1 min, seguido de 30 ciclos de desnaturação a 98 °C por 10 s, anelamento a temperatura com subtração de 15 °C da temperatura de anelamento de cada *primer* por 30 s e polimerização a 72 °C por 5 min e um ciclo final de polimerização a 72 °C por 5 min. Após a PCR, os 20 μL do produto de cada reação sintetizados separadamente foram reunidos em um mesmo microtubo, elevando a temperatura da solução a 98 °C e depois, o anelamento entre as fitas foi realizado pela redução gradativa de 1 °C/ min até 90 °C e depois, redução de 10 °C/min até 37 °C (EDELHEIT et al., 2009).

Tabela 2 - Lista com os *primers forward* e *reverse* para cada reação de mutação sítio-dirigida.

Primer	Sequência 5'-3'	Ta (°C)
F183Lf (VZO2347)	CTCAACTGTCTTTGCCACCGCTTT <u>GACTT</u> ACGTCTGTAT GAGAATTTTG	66
F183Lr (VZO2348)	CAAATTCTCATACAGACGTAAGT <u>CAAAG</u> CGGTGGCAA AGACAGTTGAG	66
C369Af (VZO2351)	CAGTAGATACATCACAATCGGT <u>GCT</u> GTTGAAAAGCTTT ATGCATG	64,1
C369Ar (VZO2352)	CATGCATAAAGCTTTTTCAAC <u>AGC</u> ACCGATTGTGATGTA TCTACTG	64,1
W417Hf (VZO2353)	CTCTTTCGGTTCACAATTG <u>CAT</u> GATGCTACTTTTGGTTT CC	64
W417Hr (VZO2354)	GGAAACCAAAGTAGCAT <u>CATG</u> CAATTGTGAACCGAAA GAG	64
D484Ef (VZO2357)	CGGTTGGCAATTAAGT <u>GAA</u> TGCACAGCAGAAGCCTTG	66
D484Er (VZO2358)	CAAGGCTTCTGCTGTGCAT <u>TCA</u> CTTAATTGCCAACCG	66
W612Ff (VZO2365)	GTTCTTGGTACGGTAACTTTGGTATCTGTTTCATATAC	59,5
W612Fr (VZO2366)	GTATATGAAACAGATACCAAAGTTACCGTACCAAGAAC	59,5

A região sublinhada indica as bases do respectivo resíduo de aminoácido a ser mutado. As letras f e r nos *primers* correspondem a *forward* e *reverse*, respectivamente.

O DNA plasmidial molde de cada mutação foi digerido com 1 µL da enzima *DpnI* (20 U/µL), em uma reação contendo o produto de anelamento das PCRs e 5 µL de tampão CutSmart (New England Biolabs) e a digestão foi incubada *overnight* a 37 °C. O DNA plasmidial resultante foi transformado em *E. coli* competente, como descrito no item 3.3.

Após crescimento das colônias, cinco foram utilizadas para confirmação dos mutantes obtidos de *friedelina* sintase. As colônias foram incubadas em meio LB com ampicilina, o plasmídeo foi obtido por miniprep e o sequenciamento foi realizado de acordo com o procedimento descrito no item 4. Os *primers* utilizados para a reação de sequenciamento das diferentes mutações estão apresentados na Tabela 3.

Tabela 3 - Lista com os *primers forward* usados no sequenciamento para cada reação de mutação sítio-dirigida.

Primer	Sequência 5'-3'	Mutações verificadas
F183seq (VZO2371)	CATTACCGGTCATTTGAACAG	F183L
369/484seq (VZO2372)	GGGACTCTTTGTATGTTGCATC	C369A, W417H, D484E
534/612seq (VZO2373)	GTTGCTTGTTAGCTGCAACC	W612F
MiseqR (VZO2377)	AATGTCCAACCTACCCTTGG	W417H, C369A

O sequenciamento automático foi realizado com 20 µL das amostras no sequenciador GeneticAnalyzer 3130 e a confirmação de cada mutação foi feita por alinhamento com a sequência selvagem de friedelina sintase utilizando a ferramenta de alinhamento global Clustal Omega.

3.6 Transformação dos mutantes em levedura

Após a confirmação das mutações, os plasmídeos foram transformados na linhagem *S. cerevisiae* CEN.PK113-5D (*MATa MAL2-8^c SUC2 ura3-52 P_{ERG7}::P_{KEX2}*) pelo método de acetato de lítio/polietilenoglicol (ITO H, et al. (1983). A transformação foi realizada a partir de um inóculo *overnight* da linhagem de *S. cerevisiae* CEN.PK113-5D em 3 mL de YPD (1% de extrato de levedura, 2% de peptona e 2% glicose). As células crescidas *overnight* foram recuperadas por centrifugação na velocidade máxima por 1 min, ressuspensas em 1 mL de acetato de lítio 100 mM e incubados por 15 min a 30 °C. As células foram novamente centrifugadas por 10 s na velocidade máxima e o sobrenadante foi descartado. Ao pellet formado foram homogeneizados os seguintes reagentes: 240 µL de PEG a 50%, 36 µL de acetato de lítio 1 M, 50 µL de solução de fita simples de DNA de esperma de salmão (2,0 mg/mL), 5 µL do plasmídeo com a mutação (em torno de 2 µg) e, 20 µL de água esterilizada. As suspensões foram incubadas por 30 min, sob agitação, a 30 °C. Posteriormente, foi realizado o choque térmico por 10 min a 42 °C. Ao término da incubação, a transformação foi centrifugada, o sobrenadante removido e o pellet ressuspensado em 200 µL de água esterilizada. Este volume foi então plaqueado em meio sintético completo (0,67% de base para levedura nitrogenada sem adição de aminoácidos - YNB, e 2% de glicose) suplementado com aminoácidos, bases e ácido p-aminobenzoico, sem adição de uracila (SC-URA) e colocado em estufa de crescimento a 30 °C por 2 a 4 dias.

3.7 Expressão heteróloga dos mutantes de friedelina sintase em *S. cerevisiae*

Para avaliar possíveis modificações na produção de friedelina pelos mutantes formados, a expressão da sequência codificadora de friedelina sintase e dos mutantes obtidos foi realizada na linhagem de *S. cerevisiae* transformada com os respectivos plasmídeos. Inicialmente, foram realizados pré-inóculos, separadamente, de uma colônia de cada levedura contendo o vetor vazio como controle negativo, a sequência selvagem de *MIFRS* e as sequências mutadas em 5 mL de SC-URA, sob agitação constante a 30°C, *overnight*. Posteriormente, os inóculos foram diluídos para uma densidade óptica a 600 nm inicial de 0,05 em 50 mL de meio mínimo de Delft (0,75% de sulfato de amônio, 1,44% de fosfato monobásico de potássio, 0,05% de sulfato de magnésio heptahidratado, 2% de glicose, 0,2% de solução contendo traços de metais: 1,5% de EDTA, 0,045% sulfato de zinco heptahidratado, 0,1% de cloreto de manganês II, 0,03% de cloreto de cobalto II hexahidratado, 0,03% de sulfato de cobre II pentahidratado, 0,05% de molibdato de sódio dihidratado, 0,045% de cloreto de cálcio dihidratado, 0,03% de sulfato de ferro II heptahidratado, 0,01% de ácido bórico, 0,01 % de iodeto de potássio; e, 0,1% de solução de vitaminas: 0,005% de biotina, 0,02% de ácido p-aminobenzoico, 0,1% de ácido nicotínico, 0,1% de ácido pantotênico, 0,1% de piridoxina-HCl, 0,1% de tiamina-HCl e 2,5% mio-inositol) (VERDUYN et al., 1992). A incubação foi realizada a 30 °C, por 72 h, a 200 rpm. Após este período, as células foram coletadas por centrifugação a 3000 rpm por 5 min em tubos tarados de 50 mL e lavadas com 10 mL de tampão fosfato salino (PBS). As células foram congeladas em ultrafreezer a -80 °C por 24 h e depois foram submetidas a liofilização.

3.8 Extração dos produtos gerados heterologicamente em *S. cerevisiae*

Após a liofilização, foram pesados cerca de 20 mg da massa de células secas em tubos de vidro de borosilicato (Pyrex, 16x100 mm), aos quais foram misturados 7 mL da solução clorofórmio:metanol (2:1, v/v) para extração dos triterpenos por ultrassom (Elmasonic S30H, ELMA) com duração de 30 min.

Para separação da fase orgânica contendo os triterpenos foram adicionados 1,7 mL de solução de NaCl 0,73%. A mistura foi agitada vigorosamente por 30 s e a fase apolar foi recuperada em um novo tubo de vidro após 30 min. As amostras foram então secas e armazenadas a -20°C até a análise por cromatografia.

3.9 Análise dos produtos gerados

As frações apolares extraídas das células de *S. cerevisiae* expressando heterologicamente a sequência selvagem de *MIFRS*, as sequências mutantes e o vetor vazio foram analisadas por cromatografia gasosa acoplada a espectrometria de massas (CG-EM). As amostras secas foram ressuspensas em 200 µL de acetonitrila. A análise cromatográfica foi realizada em um cromatógrafo a gás acoplado ao espectrômetro de massas (SHIMADZU, QP2020C W/O RP230V) com coluna HP-5 (30 m x 0,25 mm x 0,25 µm; Agilent Technologies). Temperatura do injetor: 270° C; rampa de aquecimento de 200 a 290° C (10° C/min); temperatura do trap: 200° C por 3 min; temperatura da interface: 290 °C por 18 min; volume de injeção: 1 µL, com modo de injeção do tipo Split: 1:10; fluxo do gás de arraste 1,0 mL/min; tempo total de análise de 30 min, relação massa/carga *m/z* de 35 a 600.

4 RESULTADOS E DISCUSSÃO

4.1 Determinação dos resíduos a serem mutados

Inicialmente, foi identificada a estrutura tridimensional a ser utilizada como molde para a modelagem molecular por homologia da friedelina sintase de *M. ilicifolia* (*MiFRS*). Utilizando a ferramenta Blastp, foi feita a busca de proteínas que apresentassem identidade local com a sequência de *MiFRS* no PDB, onde estão depositadas as estruturas tridimensionais das proteínas obtidas experimentalmente por cristalografia por raio-X, ressonância magnética nuclear e microscopia crioeletrônica.

O melhor alinhamento obtido da sequência primária de *MiFRS* se deu com a lanosterol sintase de humano (PDB 1W6K), que apresentou 40% de identidade. A partir da estrutura da lanosterol sintase, pôde-se montar um modelo da estrutura molecular de *MiFRS* usando o software Modeller. Este software realiza o alinhamento das estruturas secundárias da proteína molde resultante do PDB com a sequência da proteína a ser predita. O software retornou 100 modelos e estes foram avaliados conforme seu potencial DOPE.

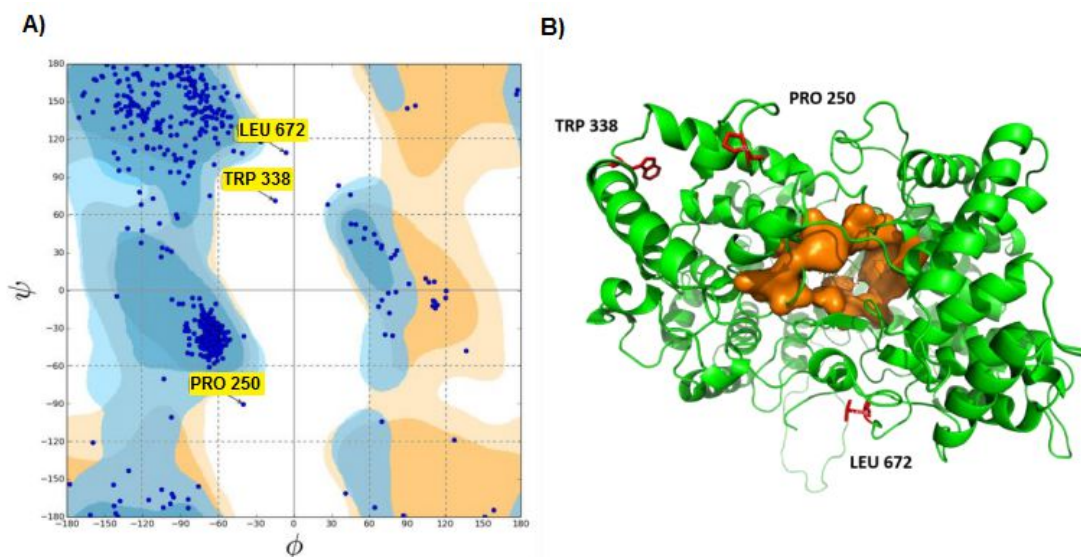
O potencial DOPE é calculado em uma modelagem comparando a estrutura modelada com a estrutura cristalográfica em relação à posição atômica das estruturas e a energia potencial em cada átomo. Assim, entre os 100 modelos foram escolhidos três modelos, aqueles com menor valor de DOPE, ou seja, os modelos em que a proteína está na menor conformação global de energia potencial e, possivelmente, os modelos em que a proteína está com a melhor estrutura validada (SHEN; SALI, 2006).

As interações entre os aminoácidos que formam a estrutura secundária de uma proteína relacionam-se entre si formando ângulos de torção e as rotações dos ângulos permitem que os resíduos se encontrem rotacionados de forma a melhor se acomodarem espacial e energeticamente. Assim, o diagrama de Ramachandran define os resíduos que se encontram nas regiões energeticamente mais favoráveis e desfavoráveis e orienta a avaliação da qualidade de modelos teóricos ou experimentais de proteínas (SANTOS; ALANCASTRO, 2003).

Deste modo, no diagrama de Ramachandran gerado para cada resíduo de aminoácido da friedelina sintase predita, o modelo gerado com o mínimo de resíduos de aminoácidos fora da região favorável foi selecionado. Os resíduos de

aminoácidos leucina na posição 672 (Leu672), triptofano na posição 338 (Trp338) e prolina na posição 250 (Pro250) estão na região desfavorável do diagrama de Ramachandran (figura 6 A), porém são resíduos de aminoácidos fora do sítio catalítico (figura 6 B) e, assim, não interferem no estudo.

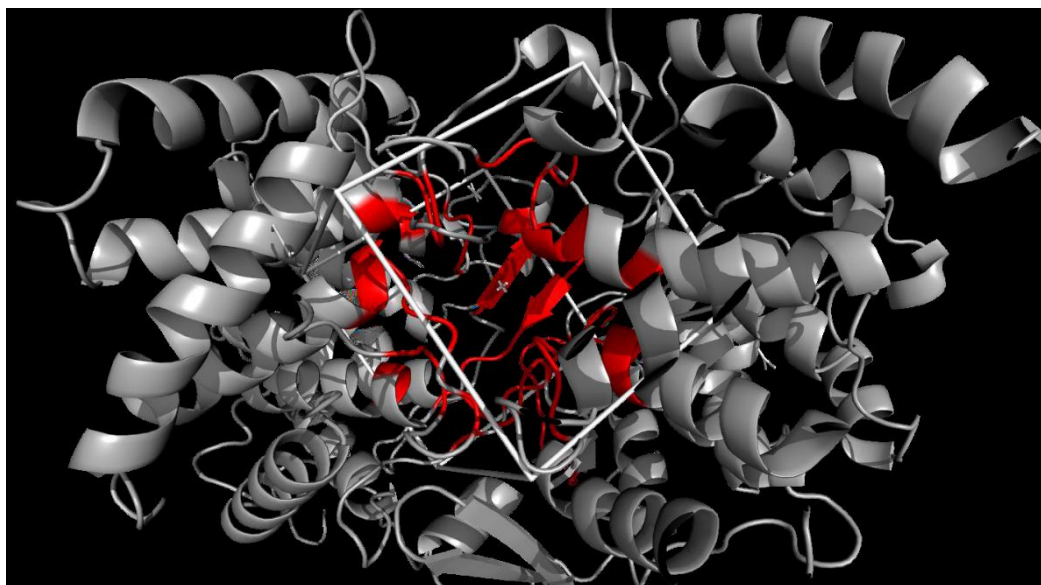
Figura 6 - Resultados da predição da estrutura molecular de friedelina sintase.



A) Diagrama de Ramachandran do melhor modelo de friedelina sintase mostrando os resíduos de aminoácidos na região desfavorável dos ângulos diédricos (Leu672, Trp338 e Pro250) B) posição dos aminoácidos na região desfavorável, em vermelho, na estrutura predita de friedelina sintase, destacando o sítio ativo em alaranjado. (Autora)

O sítio ativo foi localizado em um cubo de aresta de 12 Å em torno do resíduo de aminoácido triptofano na posição 612, região mais central do sítio ativo e, portanto, permitiu avaliar o máximo possível de resíduos ao redor do centro ativo (Figura 7).

Figura 7 - Definição do sítio de ligação para o *docking*.



Definição do sítio ativo (vermelho) realizado centrando um cubo de aresta 12 Å em torno do resíduo Trp 612. (Autora)

A estrutura monomérica globular de *MIFRS* contém 771 resíduos de aminoácidos e um desvio quadrático médio calculado (RMSD) de 0,3 Å com as coordenadas de C α alinhadas com a oxidoesqualeno ciclase 1W6K de *H.sapiens*. A topologia cilíndrica alfa de *MIFRS* abrange 21 alfa-hélices estando de acordo com a literatura de estrutura e funções de sequências genômicas CATH 1.50.10.20.

Posteriormente, o modelo de friedelina sintase foi submetido ao programa AutoDock Vina para execução de *docking* com lanosterol e friedelina posicionados no sítio ativo.

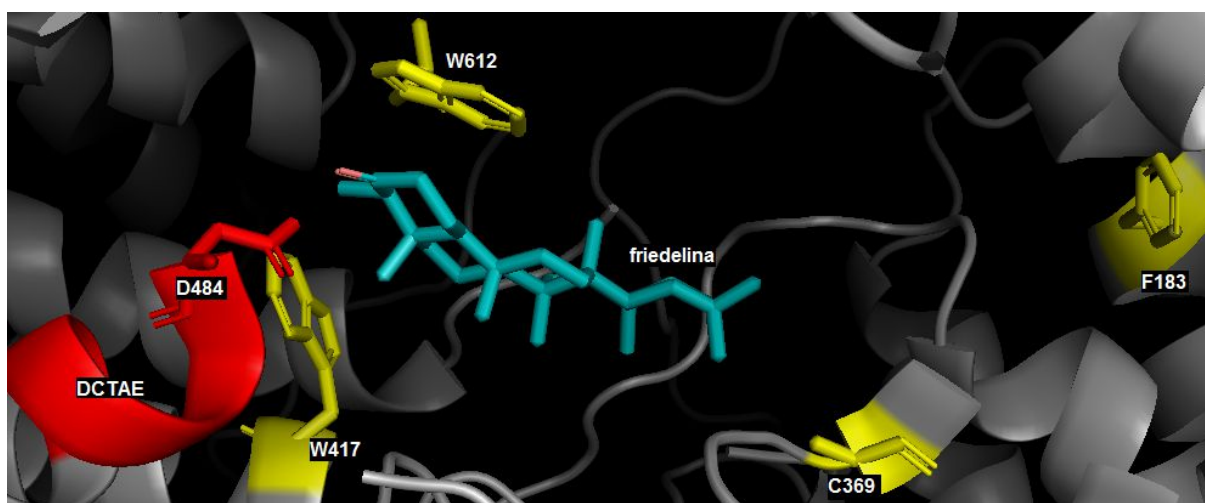
As mutações foram então selecionadas (Tabela 4) a partir da interação/acomodação dos resíduos no sítio ativo com a molécula de friedelina, como mostra a Figura 8. Para auxiliar o estudo da especificidade da friedelina sintase, também foi realizado um alinhamento múltiplo da sequência primária de várias oxidoesqualeno ciclasses depositadas no Genbank (Apêndice A) *versus* a sequência de friedelina sintase de *M. ilicifolia* (Apêndice B). Assim, a troca dos aminoácidos foi orientada tanto pela comparação com outras OSCs quanto pela similaridade entre os aminoácidos de acordo com a matriz de substituição *BLOSUM62* (*Blocks Substitution Matrix* - Anexo A), que busca regiões muito

conservadas de famílias de proteínas e exibe a frequência relativa de aminoácidos e as suas probabilidades de substituição a partir do banco de dados BLOCKS.

Tabela 4 - Resíduos de aminoácidos do sítio ativo da friedelina sintase selecionados para mutagênese sítio dirigida.

Resíduo	Troca	Posição	Justificativa
F	L	183	Na maior parte de outras OSCs ao invés de F é L
C	A	369	Em friedelina sintase de <i>K. daigremontiana</i> , ao invés de C é A
W	H	417	Interação entre o resíduo W417 com Y736
D	E	484	Resíduo doador de próton
W	F	612	Interação entre os resíduos F473 e W612 (importante a ligação H ou π stacking)

Figura 8 - Docking de friedelina no sítio ativo da enzima friedelina sintase predita evidenciando os resíduos de aminoácidos a serem mutados.



Modelo predito de friedelina sintase com os resíduos selecionados para reação de mutação sítio-dirigida (amarelo), o motivo conservado DCTAE (vermelho) e friedelina (ciano). Imagens formadas com auxílio do programa PyMOL™ Molecular Graphics System (versão 1.7.2.1). (Autora)

Desta forma, os resíduos de aminoácidos determinados pela análise *in silico* para as mutações podem ser importantes para a relação estrutura-atividade específica de *MiFRS* por diferentes motivos, como resumido na Tabela 4.

4.2 Obtenção dos mutantes do sítio ativo de *MiFRS*

Todas as mutações propostas acima foram obtidas por mutagênese sítio-dirigida e confirmadas por alinhamento global utilizando-se a ferramenta Clustal Omega das sequências obtidas pela reação de sequenciamento e o gene selvagem, como mostrados na figura a seguir. Os alinhamentos completos estão mostrados no Apêndice B.

Figura 9 - Alinhamento global múltiplo entre a sequência selvagem de *MiFRS* e as sequências mutadas obtidas.



O códon do resíduo de *MiFRS* selvagem está evidenciado em cinza e a sua troca está evidenciada em cinza. A) mutação F183L, troca de nucleotídeos TTC por TTG; B) mutação C369A, troca de nucleotídeos ACA por AGC; C) mutação W417H, troca de nucleotídeos TGG por CAT; D) mutação D484E, troca de nucleotídeos GAT por GAA; E) mutação W612F, troca de nucleotídeos TGG por TTT. (Autora)

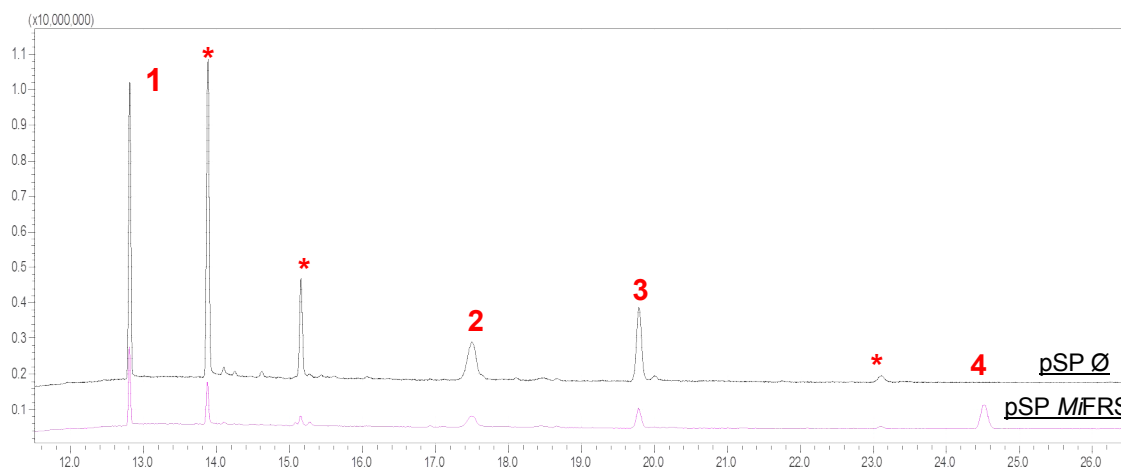
Com a confirmação das mutações dos aminoácidos selecionados, foi realizada a transformação destes em *S. cerevisiae* CEN.PK113-5D, cujo promotor constitutivo do gene *ERG7*, codificador de lanosterol sintase, foi trocado por um promotor fraco, para redução dos níveis de sua expressão e diminuição da competição da lanosterol sintase pelo substrato oxidoesqualeno (Tatiana Moreira, dados não publicados). Assim, possibilitou-se o desvio da rota de formação de ergosterol para aumento de produção de triterpenos heterólogos.

4.3 Análise dos produtos triterpênicos gerados heterologicamente

Os cromatogramas gerados a partir da injeção das frações apolares extraídas das células de *S. cerevisiae* expressando heterologicamente as sequências de *MiFRS* mutantes foram analisados separadamente e foram também comparados com os cromatogramas gerados a partir da injeção das frações apolares extraídas das

células de *S. cerevisiae* expressando a sequência selvagem de *MiFRS* e das células de controle negativo, contendo apenas com o plasmídeo vazio (figura 10).

Figura 10 - Cromatograma da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência de *MiFRS*.

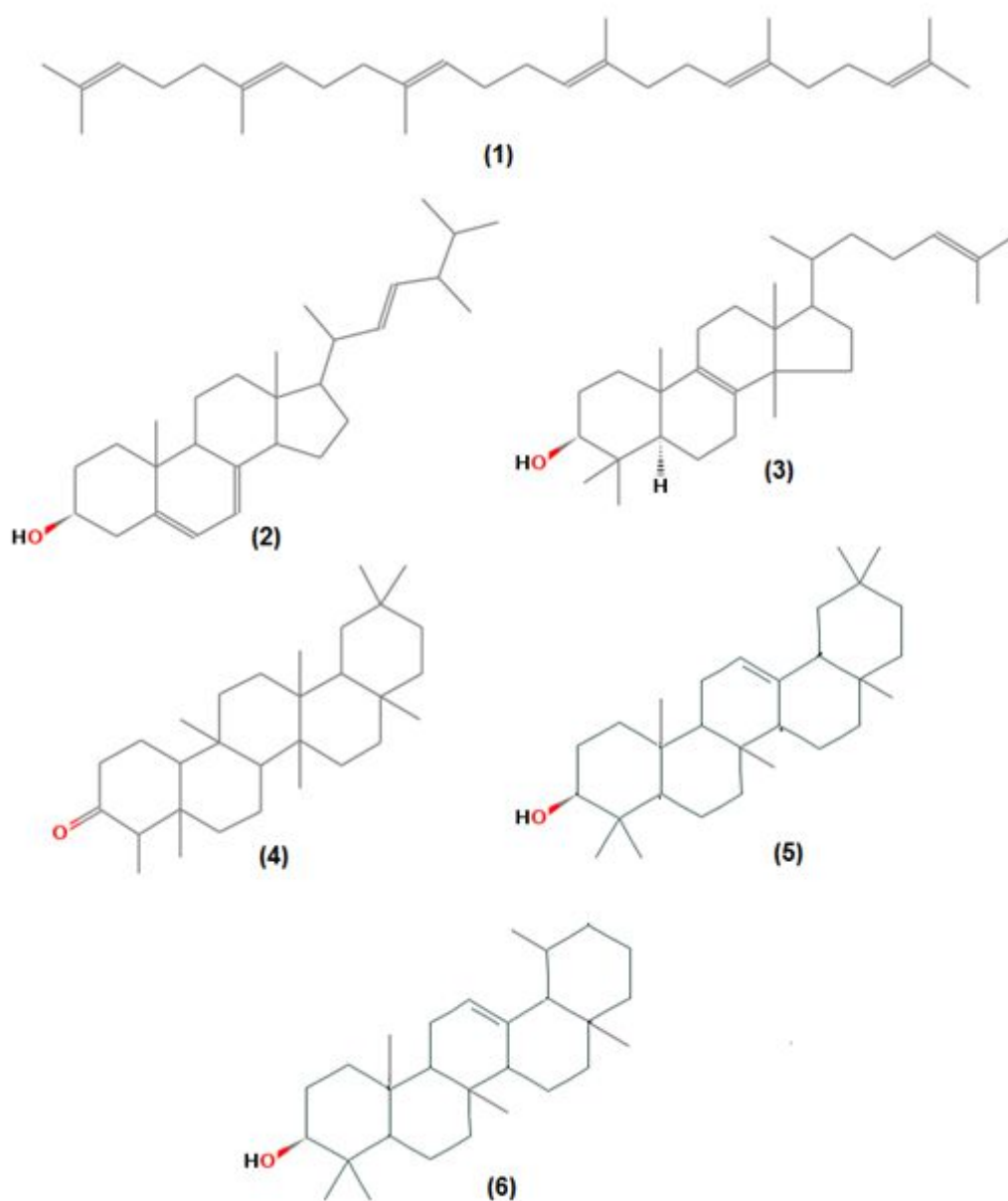


Pico 1 = esqualeno; Pico 2 = ergosterol; Pico 3 = lanosterol; Pico 4 = friedelina; * = picos resultantes de resíduos da coluna utilizada na CG-EM (oxirano e 2,2-Dimetil-3-(3,7,16,20-tetrametil-heneicosa-3,7,11,15,19-pentaenil)-oxirano) e de metabólitos gerados pelo metabolismo de ergosterol (9,19-ciclolanost-23-eno-3,25-diol) (Autora)

Foram observados metabólitos em comum decorrentes da via do ergosterol (representados pelos picos 1, 2 e 3) na amostra de *S. cerevisiae* utilizada como controle negativo e nas amostras expressando *MiFRS* selvagem e mutantes. Assim, o objetivo deste trabalho foi analisar a formação de diferentes terpenos, observando os cromatogramas das frações apolares correspondentes às produções heterólogas dos mutantes de *MiFRS* (Apêndice C) e os espectros de massa resultantes dos compostos diferencialmente produzidos. Os compostos derivados da expressão heteróloga de *MiFRS* e seus mutantes, identificados neste trabalho, estão representados e numerados na Figura 11.

Os cromatogramas gerados pelas amostras são apresentados a partir do tempo de 17 min porque esta é a região em que os compostos diferencialmente formados estão presentes. Contudo, no Apêndice X, estão apresentados os cromatogramas completos de todas as amostras analisadas neste trabalho, bem como estão apresentados os espectros dos compostos da via de ergosterol que apareceram com maior intensidade (esqualeno, lanosterol e ergosterol) e dos compostos heterologicamente produzidos (friedelina, β -amirina e α -amirina).

Figura 11 - Metabólitos gerados pela produção heteróloga do gene selvagem de *MiFRS* (1-4), dos mutantes (1-6) e, pela linhagem de levedura contendo apenas o vetor vazio (1-3).



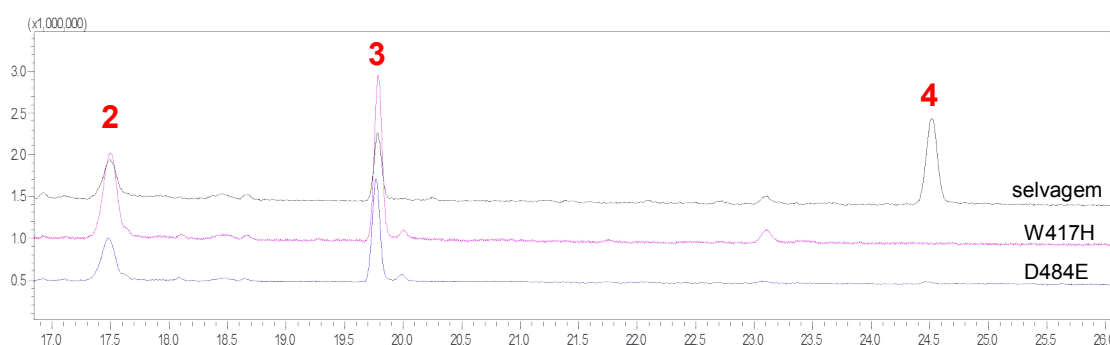
Metabólitos: (1) = esqualeno; (2) = ergosterol; (3) = lanosterol; (4) = friedelina; (5) = β -amirina; (6) = α -amirina; (disponível em: <http://webbook.nist.gov/chemistry/> e <https://pubchem.ncbi.nlm.nih.gov/>)

4.3.1 Mutantes com perda de função

Os cromatogramas foram analisados com intuito de verificar a presença de friedelina, ou outro composto não derivado do metabolismo primário celular, nas amostras de *S. cerevisiae* expressando heterologicamente as sequências mutantes de

MiFRS. Foi observada a ausência de formação de friedelina nos cromatogramas das células de *S. cerevisiae* expressando heterologicamente quatro sequências de *MiFRS* mutantes: W417H e D484E (Figura 12).

Figura 12 - Cromatogramas da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência selvagem de *MiFRS* e os mutantes que perderam a função de produção de friedelina: W417H e D484E.



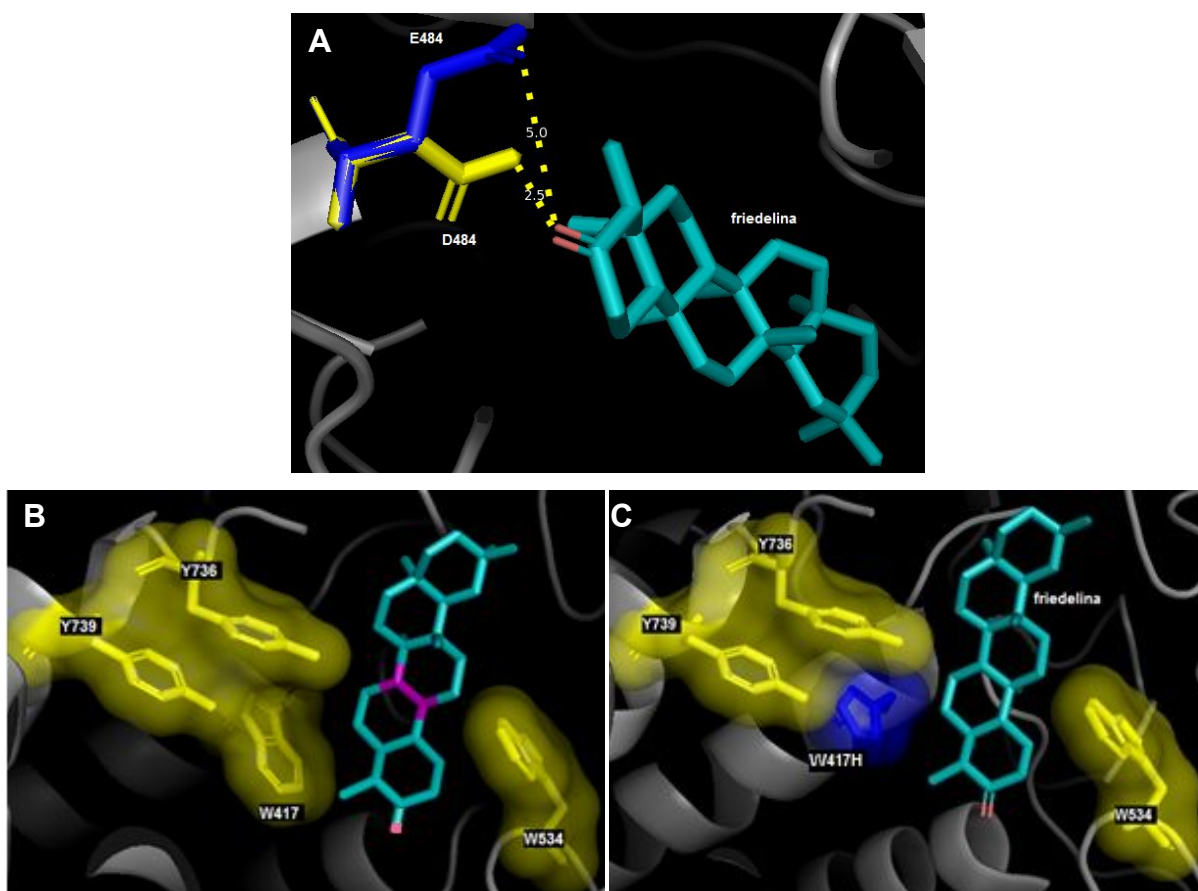
Pico 2= ergosterol; Pico 3= lanosterol; Pico 4= friedelina. (Autora)

O resíduo de ácido aspártico na posição 484 faz parte do sítio catalítico Asp-Cys-Thr-Ala-Glu (DCTAE), altamente conservado entre as oxidoesqualeno ciclases. Nele, o substrato oxidoesqualeno se ancora e, a partir da doação de próton do ácido aspártico ao substrato, é iniciada a sua ciclização (Ito et al, 2013). Para a formação de friedelina, é proposto que este mesmo resíduo seja o acceptor final do próton, formando assim, o grupo cetônico (Wang et al., 2010). A troca do ácido aspártico na posição 484 pelo ácido glutâmico impediu a enzima friedelina sintase de realizar a produção da friedelina porque o ácido glutâmico possui um grupo metilênico a mais em comparação ao ácido aspártico e essa maior cadeia carbônica influencia na acomodação do oxidoesqualeno no sítio ativo da friedelina sintase, impedindo a ciclização e consequente produção da friedelina. O grupo metileno a mais no ácido glutâmico, comparando com o ácido aspártico, não permite que o substrato se ancore. Segundo Hoshino (2017), o grupo metileno interfere no acesso do grupo carboxílico ao anel epóxido do substrato e assim, deve haver uma distância fixa entre o substrato e o aminoácido de ancoramento para que ocorra o início da policiclização. A distância entre o anel epóxido do substrato e a carbonila do aminoácido ácido aspártico na posição 484 é de 2,5 Å, enquanto que a distância

entre o anel epóxido do substrato com a troca do ácido aspártico pelo ácido glutâmico nessa posição é o dobro, 5 Å, como observado na figura 13 A , interferindo na formação da ligação de hidrogênio (~2,6 Å) entre o resíduo e o terpeno. Da mesma forma, foi visto que a mutação D485E em β -amirina sintase de *Euphorbia tirucalli* proporcionou a perda da atividade desta enzima devido à inclusão do grupo metilênico, em detrimento à presença do grupamento ácido do ácido glutâmico (Ito et al., 2013), por sua vez colaborando com a observação de que uma distância fixa deve existir para a doação do próton ao grupo epóxido.

A troca do triptofano na posição 417 por uma histidina impediu a enzima friedelina sintase de produzir friedelina. Ao estudar a estrutura molecular da friedelina sintase (figura 13 B), foi observado, através da nuvem de densidade eletrônica, que o triptofano na posição 417 forma uma cavidade rica em elétrons- π , interagindo também com outros resíduos aromáticos como Y366 e Y369, como já previamente observado em lanosterol sintase (Wu et al., 2006). Esta densidade eletrônica favorece os rearranjos durante a ciclização e a desprotonação final, resultando na produção da friedelina. A histidina, por ser um aminoácido com um imidazol parcialmente carregado, de menor volume que o triptofano, interfere na densidade eletrônica necessária para o rearranjo dos cátions intermediários no sítio ativo (figura 13 C), resultando na perda da capacidade de produção de friedelina pela enzima.

Figura 13 - Análise *in silico* por modelagem molecular de friedelina sintase e seus mutantes com perda de função: W417H e D484E.



A) mutante D484E, evidenciando os aminoácidos ácido aspártico (amarelo) e ácido glutâmico (azul), a friedelina sintase docada no sítio ativo e as distâncias das interações dos aminoácidos com a friedelina (linha pontilhada em amarelo); B) *MIFRS* selvagem evidenciando a densidade eletrônica formada pelos resíduos W417, Y736, Y739 e W534 (amarelo) no sítio ativo, com destaque para o posicionamento dos carbocátions intermediários em C8 e C9 (lilás); C) mutante W417H (azul), evidenciando a alteração da densidade eletrônica no sítio ativo (amarelo) alterando a interação com os cátions intermediários de friedelina (ciano). (Autora)

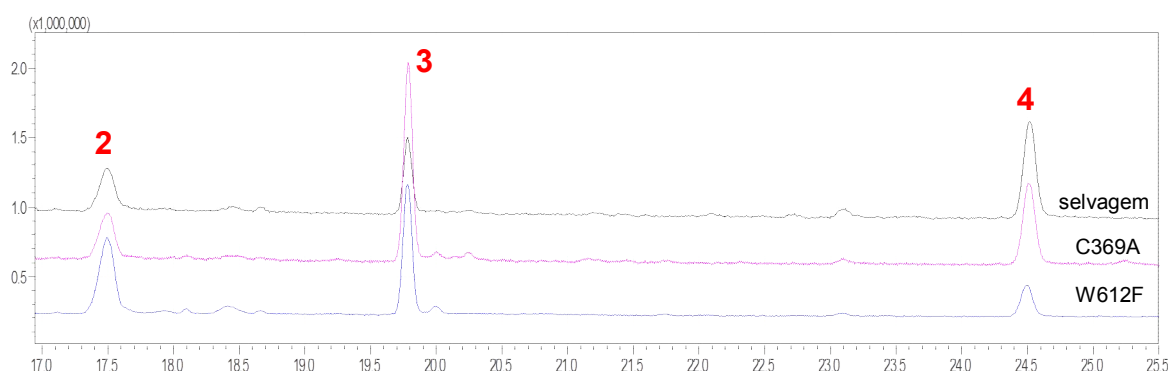
Segundo Ito et al. (2016), ao realizar a troca de resíduos de aminoácidos envolvidos com interações cátion- π , como os resíduos Y259 e W256, por aminoácidos de volume pequeno e sem capacidade de interações através dos elétrons π , ocorre a desestabilização dos carbocátions intermediários, produzindo metabólitos intermediários e/ou não ocorrendo a produção de β -amirina. O mesmo é observado no mutante de W417 em que a troca por um aminoácido de cadeia pequena, interrompeu a interação do resíduo e o substrato, provocando a interrupção da produção da friedelina.

Assim pode-se notar que a perda do distanciamento correto para a doação/aceito de próton entre a enzima e o substrato/produto, como mostrado na mutação D484E, bem como a perda de interação dos cátions intermediários com o sítio ativo, como observado no mutante W417H, foram os principais motivos de perda de função da enzima friedelina sintase de *M. ilicifolia*.

4.3.2 Mutantes com manutenção da atividade da friedelina sintase

Algumas mutações na sequência de *MIFRS* mantiveram a sua atividade formadora de friedelina, sem levar à produção de outros compostos. Tais mutantes foram W612F e C369A (Figura 14).

Figura 14 - Cromatogramas da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência selvagem de *MIFRS* e os mutantes W612F e C369A.



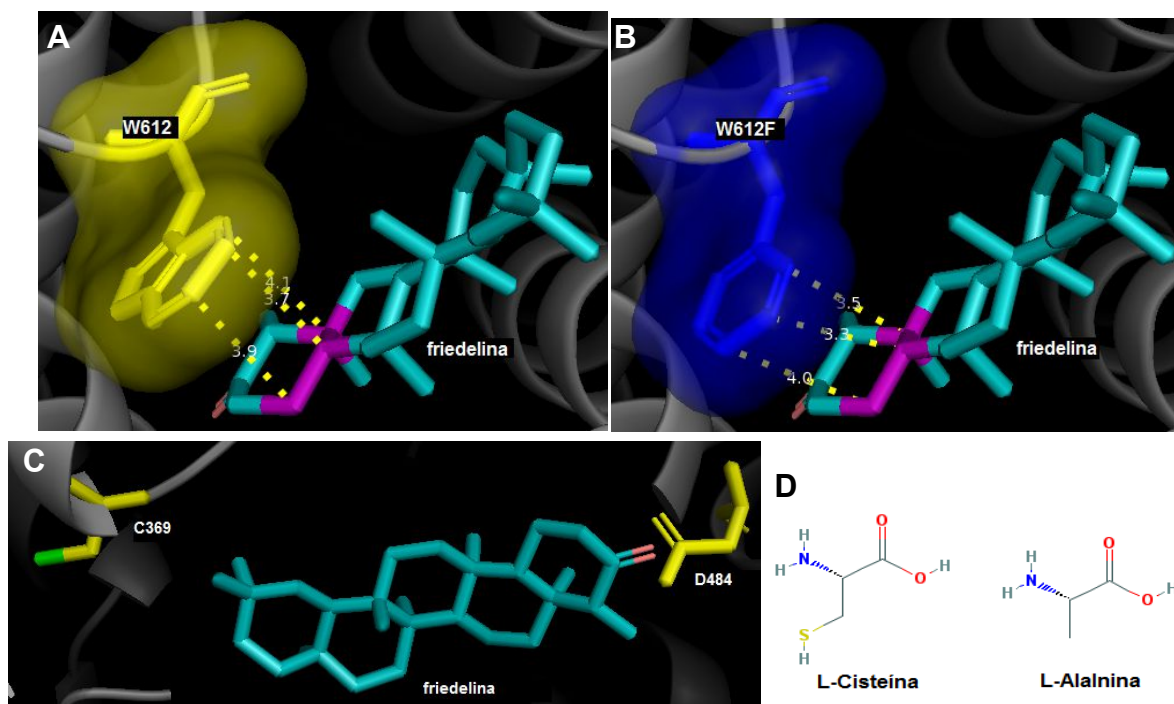
Pico 2 = ergosterol; pico 3 = lanosterol; pico 4 = friedelina. (Autora)

O resíduo W612 tem característica de realizar interações do tipo cátion- π , estabilizando os últimos carbocátions intermediários no anel A (figura 15 A). E, por isso, é um resíduo altamente conservado na produção de triterpenos pentacíclicos, como lupeol e β -amirina e tetraciclos (lanosterol e cicloartenol) (Hoshino, 2017).

Ao realizar o estudo da modelagem de friedelina sintase, notou-se que as distâncias de interações entre os elétrons π do resíduo W612 e os carbocátions no anel A (figura 15 A) são semelhantes em relação ao resíduo mutado W612F e os mesmos cátions intermediários (figura 15 B). A manutenção da formação de friedelina pelo mutante W612F demonstrou que a troca do aminoácido manteve as interações cátion- π , importantes para a estabilização dos rearranjos finais de

carbocátion na formação de friedelina. Contudo, é interessante observar que o resíduo de triptofano nesta posição é conservado entre as mais diversas OSC's, sendo que apenas a enzima β -amirina syntase multifuncional (AMY2) de *Lottus japonicus* possui uma alanina na posição correspondente (Apêndice B).

Figura 15 - Análise *in silico* por modelagem molecular de friedelina sintase e seus mutantes que favorecem a manutenção da atividade de friedelina sintase.



A) *MIFRS* selvagem evidenciando o resíduo W612 (amarelo) e o comprimento das interações cátion- π com os carbocátions intermediários (lilás) de friedelina (ciano); B) mutante W612F, evidenciando a manutenção de comprimentos semelhantes das interações cátion- π com os carbocátions intermediários (lilás) de friedelina; C) *MIFRS* selvagem evidenciando o resíduo C369A (amarelo), o e o resíduo D484 (amarelo) em que ocorre o ancoramento do substrato para formação de friedelina (ciano); D) estrutura da L-cisteína e L-alanina, evidenciando os átomos: hidrogênio (cinza), oxigênio (vermelho), enxofre (amarelo), nitrogênio (azul) e cadeia carbônica (preto) (<https://pubchem.ncbi.nlm.nih.gov>). (Autora)

A troca da cisteína na posição 369 por uma alanina (figura 15 E), um aminoácido menor, também não impediu a enzima de produzir friedelina, como observado no cromatograma da figura 14. A partir do alinhamento múltiplo entre as diversas oxidoesqualeno ciclases, foi observada a presença de uma alanina na posição correspondente da friedelina sintase de *K. daigremontiana*, enquanto que na

maioria das outras triterpeno sintases de pentacíclicos, a cisteína é conservada nesta posição (Apêndice B).

Neste mutante não foi observada a produção de nenhum outro triterpeno, embora tenha sido descrita a produção de β -amirina e glutinol em *KdFRS*, indicando que outros resíduos de *MiFRS* estão relacionados à sua especificidade.

É interessante notar, contudo, que a cadeia lateral de tamanho e volume menor do aminoácido cisteína e alanina na posição 369 é importante para a atividade de *MiFRS* e de outras OSCs que formam o cátion damarenil, visto que nas OSCs que formam o cátion protosteril, a existência de um resíduo de prolina, um aminoácido com cadeia constricta, nesta posição leva à produção de triterpenos tetracíclicos.

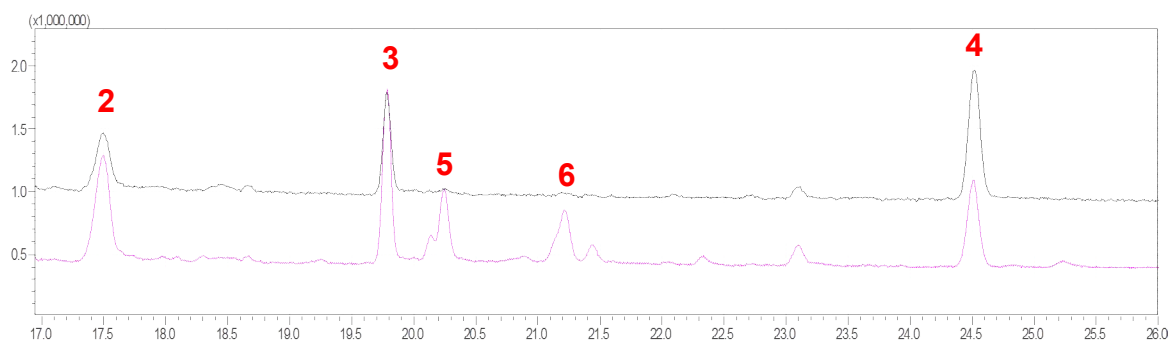
Não foram encontrados relatos na literatura de estudos mutagênicos pela troca deste resíduo de cisteína em outras OSC. Vale salientar que a enzima friedelina sintase de *M. ilicifolia* compartilha 65% de identidade com a de *K. daigremontiana* e que outro trabalho do nosso grupo, publicado recentemente, elucidou a importância do resíduo L482 para a produção de friedelina, um resíduo conservado somente nestas duas enzimas e na enzima glutinol sintase, também de *K. daigremontiana*, produtora de glutinol majoritariamente e de friedelina em menor quantidade (SOUZA-MOREIRA et al., 2016). No presente trabalho de Mestrado, avaliou-se, por outro lado, um resíduo conservado entre a maioria das triterpeno pentacíclico sintases, encontrado também no sítio ativo de *MiFRS*, porém diferente entre esta e *KdFRS* e foi demonstrado que a presença de um resíduo de pequeno volume favorece a produção de friedelina. Outros estudos mutacionais seriam interessantes a fim de avaliar a função de outros resíduos como treonina e serina, presentes em algumas OSCs, ou mesmo como metionina, aminoácido maior que a cisteína, mas também contendo enxofre em sua cadeia lateral.

Observou-se com estas análises que a troca de alguns resíduos do sítio ativo favoreceu a manutenção do mesmo tipo de interação, sem alteração estrutural de friedelina sintase, mantendo a formação de friedelina.

4.3.3 Mutante com ganho de função

Um mutante de friedelina sintase, além de friedelina, também produziu outros triterpenos pentacíclicos, como apresentados na figura 16.

Figura 16 - Cromatograma da fração apolar das células de *S. cerevisiae* expressando heterologicamente a sequência de *MiFRS* e mutante com ganho de função.



Pico 2 = ergosterol; pico 3 = lanosterol; pico 4 = friedelina; pico 5 = β -amirina; pico 6 = α -amirina.

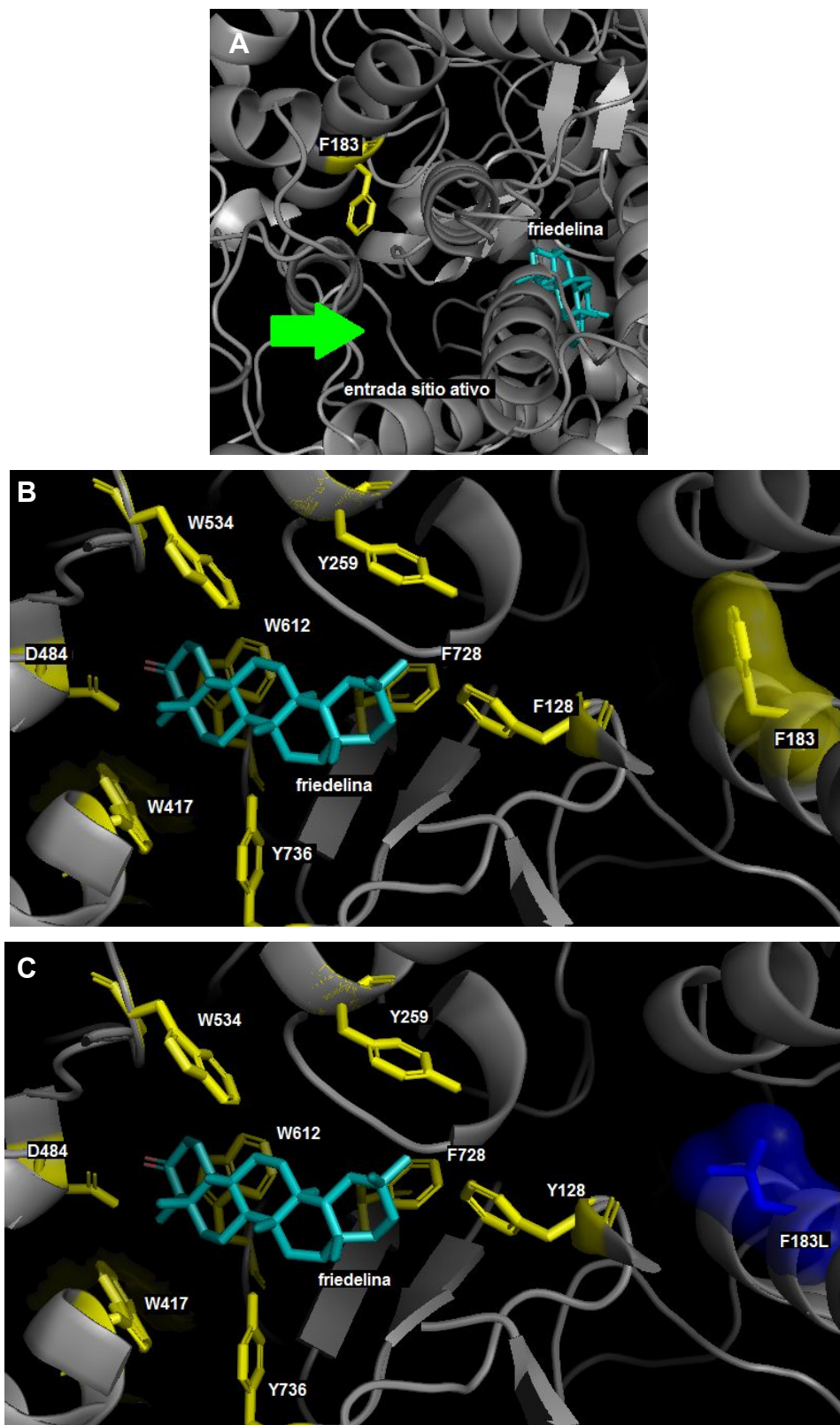
(Autora)

O resíduo de fenilalanina na posição 183 da friedelina sintase está localizado em uma região entre a entrada do sítio ativo e o sítio de ancoragem ao aspartato catalítico. A troca de fenilalanina por leucina na posição 183 não impediu a enzima friedelina sintase de produzir friedelina, porém, também possibilitou a formação de outros triterpenos pentacíclicos, como observado no cromatograma da figura 16, levando à perda de especificidade da enzima. O modelo estrutural de *MiFRS* mostra que a F183 localizada numa região próxima ao sítio ativo e em que o substrato entra para se ancorar (figura 17 A). Ao se realizar a troca de fenilalanina na posição 183 por uma leucina, um aminoácido menor e não aromático, nota-se que houve a produção não só de friedelina, mas também de β -amirina e α -amirina (figura 11), assim como nas outras triterpeno sintases de pentacíclicos. Observa-se que o modelo de estrutura molecular utilizado para a análise dos resíduos que possivelmente poderiam estar envolvidos por interações com o resíduo F183 na produção dos diferentes produtos triterpenicos, não foi suficiente para explicar o porquê das produções diferenciadas tanto de β -amirina quanto de α -amirina. Portanto, são necessários outros estudos mais aprofundados para entender qual o envolvimento do resíduo F183 com a produção diferenciada dos triterpenos, já que este resíduo está localizado somente próximo ao sítio ativo e não diretamente sobre o substrato ou interagindo diretamente com outros aminoácidos do sítio ativo.

A formação de β -amirina também ocorre pela ação da friedelina sintase de *K. daigremontiana* (*KdFRS*) (Wang et al., 2010), enzima que possui um resíduo de

leucina na posição 183. A maioria das OSCs que produzem triterpenos tetracíclicos ou pentacíclicos também possui leucina na posição correspondente, exceto as enzimas isomultiflorenol sintase (LcIMS1) de *Luffa cylindrica* e triterpeno sintase multifuncional (RsM2) de *Rhizophora stylosa* que também possuem uma fenilalanina e as enzimas arabidiol sintase (PEN1), talianol sintase (PEN4) e marneral sintase (PEN5) de *Arabidopsis thaliana*, que possuem o isômero isoleucina (Apêndice B). Assim, a produção de α -amirina e β -amirina pelo mutante F183L condiz com a função do resíduo de leucina nas outras triterpeno pentacíclico sintases, contudo, a produção concomitante e relativamente maior de friedelina, em relação aos outros triterpenos da amostra, demonstra que apenas este resíduo, não envolvido com o sítio ativo, mas direcionado à sua cavidade, não está envolvido na troca de atividade da enzima.

Figura 17- Análise *in silico* por modelagem molecular de friedelina sintase e o mutante com manutenção da atividade de friedelina sintase e o ganho de função: F183L.



A) entrada do sítio ativo (flecha em verde), aminoácido fenilalanina na posição 183 (amarelo) e a friedelina docada no sítio ativo (ciano); B) sítio ativo da friedelina sintase evidenciando o resíduo F183 (amarelo) e os resíduos participantes das interações π (amarelo), representando também a friedelina em *docking* no sítio ativo (ciano) e o resíduo D484 onde o substrato se ancora (amarelo); C) sítio ativo da friedelina sintase evidenciando o resíduo de leucina do mutante F183L (azul) e os demais resíduos e a molécula apresentados no quadro B. (Autora)

É interessante observar que diferentes resíduos presentes no sítio ativo exercem funções importantes para a atividade e especificidade catalítica da enzima. Tais aminoácidos são geralmente altamente conservados entre as OSCs e sua troca leva à perda de formação de qualquer triterpeno, como observado nos mutantes: W417H e D484E. Foi observado que os resíduos correspondentes de posição em lanosterol sintase de *S. cerevisiae* (Y510, H234, Y707, Y710 e W390) e em β -amirina sintase de *E. tirucalli* (D484, Y259, W257, F728, W612) também estão relacionados à atividade das suas enzimas.

Outros se mostram divergentes entre os grupos de enzimas OSCs, podendo a sua troca levar à formação de diferentes triterpenos e, por isso, relacionam-se com a especificidade da friedelina sintase estudada, como o mutante F183. Resíduos de aminoácidos como a mesma característica de formação de outros triterpenos, constituintes do sítio ativo, porém com uma distância maior do substrato já foram relatados, como o resíduo C393 da enzima cucurbitadienol sintase de *Cucumis sativus* e o resíduo F728 da enzima β -amirina sintase de *Avena strigosa*.

Outros resíduos, contudo, podem ser considerados menos relacionados com a especificidade da enzima, visto que sua troca por aminoácidos com características semelhantes manteve a formação de friedelina (W612 e C369).

Desta forma, é possível determinar que os resíduos W417 e D484 são essenciais para a atividade catalítica da friedelina sintase de *M. ilicifolia*, enquanto que os resíduos W612, C369 e F183 estão relacionados à estabilização e controle dos rearranjos específicos da formação de friedelina.

5 CONCLUSÕES

A predição da enzima friedelina sintase por homologia resultou em um modelo dentro dos parâmetros aceitáveis e possibilitou a realização do estudo *in silico* dos aminoácidos constituintes do sítio ativo da enzima. Assim, a análise das possíveis interações com o sítio ativo da enzima e a comparação dos resíduos dessa enzima com outras oxidoesqualeno ciclases guiaram a seleção dos resíduos a serem mutados, sendo possível avaliar o efeito na estrutura-atividade dos mutantes por meio da análise dos compostos formados heterologamente em *S. cerevisiae*.

Desta forma, foi possível observar resíduos do sítio ativo da enzima envolvidos com a sua função catalítica e especificidade. Assim, os resíduos determinantes para a atividade catalítica da enzima são:. Já os resíduos de aminoácidos C369, W612 e F183 estão envolvidos com a estabilização e controle dos rearranjos específicos da formação de friedelina.

A friedelina sintase é uma OSC singular por levar à formação de um triterpeno cetônico, com maior número de rearranjos. A compreensão da sua atividade singular entre as OSC, de acordo com a sua estrutura, é ainda modesta e com este trabalho foram determinados resíduos essenciais para a sua função catalítica bem como resíduos envolvidos com a sua estabilidade.

Tais resíduos, contudo, não são os únicos determinantes para a sua atividade específica de formação de friedelina e futuros estudos com outros resíduos podem esclarecer melhor a relação estrutura-atividade desta enzima.

REFERÊNCIAS

- ABE, I.; PRESTWICH, G. Identification of the active site of vertebrate oxidosqualene cyclase. **Lipids**, v. 30, n. 3, p. 231-234, 1995.
- ANDRE, C. M.; GREENWOOD, J. M.; WALKER, E. G.; RASSAM, M.; SULLIVAN, M.; EVERS, D.; PERRY, N. B.; LAING, W. A. Anti-inflammatory procyanidins and triterpenes in 109 apple varieties. **Journal of Agricultural and Food Chemistry**, v. 60, p. 10546-10554, 2012.
- BERO, J.; FREDERICH, M.; QUETIN-LECLERCQ, J. Antimalarial compounds isolated from plants used in traditional medicine. **Journal of Pharmacy and Pharmacology**, v. 61, p. 1401-1433, 2009.
- CHANG, C. H.; WEN, H. Y.; SHIE, W. S.; LU, C. T.; LI, M. E.; LIU, Y. T.; LI, W. H.; WU, T. K. Protein engineering of oxidosqualene-lanosterol cyclase into triterpene monocyclase. **Organic & biomolecular chemistry**, v. 11, n. 25, p. 4214-4219, 2013.
- CORSINO, J.; CARVALHO, P. R. F.; KATO, M. J.; LATORRE, L. R.; OLIVEIRA, O. M. M. F.; ARAUJO, A. R.; BOLZANI, V. da S.; FRANÇA, S. C.; PEREIRA, A. M. S.; FURLAN, M. Biosynthesis of friedelane and quinonemethide triterpenoids is compartmentalized in *Maytenus aquifolium* and *Salacia campestris*. **Phytochemistry**, v. 55, n. 7, p. 741-748, 2000.
- DEEB, D.; GAO, X.; LIU, Y.; PINDOLIA, K.; GAUTAM, S. C. Inhibition of hTERT/telomerase contributes to the antitumor activity of pristimerin in pancreatic ductal adenocarcinoma cells. **Oncology Reports**, v. 35, p. 518-524, 2015.
- DOSHI, G. M.; NALAWADE, V. V.; MUKADAM, A. S.; CHASKAR, P. K.; ZINE, S. P.; SOMANI, R. R.; UNE, H. D. Structural elucidation of chemical constituents from *Benincasa hispida* and *Carissa congesta* roots by gas chromatography: Mass spectroscopy. **Pharmacognosy research**, v. 7, n. 3, p. 283-293, 2015.
- EDELHEIT, O.; HANUGOKLI, A.; HANUGOKLI, I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. **BMC Biotechnology**, v. 9, 2009. doi:10.1186/1472-6750-9-61.
- GAS-PASCUAL, E.; BERNA, A.; BACH, T. J.; SCHALLER, H. Plant oxidosqualene metabolism: cycloartenol synthase-dependent sterol biosynthesis in *Nicotiana benthamiana*. **PLoS One**, v. 10, 2014. doi: 10.1371/journal.pone.0109156.
- GERMANN, M.; GALLO, C.; DONAHUE, T.; SHIRZADI, R.; STUKEY, J.; LANG, S.; RUCKENSTUHL, C.; OLIARO-BOSSO, S.; McDONOUGH, V.; TURNOWSKY, F.; BALLIANO, G.; NICKELS, J. T. Jr. Characterizing sterol defect suppressors uncovers a novel transcriptional signaling pathway regulating zymosterol biosynthesis. **The Journal of Biological Chemistry**, v. 280, p. 35904-35913, 2005.

- HART, E. A.; HUA, L.; DARR, L. B.; WILSON, W. K.; PANG, J.; MATSUDA, S. P. T. Directed evolution to investigate steric control of enzymatic oxidosqualene cyclization. An isoleucine-to-valine mutation in cycloartenol synthase allows lanosterol and parkeol biosynthesis. **Journal of the American Chemical Society**, v. 121, p. 9887-9888, 1999.
- HE, P.; YE, F.; HUANG, S.; GUO, Y.; WANG, H.; WU, Y. Anti-inflammatory effect of pristimerin on TNF α -induced inflammatory responses in murine macrophages. **International Journal of Clinical and Experimental Pathology**, v. 9, p. 1186-1194, 2016.
- HOSHINO, T. β -amyrin biosynthesis: catalytic mechanism and substrate recognition. **Organic&Biomolecular Chemistry**, 2017. doi:10.1039/C7OB00238F.
- ITO, H.; FUKUDA, Y.; MURATA, K.; KIMURA, A. Transformation of intact yeast cells treated with alkali cations. **Journal of Bacteriology**, v. 153, p. 163-168, 1983.
- ITO, R.; MASUKAWA, Y.; HOSHINO, T. Purification, kinetics, inhibitors and CD for recombinant b-amyrin synthase from *Euphorbia tirucalli* L and functional analysis of the DCTA motif, which is highly conserved among oxidosqualene cyclases. **FEBS Journal**, v. 280, p. 1267-1280, 2013.
- KUSHIRO, T.; SHIBUYA, M.; EBIZUKA, Y. Cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. **European Journal of Biochemistry**, v. 256, p. 238-244, 1998.
- LIU, H.; NIUSMITH, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. **BioMedCentral Biotechnology**, 2008. doi:10.1186/1472-6750-8-91
- MATSUDA, S. P. T.; DARR, L. B.; HART, E. A.; HERRERA, J. B. R.; McCANN, K. E.; MEYER, M. M.; PANG, J.; SCHEPMANN, H. G. Steric bulk at cicloartenol synthase at position 481 influences cyclization and deprotonation. **Organic Letters**, v. 2, n. 15, p. 2261-2263, 2000.
- MOORE, P. A.; RUBEN, S. M.; ROSEN, C. A. Conservation of transcriptional activation functions of the nf-kb p50 and p65 subunits in mammalian cells and *Saccharomyces cerevisiae*. **Molecular and Cellular Biology**, v. 13, n. 3, p. 1666-1674, 1993.
- PARTOW, S.; SIEWERS, V.; BJORN, S.; NIELSEN, J.; MAURY, J. Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. **Yeast**, v. 27, n. 11, p. 955-964, 2010.
- POMMIÉ, C.; LEVADOUX, S.; SABATIER, R.; LEFRANC, G.; LEFRANC, M. P. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. **Journal of Molecular Recognition**, v. 7, n. 1, p. 17-32, 2004.

- SALMON, M.; THIMMAPPA, R. B.; MINTO, R. E.; MELTON, R. E.; HUGHES, R. K.; O'MAILLE, P. E.; HEMMINGS, A. M.; OSBOURN, A. A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. **Proceedings of the National Academy of Science of the United States of America**, v. 113, n. 30, p. e4407-e4414, 2016.
- SALVADOR, J. A.; MOREIRA, V. M.; GONCALVES, B. M.; LEAL, A. S.; JING, Y. Ursane-type pentacyclic triterpenoids as useful platforms to discover anticancer drugs. **Natural Product Reports**, v. 29, p. 1463-1479, 2012.
- SANTOS, O. A. S. F.; ALANCASTRO, R. B. Modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p. 253-259, 2003.
- SANTOS, V. A. F. F. M.; SANTOS, D. P.; CASTRO-GAMBOA, I.; ZANONI, M. V. B.; FURLAN, M. Evaluation of antioxidant capacity and synergistic associations of quinonemethide triterpenes and phenolic substances from *maytenus ilicifolia* (Celastraceae). **Molecules**, v. 15, n. 10, p. 6956-6973, 2010.
- SHEN, M. Y.; SALI, A. Statistical potential for assessment and prediction of protein structures. **Protein Science**, v. 15, p. 2507-2524, 2006.
- SHIBUYA, M.; ZHANG, H.; ENDO, A.; SHISHIKURA, K.; KUSHIRO, T.; EBIZUKA, Y. Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclases. **European Journal of Biochemistry**, v. 266, p. 302-307, 1999.
- SOUZA-MOREIRA, T. M.; ALVES, T. B.; PINHEIRO, K. A.; FELIPPE, L. G.; LIMA, G. M. A. de; WATANABE, T. F.; BARBOSA, C. C.; SANTOS, V. A. F. F. M.; LOPES, N. P.; VALENTINI, S. R.; GUIDO, R. V. C.; FURLAN, M.; ZANELLI, C. F. Friedelin synthase from *Maytenus ilicifolia*: leucine 482 plays an essential role in the production of the most rearranged pentacyclic triterpene. **Scientific Reports**, 2016. doi:10.1038/srep36858.
- THIMMAPPA, R.; GEISLER, K.; LOUVEAU, T.; O'MAILLE, P.; OSBOURN, A. Triterpene biosynthesis in plants. **Annual Review of Plant Biology**, v. 65, p. 225-257, 2014.
- VERDUYN, V.; POSTMA, E.; SCHEFFERS, W. A.; VAN DIJKEN, J. P. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. **Yeast**, v. 8, p. 501-517, 1992.
- WANG, Z.; YEATS, T.; HAN, H.; JETTER, R. Cloning and characterization of oxidosqualene cyclases from *Kalanchoe daigremontiana*. **The Journal of Biological Chemistry**, v. 285, n. 39, p. 29703-29712, 2010.
- WU, T. K.; WEN, H. Y.; CHANG, C. H.; LIU, Y. T. Protein plasticity: a single amino acid substitution in the *Saccharomyces cerevisiae* oxidosqualene lanosterol cyclase generates protosta-13(17),24-dien-3 β -ol, a rearrangement product. **Organic Letters**, v. 10, p. 2529-2532, 2008.

XUE, Z.; DUAN, L.; LIU, D.; GUO, J.; GE, S.; ÓMÁILLE, P.; DICKS, J.; OSBOURN, A.; QI, X. Divergent evolution of oxidosqualene cyclases in plants. **New Phytologist**, v. 193, p. 1022-1038, 2012.

APÊNDICE A - Sequências das enzimas oxidoesqualeno ciclases usadas no alinhamento múltiplo com seu número de acesso no GenBank, espécie e função.

(continua)

Nº de acesso	Espécies	Função
AB257562.1	<i>Arabidopsis thaliana</i>	Arabidiol sintase (PEN1)
AY327541.1	<i>Arabidopsis thaliana</i>	Talianol sintase (PEN4)
BT020456.1	<i>Arabidopsis thaliana</i>	Marneral sintase (PEN5)
AB274959.1	<i>Arabidopsis thaliana</i>	Triterpeno sintase multifuncional (PEN6)
NM_179572.1	<i>Arabidopsis thaliana</i>	Lupeol sintase 1 (LUP1)
NM_106545.3	<i>Arabidopsis thaliana</i>	Lupeol sintase 2 (LUP2)
NM_126681.2	<i>Arabidopsis thaliana</i>	Cicloartenol sintase (CAS1)
AB263204.1	<i>Rhizophora stylosa</i>	Triterpeno sintase multifuncional (RsM2)
AB289586.1	<i>Bruguiera gymnorhiza</i>	Lupeol sintase (BgLUS)
AB257507.1	<i>Kandelia candel</i>	Triterpeno sintase multifuncional (KcMS)
DQ268869.1	<i>Ricinus communis</i>	Lupeol sintase (RcLUS)
AB058643.1	<i>Luffa cylindrica</i>	Isomultiflorenol sintase (LcIMS1)
AB037203.1	<i>Glycyrrhiza glabra</i>	β -amirina sintase (GgbAS1)
AB181244.1	<i>Lotus japonicus</i>	β -amirina sintase (OSC1)
AB034802.1	<i>Pisum sativum</i>	β -amirina sintase (PSY)
AF478453.1	<i>Medicago truncatula</i>	β -amirina sintase (AMY1)
AF478455.1	<i>Lotus japonicus</i>	β -amirina sintase multifuncional (AMY2)
AB034803.2	<i>Pisum sativum</i>	amirina sintase mista (PSM)
AB009030.1	<i>Panax ginseng</i>	β -amirina sintase (PNY1)
AB014057.1	<i>Panax ginseng</i>	β -amirina sintase (PNY2)
AB055512.1	<i>Betula platyphylla</i>	β -amirina sintase (BPY)
HM623868.1	<i>Kalanchoe daigremontiana</i>	Taraxerol sintase (KdTAS)
AB263203.1	<i>Rhizophora stylosa</i>	Triterpeno sintase multifuncional (RsM1)
AB206469.1	<i>Medicago tirucalli</i>	β -amirina sintase (EtAS)

(conclusão)

Nº de acesso	Espécies	Função
HM623870.1	<i>Kalanchoe daigremontiana</i>	Friedelina sintase (KdFRS)
HM623869.1	<i>Kalanchoe daigremontiana</i>	Glutinol sintase (KdGLS)
HM623871.1	<i>Kalanchoe daigremontiana</i>	Lupeol sintase (KdLUS)
AB025343.1	<i>Olea europaea</i>	Lupeol sintase (OEW)
AB025345.1	<i>Taraxacum officinale</i>	Lupeol sintase (TRW)
AB181245.1	<i>Lotus japonicus</i>	Lupeol sintase (OSC3)
AB116228.1	<i>Glycyrrhiza glabra</i>	Lupeol sintase (GgLUS1)
AB055511.1	<i>Betula platyphylla</i>	Lupeol sintase (BPW)
AB181246.1	<i>Lotus japonicus</i>	Cicloartenol sintase (OSC5)
AB025968.1	<i>Glycyrrhiza glabra</i>	Cicloartenol sintase (GgCAS1)
AB009029.1	<i>Panax ginseng</i>	Cicloartenol sintase (PNX)
HM623872.1	<i>Kalanchoe daigremontiana</i>	Cicloartenol sintase (KdCAS)
AB055509.1	<i>Betula platyphylla</i>	Cicloartenol sintase (BPX1)
APG38073.1	<i>Maytenus ilicifolia</i>	Friedelina sintase (MiFRS)
AAA16975.1	<i>Saccharomyces cerevisiae</i>	Lanosterol sintase (ScLAS)
P48449.1	<i>Homo sapiens</i>	Lanosterol sintase (1W6K)
BAK52535.1	<i>Aster tataricus</i>	Shionona sintase (SS)

APÊNDICE B - Alinhamento entre oxidoesqualeno ciclases e friedelina sintase de *M. ilicifolia*, cuja identificação e número de acesso estão apresentados na tabela do apêndice A.

ScLAS	1	MTE-----FYSDT-----GLPKTDPRLWRLRTDELGRESWEYL
1W6K	1	---MTEGTCL-RRRGGPYKTEPA-----TDLGRWR-LNCERGRQTWTYLYL
BPX1	1	-----MWKLIKCAETARGDGGGGGSETWLRS-LNNHILGRQIWEFHH
KdCAS	1	-----MWKLIKAD-----AGGSQWLRS-VNNHILGRQIWDFFD
CAS1	1	-----MWKLIKAE-----GGSPWLRT-TNNHIVGRQVWEFDD
PNX	1	-----MWKLIKAE-----GGNPWLRT-LNDHIVGRQIWEFDD
PSX	1	-----MWKLIKAE-----GGTPWLRT-LNNHIVGRQVWEFDD
GgCAS1	1	-----MWKLIKAE-----GGSPWLRT-VNNHIVGRQVWEFDD
OSC5	1	-----MWKLIKAE-----GGNPWLRS-TNSHIVGRQVWEFDD
BPW	1	-----MWKLIKAE-----GGPG--LVS-GNDFIGRQHWEFDD
GgLUS1	1	-----MWKLIKGE-----GGAG--LIS-VNNFIGRQHWEFDD
OSC3	1	-----MWKLIKAE-----GGKG--LVS-VSNIIGRQHWVFFD
TRW	1	-----MWKLIKAE-----GGDEWLTT-TNNHIVGRQHWQFFD
OEW	1	-----MWKLIKAD-----G-TGPWLTT-TNNHIGRQHWEFDD
PEN5	1	-----MWLRIGA-----EAR--QDPHLFT-TNNFAGRQIWEFDD
PEN6	1	-----MWLRIGA-----KCG--DETHLFT-TNNYTGROTWEFDD
PEN4	1	-----MWLRITGP-----KAG--EDTHLFT-TNNYAGRQIWEFDD
PEN1	1	-----MWLRIGA-----KAG--NDTHLFT-TNNYVGRQIWEFDD
Rsm2	1	-----MGVWRLKIGE-----GA--NNPYLTS-TNNFVGRQTWVFE
LcIMS1	1	-----MWRLKVAD-----GG--NDPYIYS-MNNFIGRQIWEFDD
MiFRS	1	-----MWKLIKAD-----RCNCPYNEYLYT-TNDFVGRQIWEFDD
LUP2	1	-----MWKLIKGE-----GNG--EDPYLFS-SNNFVGRQTWEFDD
LUP1	1	-----MWKLIKIK-----GNG--EDPHLFS-SNNFVGRQTWKFFD
RcLUS	1	-----MWRLKIAE-----GG--NNPYIYS-TNNFQGRQIWVFFD
KcMS	1	-----MWRLKIAE-----GG--DNPYIYS-TNNFLGRQTWEFEE
BgLUS	1	-----MWRLKIAE-----GG--NNPYIYS-TNNFVGRQTWEFDD
SS	1	-----MWRLKIAD-----GG--NNPYLYS-TNNFIGRQTWEFDD
KdFRS	1	-----MWKLIKAE-----GG--SDPYIYT-TNNFVGRQIWEFDD
KdLUS	1	-----MWKLIKAD-----GG--SNPYIFT-TNNFVGRQIWEFDD
KdGLS	1	-----MWKLIKAD-----GG--SNPYIFT-TNNFVGRQIWEFDD
EtAS	1	-----MWKLIKAE-----GG--NDEYLYS-TNNYVGRQTWVFFD
Rsm1	1	-----MWRLKIAE-----GG--NDPYLYS-TNNYVGRQIWEFDD
PNY1	1	-----MWKLIKAE-----GNK--NDPYLYS-TNNFVGRQIWEFDD
PNY2	1	-----MWRLMTAK-----GG--NDPYLYS-TNNFIGRQIWEFDD
KdTAS	1	MSFVWVEESKECSEQRKGS MWKLIKIAQ-----GG--KDPYLYS-TNNYVGRQIWEFDD
BPY	1	-----MWRLKIAD-----GG--SDPYIYS-TNNFVGRQIWEFDD
PSM	1	-----MWKLIKIGD-----GG--KDRNIFS-TNNFVGRQIWEFDD
AMY2	1	-----MWKLVAD-----GG--KNPYIFS-INNFVGRQIWEYDD
AMY1	1	-----MWKLIKGE-----GK--NEPYLFS-TNNFVGRQIWEYDD
PSY	1	-----MWRLKIAE-----GG--NDPYLFS-TNNFVGRQIWEYDD
OSC1	1	-----MWKLVAD-----GG--KDPYIFS-TNNFVGRQIWEYDD
GgbAS1	1	-----MWRLKIAE-----GG--KDPYIYS-TNNFVGRQIWEYDD

ScLAS	35	TPQQAANLP-----PSTFTQWLL---QDPK-----FPQPNPERN
1W6K	40	QDER-AGREQTGLEAYA-----LGLDTKNYFK---DLPKA-----
BPX1	41	PELG-TQEEELQQIDDARRRFWRERFERRHSSDLLMRIQFAKENPSS---ANIPQVKIKDT
KdCAS	31	PALG-SPEELAQIEDARDNFARHRFDKKHSADLLMRFQLTKENPQS---DLLPKVNIIGKI
CAS1	30	PNIG-TPEDIAAVEEARKSFSDNRFVQKHSADLLMRIQFSRENLSIS---PVLPOVKIEDT
PNX	30	PNIG-SPEELAEVEKVRNENFRNHRFEKKHSADLLMRIQFANENPGS---VVLPOVKVNDG
PSX	30	PHSG-SPOLDDIETARNFHDNRFTHKHSADLLMRIQFAKENPMN---EVLPKVKVKDV
GgCAS1	30	EKLG-SPELLEIEKARONFHDNRFTHKHSADLLMRIHFAKENPMN---EVLPKVRVKDI
OSC5	30	EKLG-SPOLAEIETARNFHDNRFTHKHSADLLMRIQFSKENPIG---EVLPKVKVKDV
BPW	29	PDAG-TPQERAEVEKVRREEFTKNRFQMKQSADLLMRMQLRKENPCQ---PIPPPVKVKET
GgLUS1	29	PNAG-TPQEHAEIERLRREFTKNRFSEIKQSADLLMRMQLRKENHYGTNNNIPAAVKLSDA
OSC3	29	PNAG-TPQEHAEIERLRREFTKNRFSEIKQSADLLMRMQLRKENPCG---PIPPAVKLRDV
TRW	31	PDAG-TPQERAEIEKIRLNFKLNRFQFKQSADLLMRTQLRKENPIN---KIPDAIKLNET
OEW	30	PEAG-TPDERVEVERLRREEFKKNRFRTKQSADLLMRMQLVKENQRV---QIPPAIKIKET
PEN5	32	ANGG-SPEELAEVEEARLNFNANNRSREKASPDLFWRRFQFLREKKFE---QKIPRVRIEDA
PEN6	32	ADAC-SPEELAEVDEARONFSINRSREKISADLLWRMQLREKKFE---QKIPRVEIGDA
PEN4	32	ANAG-SPEELAEVEEARHFSNNTSREKITTADLLWRMQLREKKFE---QKIPRVIIEDA
PEN1	32	ANAG-SPEELAEVEEARNFENNRSHYKASADLLWRMQLREKGFEE---QKIPRVVEDA
RsM2	33	PDGG-TPPEERDQVEEARONFYNKRFVRVPCSADLLWQMFQFLREKNFR---QKIPQVKVRDG
LcIMS1	31	PNAG-TPPEERAEIERLRHHEFTKNRHKGFPSADLLWRVQLLREKNFK---QSIPAVKVGDG
MiFRS	34	PNSG-TPPEELAEIEEARRKFETENRVEVKPASADLLWMMQFLRKNNFK---QTIPLRIGEK
LUP2	32	PKAG-TPPEERAAVEDARNMLDNRPVKGCSADLLWRMQLKEAKFE---QVIIPVKIDDG
LUP1	32	HKAG-SPEERAAVEEARRGFLDNRFVKGCSADLLWRMQLREKKFE---QGIPOQKATNI
RcLUS	31	PNAG-TPPEERAEVEEARONFYNKRFQVKPNSADLLWQLQFLREKNFK---QKIPKVKVEDG
KcMS	31	PEAG-TPPEERAEVEEARONFWRDRFRKPCSADLLWRFQFLREKKFK---QIIPQGVQDG
BgLUS	31	PEAG-TPPEERAEVEEARONFWRDRFLKPSADLLWRFQFLSEKKFK---QRIPOQKQDG
SS	31	PNYG-TPPEERDEVEEARLHFVNNRHEIKPSGDTLWRMQLFIREKKFK---QTIPOVKIEDD
KdFRS	31	PQAT-DEQQIAKVEEARLNFNHHRHKIKPSSADLLWRLOFLEEKDFR---QNTAQVKVEDG
KdLUS	31	PQAT-DEQQIAKVEEARLDFMHNRYKIKPNSADLLWRMQLFLEKAFT---QTIPOVKVEDG
KdGLS	31	PQAT-DEQQIAKVEEARLDFMHNRYKIKPNSADLLWRMQLFLEKDFR---QNTIPQKVEDG
EtAS	31	PQPP-TPQELAQVQCARLNFMNRMHVKPSADLLWRFQFLREKNFK---QTIPOAKINEG
RsM1	31	PDAG-TPPEERAKAEEARONFYNKRFQVKPSGADLLWRLOFLEKNFK---QTIPOVRIEAG
PNY1	32	PDYVAPGEELEVEQVRRQFWDNRYQVKPSGADLLWRMQLFLEKNFR---QTIPOVKVGDG
PNY2	31	PDYG-TPAERAEVEEARLHFVNNRYQVKPSADLLWRMQLFLEKNFK---QIIPQVKVEDG
KdTAS	50	PEAG-TPPEERAEVEEARLNFNRYVKNRQVKPSADLLWRMQLFLEKNFK---QTIIPVKVEDG
BPY	31	PQAG-SPEERAEVEEARNFYDNRYQVKPSGADLLWRMQLFLEKNFK---QTIIPVKVEDG
PSM	31	PDAG-TSQEKAQVEEARQHFYDNRFVVKACSADLLWRFQILKEKNFK---QTIESVKIKDE
AMY2	31	PDAG-TPPEERAEVEEARQDFYNNRYKVKTCGDRLLWRFQVMRENNFK---QTIIPSVKIEDG
AMY1	31	PEAG-SPEERAEVEEARNFYDNRFVVKPCGADLLWRFQVLRENNFM---QTIIDGVKIEDG
PSY	31	PEAG-SPEERAEVEEARNFYNNRYVKNRQVKPCGADLLWRFQVLRENNFK---QTIIGGVKIEDE
OSC1	31	PDAG-TPPEERAEVEEARQDFYNNRYKVKPCGADLLWRFQVLRENNFK---QTIIPSVKIEDG
GgbAS1	31	PDGG-TPPEERAEVEEARLHFVNNRYQVKPCGADLLWRFQILRENNFK---QTIASVKIGDG

ScLAS 66 KHSPDFSAFDACHNCASFFKLLQEPDSGLFFPCQYKGPMTIGYVAVNYIAGI---EIPE
 1W6K 71 -----HTAFECALNGMTFYVGLQA-EDGHWTGDYGGPLFLIPGLLITCHVA---RIPLPA
 BPX1 97 EEVREEA VGMTLRRAINFYSTIQA-DDGHWP GDYGGPMFLIPGLVITLSITGTLNAFLSK
 KdCAS 87 EDITTEDAVNTLRRAINFHSSTIQA-HDGHWP GDYGGPLFLMPGLVITLSITGALNAVLSK
 CAS1 86 DDVTEEMVETTLKRGIDFYSTIQA-HDGHWP GDYGGPMFLIPGLIITLSITGALNTVLSE
 PNX 86 EDISEDKVTITLKRAMSFYSTIQA-HDGHWP GDYGGPMFLMPGLVITLSITGVLNVVLSK
 PSX 86 EDVTEEA VATT LRRGLNFYSTIQA-HDGHWP GDYGGPMFLMPGLVITLSITGALNAVLT D
 GgCAS1 86 EDVTEETVKITLRRAINFHSSTIQA-HDGHWP GDYGGPMFLMPGLVITLSITGALNAVLT E
 OSC5 86 EDVTEEA VVT LRRAINFHSSTIQA-HDGHWP GDYGGPMFLMPDLVITLSITGALNAVLT D
 BPW 85 EVITTEEA VIT LRRSISFYSSIQA-HDGHWP GESAGPLFFLQPFVMA LYITGDLNTIFSP
 GgLUS1 88 ENITVEALVTTITRAISFYSSIQA-HDGHWP AESAGPLFFLQPLVMA LYITGSLDDVLGP
 OSC3 85 EKVTAEALITITIRRSITFYSSIQA-HDGHWP AESAGPLFFVQPLVMA LYITGSLDDVLGP
 TRW 87 EEVTNDAVTITLKR AISFYSTIQA-HDGHWP AESAGPLFFLPPVLIALYVTGAMNDILTP
 OEW 86 EGITTEEA VIT LRR AISFYSTIQA-HDGHWP AESAGPLFFLPPVLIALYVTGAINVLSR
 PEN5 88 EDITYEDAKTALRRGVLYAACQA-NDGHWP SEVSGSMFLDAPFVICLYITGHLEKIFTL
 PEN6 88 ENITYKDAKTALRRGILYFKALQA-EDGHWP AENSGLFFEAPFVICLYITGHLEKILTL
 PEN4 88 RKIKYEDAKTALRRGILYFTALQA-DDGHWP AENSGPNFYTPPFLICLYITGHLEKIFTP
 PEN1 88 AKIRYEDAKTALRRGILYFTALQA-DDGHWP AENSGPNFYTPPFLICLYITGHLEKIFTV
 RsM2 89 EEINYE TVTNAIRRS AHYLSATQA-SDGHWP ADASAPVFYIAPWVIGLYVIGHLNTVFP A
 LcIMS1 87 EEISYEMALDAMRRCAHFLAAIQA-SDGHWP SETSGPLFYVCPPLICMYIMGFMDKVFSP
 MiFRS 90 EQVTYEDVTTALRRASSFFSALQA-SDGHWP AENAGVSFFLPPFIFCLYITGHLSIITP
 LUP2 88 EGITYEKNATDALRRAVSFYSALQA-SDGHWP AEITGTLFFLPPLVFCFYITGHLEKIFDA
 LUP1 88 EEITYETTALRRGVRYFTALQA-SDGHWP GEITGPLFFLPPLI FCLYITGHLEEVFDA
 RcLUS 87 EEITSEIAAAALRRSVHLSALQA-SDGHWCAENGGLFFLPPLVFAVYITGHLNTVFP S
 KcMS 87 EEITREIATTAALRRSVHLSALQA-SDGHWCAENSGPMFFVPPMVFALYITGHLTTFVSA
 BgLUS 87 EEITREIATTAALRRSVHLSALQA-SDGHWCAENSGPMFFVPPMVFSLYITGHLNAVFS A
 SS 87 EEISYDKVTATMRRSVHLLAALQA-DDGHWP AENSGPSFFLQPLVMCLYITGHLSNVFP A
 KdFRS 87 EEVSYEAATAALRRGVHFYSALQA-SDGHWP AENAGPMFFMSPLVMCLYITGHLNTIFTE
 KdLUS 87 EEVSYEA VTAALRRGVHFYSALQA-SDGHWP AENAGPMFFMPPVMCLYITGHLNAIFTE
 KdGLS 87 EEVSYEA VTAALRRGVHFYSALQA-SDGHWP AENAGPMFFMPPVMCLYITGHLNAIFTE
 EtAS 87 EDITYEKATTAALRRAVHFFSALQA-SDGHWP AENAGPLFFLPPVMCLYITGHLDTVFP A
 RsM1 87 EEITREKATTAALRRAVHFFSALQA-SDGHWP AENAGPLFFLPPVMCMCITGHLDTVFP A
 PNY1 89 EAVTYEAATTTLRRAVHFFSALQA-SDGHWP AENSGPLFFLPPVMCVYITGHLDTVFP A
 PNY2 87 EEITYEAATTTLRRAVHFFSALQA-DDGHWP AENAGPLFFLPPVMCLYITGHLNTVFP A
 KdTAS 106 EEITYETATTAALRRAVHFFSALQA-SDGHWP AENSGPLFFLPPVMCLYITGHLNTVFP A
 BPY 87 EEITYEKSTAAALRRAVHFFSALQA-SDGHWP AENAGPLFFLPPVMCMYITGHLNTVFP A
 PSM 87 EEISEENVAITLRRAVHHSSTIQA-NDGHWP ALNAGPLFYFPPLVFCMYITGHLD SIFPY
 AMY2 87 EKVTYDKVTITVRRAAHHLAALQT-SDGHWPAQIAGPLFFTPPLIFCMYITGHLD S VFPE
 AMY1 87 EEITYEKATTTLRRCTHHLAALQT-SDGHWPAQIAGPLFFMPPLVFCVYITGHLD S VFPR
 PSY 87 EEITYEKTTTTLRRCTHHLAALQT-SDGHWPAQIAGPLFFMPPLVFCVYITGHLD S VFPP
 OSC1 87 EEITYEKATTTLKRAAHHLAALQT-SDGHWPAQIAGPLFFQPPLVFCMYITGHLD S VFPE
 GgbAS1 87 EEITYEKATTAARRAAHHLAALQT-SDGHWPAQIAGPLFFLPPLVFCMYITGHLD S VFPE

ScLAS	123	HERIEILRYIVNTAHPVDGGWGLH	SVDRKSTVFGT	VILRLGLPKD	-----HPVCAK
1W6K	122	GYREEIVRYLRSVQ-	LEDGGWGLHIEDKSTVFGT	ALNYVSLRILGVGPD	-----DPDLVR
BPX1	156	EHQCEICRYLYNHQ-	NEDGGWGLHIEG	PSTMFGTALNYITLRLILGE	-PE--DG-MGAVEK
KdCAS	146	EHKKEICRYLYNHQ-	NEDGGWGLHIEG	PSTMFGSVILNYVTLRLILGEDVN	--GG-DGEIER
CAS1	145	QHKQEMRRYLYNHQ-	NEDGGWGLHIEG	PSTMFGSVILNYVTLRLILGEGPN	--DG-DGDMEK
PNX	145	EHKREICRYLYNHQ-	NRDGGWGLHIEG	PSTMFGTVILNYVTLRLILGEGAN	--DG-QGAMK
PSX	145	EHRKEMRRYLYNHQ-	NKDGGWGLHIEG	PSTMFGSVLCYVTLRLILGEGPN	--DG-EGDMER
GgCAS1	145	EHRKEICRYLYNHQ-	NKDGGWGLHIEG	PSTMFGSVILNYVALRILGEGPN	--DR-QGEMEK
OSC5	145	EHRKEMCRYLYNHQ-	NKDGGWGLHIEG	PSTMFGSVILNYVTLRLILGEGPN	--DG-QGDMEK
BPW	144	AHQKEILRYLYNHQ-	NEDGGWGLHIEGH	STMFGSALS	YIALRILGEGLE-DGE-DGAMAK
GgLUS1	147	EHKKEIVRYLYNHQ-	NEDGGWGLHIEGH	STMFGSALS	YVALRILGEGP----Q-DKAMAK
OSC3	144	QHKKEILRYLYNHQ-	NEDGGWGLHIEGH	STMFGSALS	YIALRVLGQSLE-DGE-DMAVAR
TRW	146	AHQLEIKRYLYNHQ-	NEDGGWGLHIEGH	STIFGSVLS	YITLRLILGEEADSVAE--DMALVK
OEW	145	EHRKEITRYLYNHQ-	NEDGGWGLHIEGH	STMFGSVLS	YITLRLILGEGQE--DGE-DKAVAR
PEN5	147	EHVKEILLRYMYNHC-	NEDGGWGLHIEGH	SMFCTVILNYICLRILGVEPDH	DG-QKSACAR
PEN6	147	EHRKEILLRYMYNHC-	NEDGGWGLHIEGH	QSAMFCTVINYICLRILGVEAD	LDLDDIKGSGCAR
PEN4	147	EHVKEILLRHIYNHQ-	NEDGGWGLHIEGH	SMFCTVINYICLRILGVEVGH	DD-QRNGCAK
PEN1	147	EHRIEILLRYMYNHC-	NEDGGWGLHIEGH	SMFCTVINYICLRILGVEAGH	DDDDQGSTCTK
RsM2	148	EHQKEILRYIYCHO-	NEDGGWGLHIEGH	SMFCTALNYICMRILGEGPG	--GGRDNACER
LcIMS1	146	EHKKEIMRYIYCHO-	NEDGGWGLHIEGH	SMFCTTINYISLRLILGEEP	-----VEAVCK
MiFRS	149	EHRKEILLRHIYCHO-	NEDGGWGLHIEGH	STVATAITYICMRILGVGPD	-----EDACAR
LUP2	147	EHRKEMLRHIYCHO-	NEDGGWGLHIEGH	SKSMFCTVILNYICLRMLGEGPN	--GGRNNACKR
LUP1	147	EHRKEMLRHIYCHO-	NEDGGWGLHIEGH	SKSMFCTVILNYICLRMLGENPE	-----QDACKR
RcLUS	146	EHRKEILRYIYCHO-	NEDGGWGLHIEGH	STMFCTVILNYICMRILGEARD	--GGIENACER
KcMS	146	EHQKEILRYIYCHO-	NEDGGWGLHIEGH	STMFCTVILNYICMRILGEGRD	--GGKDNACER
BgLUS	146	EHQKEILRYIYCHO-	NEDGGWGLHIEGH	SMFSTVILNYINWLGKLGEGRD	--GGKDNACER
SS	146	EHRKEILRYIYSHQ-	NKDGGWGLHIEGH	SMFCTILSYICMRILGEGPD	--GGLNGACTR
KdFRS	146	EHRRETLRYIYYHO-	NEDGGWGLHIEGH	QSTMFCTVILNYICMRILGEGPE	--GGQDNAVSR
KdLUS	146	EHRSETLRYIYYHO-	NEDGGWGLHIEGH	STMFCTVILNYICMRILGEGPE	--GGQDNAVSR
KdGLS	146	EHRSETLRYIYYHO-	NEDGGWGLHIEGH	STMFCTVILNYICMRILGEGPE	--GGQDNAVSR
EtAS	146	PHRLEILRYIYCHO-	NEDGGWGLHIEGH	STMFCTVLSYICMRILGEGPN	--GGQDNACSR
RsM1	146	EHRKEILRYIYYHO-	NEDGGWGLHIEGH	STMFCTALNYICMRILGEGPN	--GGQDDACTR
PNY1	148	EHRKEILRYIYCHO-	NEDGGWGLHIEGH	STMFCTILSYICMRILGEGPD	--GGVNNACAR
PNY2	146	EHRTEILRYIYCHO-	NEDGGWGLHIEGH	STMFCTALS	YICMRILGEGRD--GGENNACAR
KdTAS	165	EHQREILRYIYYHO-	NEDGGWGLHIEGH	STMFCTALS	YICMRILGEGPD--GGLDNAVAR
BPY	146	EHQKEILRYIYYHO-	NEDGGWGLHIEGH	STMFCTALS	YICMRILGEGPD--GGQDNACAR
PSM	146	EYRKEILRYIYCHO-	NEDGGWGLHIEGH	SMFCTVILNYICMRILGEGPN	--GGKEDACAR
AMY2	146	VYRKEILRYIYVHO-	NEDGGWGLHIEGH	SMFCTVILNYICMRILGEGPD	--GGQDNACAR
AMY1	146	EHRKEILRYIYCHO-	NEDGGWGLHIEGH	STMFCTALNYICMRILGEGPD	--GGQDNACAR
PSY	146	EHRKEILRYIYCHO-	NEDGGWGLHIEGH	STMFCTALNYICMRILGEGPD	--GGEDNACVR
OSC1	146	EYRKEILRYIYVHO-	NEDGGWGLHIEGH	SMFCTALNYICMRILGEGPD	--GGQDNACAR
GgbAS1	146	EYRKEILRYIYYHO-	NEDGGWGLHIEGH	SMFCTALNYICMRILGEGPD	--GGQDNACAR

ScLAS	178	ARSTILRLGGALGSEPHWGKIWLSALNLYKWEGVNPAPPETWLLPYSLPMHPGRWWVHTRG
1W6K	176	ARNILHKKGGAVAIPSWGKEWLVAVLNVSWEGLNTLFPPEMWFDPWAPAH PSTLWCHCRO
BPX1	211	ARKWILDHGGATAITPSWGKMWLSVLGVYEWSCGNNPLPPEVWLCPPYLLPCHPGRMWCHCRM
KdCAS	202	ARKWILDHGGATAITPSWGKMWLSVLGVFEWCGNNPLPPEMWFPPYLLPCHPGRMWCHCRM
CAS1	201	GRDWILNHGGATNITPSWGKMWLSVLGAFEWSCGNNPLPPEIWLPLPYFLPIHPGRMWCHCRM
PNX	201	GRDWILDHGSATAITPSWGKMWLSVLGVFEWSCGNNPLPPEIWLPLPYILPIHPGRMWCHCRM
PSX	201	GRDWILDHGGATYITPSWGKMWLSVLGVFEWSCGNNPMPPEIWLPLPYALPVHPGRMWCHCRM
GgCAS1	201	GRDWILDHGGATFITPSWGKMWLSVLGVYEWSCGNNPLPPEIWLPLPYVLPPIHPGRMWCHCRM
OSC5	201	ARDWILDHGGATYITPSWGKMWLSVLGVFEWSCGNNPLPPEIWLPLPYALPFIHPGRMWCHCRM
BPW	201	SRKWILDHGGGLVAIPSWGKEWTVLGLYEWSCGNNPLPPEFWLPLDIFPIHPGKMLCYCRL
GgLUS1	201	GRKWILDHGGGLVAIPSWGKEWTVLGLAYEWSCGNNPLPPELWLLPKFAPFIHPGKMLCYCRL
OSC3	201	GRKWILDHGGGLVAIPSWGKEWTVLGVYEWSCGNNPLPPEFWLLPKIFPIHPGKMLCYCRL
TRW	204	GRKWILDHGGAVGIPSWGKEWTLILGVYEWGCGNPMPEFWLMPKFFPIHPGKMLCYCRL
OEW	202	GRKWILDHGGAVGIPSWGKEWTLVGVYEWDCGNPMPPEFWLLPNFSPFIHPGKMLCYCRL
PEN5	205	ARKWILDHGGATYAFMVAKAWLSVLGVYDWSGCKPLPPEIWMPLSPFSPINGGTLWIYIRD
PEN6	206	ARKWILDHGGATYTEPLIGKAWLSILGVYDWSGCKPIPPEVWMLPTEFSPFNGGTLWIYFRD
PEN4	205	AHKWILDHGGATYTEPLIGKAILSVLGVYDWSGCPMPPEFWLLPSSFPVNGGTLWIYLRD
PEN1	206	ARKWILDHGGATYTEPLIGKACL SVLGVYDWSGCKPMPPEFWLLPSSFPINGGTLWIYLRD
RsM2	205	ARKGILDHGGVTYIPSCGKTWLA MLGVFDWSGCGNPMPEFWMLPPEFPMHPAQMWCYCRI
LcIMS1	200	ARNWIHDHDCVTSIIPSWGKTWLSILNVEFDWSASNPMPPEYVWMLPTWVPIHPSNMMCYTRI
MiFRS	203	ARKWILDRGGITYMASWGKTWLSVLGIFDWYCGNPMPEFWILPSYLPPIHPAKMWCYCRM
LUP2	204	ARQWILDHGGCVTYIPSWGKTWLSILGIYDWSGCTNPMPEIWLPLPSFFPIHLGKTLCYTRM
LUP1	201	ARQWILDRGGVTHIPSWGKEWLSILGVYDWSGCTNPTEPPELLMPLSFLPIHPGKILCYSRM
RcLUS	203	GRKWILDHGGATGISSWGKTWLSILGVYEWDCGTPMPPEFWAFPSFPLHPAKMFCYCRI
KcMS	203	ARKWILDHGSATAISSWGKTWLA ILGVYEWDCGNPMPPEFWVFPTEFFPIHPAKMLCYCRL
BgLUS	203	ARFRILDHGSATAISSWGKTWLA ILGVYEWDCGNPMPPEFWAFPTFFPIHPARMMLCYCRL
SS	203	ARKWILDHGGAIANPSWGKVWLSILGVHEWVCGNPLPPEFWLFP SFPLMSPGKMWSYCRL
KdFRS	203	GRKWILDHGGATAIPSWGKTWLSIMGLCDWSGCGNPMPEFWLLPSYLPMPHPAKMWCYCRM
KdLUS	203	GRKWILDHGGATSIPSWGKTWLSIMGLCDWSGCGNPMPEFWLLPSYLPMPHPGKMWCYCRM
KdGLS	203	GRKWILDHGGATSIPSWGKTWLSIMGLCDWSGCGNPMPEFWLLPSYLPMPHPGKMWCYCRM
EtAS	203	ARKWILDHGGATYIPSWGKTWLSILGVYEWSCGNNPMPPEFWILPLFLPMHPAKMWCYCRM
RsM1	203	ARKWILDHGSVTNIPSWGKTWLSILGVYDWSGCGNPMPEFWMLPSFLPMHPAKMWCYCRM
PNY1	205	GRKWILDHGSVTAIPSWGKTWLSILGVYEWLIGSNPMPPEFWILPSFLPMHPAKMWCYCRM
PNY2	203	ARKWILDHGSVTAIPSWGKTWLSILGLEFDWSGSGNPMPEFWILPPEFPLPMHPAKMWCYCRM
KdTAS	222	GRKWILDHGVTVTAMP SWGKTWLSIMGLEFDWSGSGNPMPEFWLLPSFLPMYPAKMWCYCRM
BPY	203	ARKWILDHGGVTHMPSWGKTWLSILGIFEWLIGSNPMPPEFWILPSFLPMHPAKMWCYCRM
PSM	203	ARKWIHDHGSVTHVSSWGKTWLSVLGIFDWCASNMPMPPEFWMLPSFLPKHPAKMLCYCRL
AMY2	203	ARKWIHDHGGATHIASWGKTWLSILGIFDWSGCTNPMPEFWILPSFLPMHPAKMWCYCRM
AMY1	203	ARNWIRAHGCVTYIPSWGKTWLSILGLEFDWLGSNPMPPEFWILPSFLPMHPAKMWCYCRM
PSY	203	ARNWIRQHGCVTHIPSWGKTWLSILGVFDWLGSNPMPPEFWILPSFLPMHPAKMWCYCRM
OSC1	203	ARKWILDHGGVTHIPSWGKTWLSILGIFDWKGSNPMPPEFWILPSFLPMHPAKMWCYCRM
GgbAS1	203	ARKWIHDHGGVTHIPSWGKTWLSILGVFDWCGSNPMPPEFWILPSFLPMHPAKMWCYCRM

ScLAS	238	VYLPMSYLSLVKFSQPMTPLEEELRNELYTKPFDKINFSKNRNTVCGVDLYYPHSTTLNI
1W6K	236	VYLPMSYCYAVRLSAAEDPLVQSLRQELYVEDEASIDWLAQRNNVAPDELYTPHSWLLRV
BPX1	271	VYLPMSYLYGKRFVGPITSTIQSLRKELYTVPYHEIDWNKARNDCAKEDLYYPHPLVQDI
KdCAS	262	VYLPMSYLYGKRFVGPITPTVLSLRKELFTVPYHEIDWNEARSLCAKEDLYYPHPVVDI
CAS1	261	VYLPMSYLYGKRFVGPITSTVLSLRKELFTVPYHEVNWNEARNLCAKEDLYYPHPLVQDI
PNX	261	VYLPMSYLYGKRFVGPITPTVLSLRKEVFSVPYHEIDWNQARNLCAKEDLYYPHPLIQDI
PSX	261	VYLPMSYLYGKRFVGPITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDI
GgCAS1	261	VYLPMSYLYGKRFVGPITPTILSLRKELYTVPYHDIDWNQARNLCAKEDLYYPHPLVQDI
OSC5	261	VYLPMSYLYGKRFVGPITPTILSLRKELFTVPYHDIDWNQARNLCAKEDLYYPHPLVQDI
BPW	261	VYMPMSYLYGKRFVGPITGLIQSLRQELYNEPYHQINWNKARSTVAKEDLYYPHPLIQDL
GgLUS1	261	VYMPMSYLYGKRFVGPITALIRSLREELYNEPYNQINWNTARNTVAKEDLYYPHPLIQDM
OSC3	261	VYMPMSYLYGKRFVGPITALVRSLRKELYNEPYDRVDWNKARNTVAKEDLYYPHPLIQDM
TRW	264	VYMPMSYLYGKRFVGPITELVRDLRQELYDPEYDEINWNKARNTCAKEDLYYPHPFVQDM
OEW	262	VYMPMSYLYGKRFVGPITGLIQLRQELYTEPYHGINWNRARNTCAKEDLYYPHPLAQDM
PEN5	265	ILMGMSYLYGKRFVATPTLILQLREELYPOPYSKIINWSKARNRCAKEDLLYPKSFQDL
PEN6	266	IFMGVSYLYGKRFVATPTPLIILQLREELYPOPYDKILWSQARNQCAKEDLYYPQSFQEM
PEN4	265	TEMGLSYLYGKRFVATPTPLIILQLREELYPEPYAKINWTQTRNRCGKEDLYYPRSFQDL
PEN1	266	IFMGVSYLYGKRFVATPTPLIILQLREELYPEPYTKINWRLTRNRCAKEDLCYPSSFQDL
RsM2	265	VYMPMSYLYGKRFVGPITPLVQLREELHTOPHEIEWSKARHLCAKEDLFHRRPWIQEL
LcIMS1	260	TYMPMSYLYGKRFVQAPITPLVQLRDELHTOPYDQINWRKVRHMCATEDLYFPHFPVQDL
MiFRS	263	VYMPMSYLYGKRFVAPITPLIILQLREELHTOPYHEIEWRKMRHRCAEEDLYFPHSLIQNF
LUP2	264	VYMPMSYLYGKRFVGPITPLIMLRKELHLOPYEINWNKARRLCAKEDMIYPHPLVQDL
LUP1	261	VSTPMSYLYGKRFVGPITPLIILQLREELYLEPYEINWKKSRRLYAKEDMYAHPLVQDL
RcLUS	263	TYMPMSYLYGKRFVGPITPLIILQLREELYNEPYNKIKWNSVRHLCAKEDNYFPHPTIQKL
KcMS	263	TYLAMSYSYLYGKRFVGPITPLIILQLREELYNEPYDEINWSRMRHLCAKEDNHYPHTLTQII
BgLUS	263	TYMAMSYSYLYGKRFVGPITPLIILQLREELYNEPYDQINWSRMRHLCAKEDNYYAHTLTQII
SS	263	VYMPMSYLYGKRFVGPITPLVQLREELYAQPYNDIKWKSSRHVCAKEDIYYPHPLIQDL
KdFRS	263	VYMPMSYLYGKRFVTHITPLIILQLREELHTOPYDQINWKKVRHVCKEDTYYPHPLIQDL
KdLUS	263	VYMPMSYLYGKRFVARTITPLIILQLREELHTOPYDQIDWKKVRHVCKEDMYYPHPLIQDL
KdGLS	263	VYMPMSYLYGKRFVARTITPLIILQLREELHTOPYDQIDWKKVRHVCKEDMYYPHPLIQDL
EtAS	263	VYMPMSYLYGKRFVGPITPLIILQLREELHTOPYHHINWTKTRHLCAHEDVYYPHPLIQDL
RsM1	263	VYMPMSYLYGKRFVGLITPLIILQLREELHTOPYDQINWKKNCHQCAPEDLYYPHPLIQDL
PNY1	265	VYMPMSYLYGKRFVGPITPLIILQLREELYGOPYNEINWRKTRRVCAKEDIYYPHPLIQDL
PNY2	263	VYMPMSYLYGKRFVGPITPLIILQLREELYAQAYDEINWRKVRHNCAKEDLYYPHPLIQDL
KdTAS	282	VYMPMSYLYGKRFVGPITPLIILQLREELYDOPYEQNWKQVRHECAKEDIYYPHPLIQDL
BPY	263	VYMPMSYLYGKRFVGPITPLIILQLREELYTOPYHQVNWKKVRHLCAKEDIYYPHPLIQDL
PSM	263	VYMPMSYLYGKRFVGPITPLIILQLREELHTOPYEKVNWKKTRHLCAKEDLYYPHPLIQDL
AMY2	263	VYMPMSYLYGKRFVGPITPLIILQLREELHTOPYEKVNWKKARHQCACEDLYYPHPLIQDL
AMY1	263	VYMPMSYLYGKRFVGPITPLIILQLREELHTOPYEKINWTKSRHLCAKEDIYYPHPLIQDL
PSY	263	VYMPMSYLYGKRFVGPITPLIILQLREELHTOPYEKINWTKTRHLCAKEDIYYPHPLIQDL
OSC1	263	VYMPMSYLYGKRFVGPITPLIILQLREELHTOPYEKVNWKKARHQCACEDIYYPHPLIQDL
GgbAS1	263	VYMPMSYLYGKRFVGPITPLIILQLREELHTOPYEKVNWKKARHQCACEDLYYPHPLIQDL

ScLAS	298	ANSIIVVFYEKYLIR-----NRFYLSLKKKKVYDLIKTELQNTDSICTAPVNOAFCALVTLI
1W6K	296	VYALLNLYEHH-----HSAHLRQRAVQKLYEHTVADDFTKSISIGPISKTINMLVRWY
BPX1	331	LWASLYYAYEPTFMYWPAKRI-REKALDTVMQHIHYEDENTRYICIGPVNKVLNMLCCWA
KdCAS	322	LWATLHKVVEPVLNWPFGKLI-REKALCSAIEHIHYEDENTRYICIGPVNKVLNMLCCWV
CAS1	321	LWASLHKIVEPVLMRWPGANL-REKALRTAIEHIHYEDENTRYICIGPVNKVLNMLCCWV
PNX	321	LWASLDKVVVEPTFMHWPAKRI-REKSLRTVMQHIHYEDENTRYICIGPVNKVLNMLCCWV
PSX	321	LWATLHKFVEPVMHWPGKLI-REKALKTATEHIHYEDENTRYICIGPVNKVLNMLCCWV
GgCAS1	321	LWASLHKFLEPTLMHWPGKLI-REMAIKTATEHIHYEDENTRYICIGPVNKVLNMLCCWV
OSC5	321	LWASLHKVVEPVLQWPGKLI-REKALNSVMEHIHYEDENTRYICIGPVNKVLNMLCCWV
BPW	321	LWGFLHHVAEPVLTTRWPF SML-REKALKAAIGHVHYEDENSRYLICIGSVEKVLCLIACWA
GgLUS1	321	LWGFLYHVGERELNCPWPF SML-REKALEIATNHVHYEDENSRYLICIGSVEKVLCLIARWV
OSC3	321	LWGFLHHVGERVINTWPF SML-RQKALEVATNHVRYEDETTRYLCIGSVEKVLCLIARWV
TRW	324	VWGVLHNVEPVLTSRPISTL-REKALKVAMDHVHYEDKSSRYLCIGSVEKVLCLIATWV
OEW	322	LWGFLHHFAEPVLTTRWPF SKL-REKALKVAMEHVHYEDMNSRYLCIGSVEKVLCLIACWV
PEN5	325	FWGCVHMLSENLNRWPNKLFVROALRTMELVHYHDETHYITGACVAKPFHMLACWV
PEN6	326	FWKCVHILSENILNRWPNKLIROKALRTMELVHYQDEASRYFTGGCVPKPFHMLACWV
PEN4	325	FWKSVHMFSESTLDRWPNKLIROKALQSTMALHYHDESTRYITGGCLPKAFHMLACWI
PEN1	326	FWKGVHIFSESTLNRWPNKLIROKALRTMELVHYQDEANRYITGGCVPKAFHMLACWV
RsM2	325	FWDCVHTFAEPVLTTRWPFNNFIREKALKITMEHVHYDDKASHYINPGSVEKVICMVAACWV
LcIMS1	320	LWDTLYLLEPPLMTRWPFNKLIRKALNETMRHIHYEDENSRYITIGSVEKPLCMLACWV
MiFRS	323	LWDSLYVASEPPLTRWPF SKL-REKALEKAMEHIHYEDENSRYITIGSVEKALCMLCCWV
LUP2	324	LWDTLHNFVEPILTNWPLKLVREKALRVAMEHIHYEDENSHYITIGSVEKVLCLACWI
LUP1	321	LSDTLQNFVEPPLTRWPFNKLIRKALQITMKHIHYEDENSHYITIGSVEKVLCLACWV
RcLUS	323	LWDALYTFSEPLESRWPFNKL-REKALKITMDHIHYEDENSRYITIGSVEKPLCMLACWI
KcMS	323	LWDALYLLSEPLLKRWPF SKL-REKALKITMDHIHYEDENSRYITIGSVEKPLNMLACWH
BgLUS	323	LWDALYMLGEPPLKRWPFNKL-REKALKITMDHIHYEDENSRYITIGSVEKPLMLACWH
SS	323	MWDSLYLLEPPLTRWPFNKL-REKALATMRHIHYEDENSRYITIGSVEKILCMLACWD
KdFRS	323	IWDTLYLLEPPLTRWPFNKLIRKALKKTMKHIHYEDENSRYITIGSVEKVLCLACWV
KdLUS	323	LWDTLYLLEPPLTRWPFNKLIRKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
KdGLS	323	LWDTLYLLEPPLTRWPFNKLIRKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
EtAS	323	MWDSLYLLEPPLTRWPFNKLIRKALEVTMKHIHYEDENSRYITIGSVEKVLCLACWA
RsM1	323	IWDCVHTSMEPPLTRWPFNKLIRKALEVTMKHIHYEDGSSRYITIGSVEKVLCLACWV
PNY1	325	LWDSLYVLEPPLTRWPFNKL-REKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
PNY2	323	MWDSLYLLEPPLTRWPFNKL-REKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
KdTAS	342	LWDTLYLLEPPLTRWPFNKLIRKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
BPY	323	LWDSLYLLEPPLTRWPFNKLIRKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
PSM	323	IWDSLYLLEPPLTRWPFNKLIRKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
AMY2	323	MWDSLYLLEPPLTRWPFNKLIRKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV
AMY1	323	IWDSLYLLEPPLTRWPFNKLIRKALEVTMKHIHYEDENSRYITIGSVEKVLCLACWV
PSY	323	IWDSLYLLEPPLTRWPFNKLIRKALEVTMKHIHYEDENSRYITIGSVEKVLCLACWV
OSC1	323	MWDSLYLLEPPLTRWPFNKLIRKALEVTMKHIHYEDENSRYITIGSVEKVLCLACWV
GgbAS1	323	IWDSLYLLEPPLTRWPFNKLIRKALQITMKHIHYEDENSRYITIGSVEKVLCLACWV

ScLAS	353	EEGVDSEAFQRLQYRFKDALEHGPOGMTIMGTNGVQIWDCAFATQYFFVAGL--AERPEF
1W6K	350	VDGPASTAFQEHVSRI PDYLW MGLDGMKMQGTNGSOLWDTAFATQALLEAGG--HHRPEF
BPX1	390	-EDPNSEAFKIHLPRIIDYIWI AEDGMKMQGTNGSOLWDTTFAVQAIIST----NI AEEY
KdCAS	381	-EDPNSEAFKIHLPRIYDYIWI AEDGMKMQGTNGSOLWDTAFSVQAIIVAT----KLVEEF
CAS1	380	-EDPNSEAFKIHLPRIHDFIWI AEDGMKMQGTNGSOLWDTGFATQAILLAT----NLVEEY
PNX	380	-EDPNSEAFKIHLPRIHDFIWI AEDGMKMQGTNGSOLWDTAFAVQAIIST----NL AEEY
PSX	380	-EDPNSEAFKIHLPRIYDYIWI AEDGMKMQGTNGSOLWDTAFAAQAIIST----NLIDEF
GgCAS1	380	-EDPNSEAFKIHLPRIYDYIWI AEDGMKMQGTNGSOLWDTAFTAQAIISS----NLIEEY
OSC5	380	-EDPNSEAFKIHLPRIYDYIWI AEDGMKMQGTNGSOLWDTAFAAQAIIST----NLIEEY
BPW	380	-EDPNGEAYKIHLPRI PDYIWI AEDGLKIQSF-GCOMWDAGFATQAILSC----NLNEEY
GgLUS1	380	-EDPNSEAYKIHLPRI PDYIWI AEDGLKIQSF-GCOMWDAAFATQAILAC----NVSEEY
OSC3	380	-EDPNSEAYKIHLPRI PDYIWI AEDGLKIQSF-GCOMWDAAFATQAILSG----NVSEEY
TRW	383	-EDPNGDAYKRHLARI PDYIWI AEDGMKMQSF-GCOMWDAAFATQAI FSS----NLTEEY
OEW	381	-EDPNSEANKRHLPRI PDYIWI AEDGLKIQSF-GCOMWDAAFATQAILSS----NL AEEY
PEN5	385	-EDPDGDYFKKHLARV PDIWI AEDGLKIQFLM-GMOSWNAALSLOVMLAA----NMDDEI
PEN6	386	-EDPDGDYFKKHLARV PDIWI AEDGLKIQSF-GSOLWDTAFSLQVMLAYQDVDDDDDEI
PEN4	385	-EDPKSDYFKKHLARV PDIWI AEDGLKIQSF-GSOLWDTALSHALLDGDIDHDVDDDEI
PEN1	386	-EDPEGDYFKKHLARV PDIWI AEDGLKIQSF-GSOLWDTVMSLHFLLDGVED-DVDDEI
RsM2	385	-EDPSCGFQHLARI SDYIWI AEDGMRTGI-GSQTWDAALSIQAILAC----NLIEEM
LcIMS1	380	-EDPNSEYVKKHLARI PDYIWI AEDGMKMQSF-GSQSWDAALAMQALLSC----NITREI
MiFRS	382	-EDPNGEYFKKHLARI PDYIWI AEDGMKIQSF-GSOLWDTAFGFQALVAS----NLTDEV
LUP2	384	-ENPNGDHFKKHLARI PDIWI AEDGLKIQSF-GSOLWDTVFAIQALLAC----DLSDET
LUP1	381	-ENPNGDYFKKHLARI PDYIWI AEDGMKMQSF-GCOLWDTGFATQALLAS----NLPDET
RcLUS	382	-EDPHGEAFKKHLARI ADYIWI AEDGLKIQSF-GSQTWDTSLALQALLIAS----DLSHEI
KcMS	382	-EDPNGDAFKKHLARI SDYIWI AEDGMKIQSF-GSQAWDTSEFVLQALLIAS----NLLSET
BgLUS	382	-EDPNGDAFKKHLARI PDYIWI AEDGLKIQSF-GSQVWDTSEFVLQALLIAS----NLPSET
SS	382	-EDPNGVCFKKHLARI PDYIWI AEDGMKIQSF-GSQVWDASIGIQALLAT----ELTHDI
KdFRS	383	-EDPNGDYFKKHLARV PDIWI AEDGMKIQSF-GSCHWDTAFVQALLAS----DMTDEI
KdLUS	383	-EDPNGDYFKKHLARI PDYIWI AEDGMKMQSF-GSCHWDTAFSIQALLAS----NMAEEI
KdGLS	383	-EDPNGDYFKKHLARV PDIWI AEDGMKIQSF-GSCHWDTVFSQAALLAS----DMADEI
EtAS	383	-EDPNGVDFKKHLARI PDYIWI AEDGMKMQSF-GSQQWDTGFATQALLAS----NLTEEI
RsM1	383	-EDPNGDYFKKHLARI PDYIWI AEDGMKMQSF-GSQQWDTGFATQALLAT----NLTDEI
PNY1	384	-EDPNGDYFKKHLARI PDYIWI AEDGMKMQSF-GSQEWDTFESIQALLDS----DLTHEI
PNY2	382	-EDPNGDYFKHLARI PDYIWI AEDGMKMQSF-GSQEWDTFATQALLAS----DLIDEI
KdTAS	402	-EDPNGDYFKKHLARV PDIWI AEDGMKMQSF-GSQQWDTGFATQALLAS----NMSDEI
BPY	383	-EDPNGDYFKKHLARI PDYIWI AEDGLKIQSF-GSQEWDTFATQALLAS----NLTDEI
PSM	383	-EDPNGDAFKKHLARI PDYIWI AEDGMTHISF-GSQTWDASLIHQALLAT----NLIEDV
AMY2	383	-EDPNGIAFKRHLARV PDIWI AEDGMCMQSF-GSQEWDAGFAVQALLST----NLIDEL
AMY1	383	-EDPNGDAYKKHLARV QDYIWI AEDGMTHISF-GSQEWDAGFAVQALLAA----NLNDEI
PSY	383	-EDPNGDAFKKHLARV PDIWI AEDGMTHISF-GSQEWDAGFAVQALLAT----NLIEEI
OSC1	383	-EDPNGDAFKKHLARI PDYIWI AEDGMCMQSF-GSQEWDAGFAVQALLAT----NLVDEL
GgbAS1	383	-EDPNGDAFKKHLARV PDIWI AEDGMTHISF-GSQEWDAGFAVQALLAT----NLVEEI

ScLAS 411 YNTIVSAAYKFLCHAOFDTECV---PGSYRDKRKGAWGFSTKTQGYTVADCTAEAIKAIIM
 1W6K 408 SSCLQKAHEFLRISQVDPNPP-DYQKYRQMRKGGFSFSTLD CGWIVSDCTAEALKAVLL
 BPX1 445 GQTLRKAHEFIKDSQVLEDCPGDLNFWYRHISKGAWPFSTADHGWPI SDCTAEGLKAVIL
 KdCAS 436 SSTISKAHEFMKNSQVLEEDYFGDLSYWRHISKGAWPFSTADHGWPI SDCTAEGLKVVLLK
 CAS1 435 GFVLEKAHSFVKN SQVLEDCPGDLNFWYRHISKGAWPFSTADHGWPI SDCTAEGLKAAALL
 PNX 435 GPTLRKAHTFMKNSQVLEDCPGDLDAWYRHISKGAWPFSTADHGWPI SDCTAEGFKAVLQ
 PSX 435 GPTLKAHAHF IKN SQVSEDCPGDL SKWYRHISKGAWPFSTADHGWPI SDCTAEGLKAVLL
 GgCAS1 435 GPTLRKAHTY IKN SQVLEDCPGDL SKWYRHISKGAWPFSTADHGWPI SDCTAEGLKAVLL
 OSC5 435 GPTLRKAHTFIKN SQVLEDCPGDLN KWYRHISKGAWPFSTADHGWPI SDCTAEGLKAILS
 BPW 434 WPTLRKAHEFVKASQVPE NPSGDFKAM YRHIN KGAWTF SMQDHGWQV SDCTAEGLKVAIIL
 GgLUS1 434 GPTLRKAHHEFVKASQVRENPSGDFNAMYRHISKGAWTF SMHDHGWQV SDCTAEGLKAAALL
 OSC3 434 GPTLKAHHEFVKASQVRENPSGDFKAMYRHISKGAWTF SMHDHGWQV SDCTAEGLKVALL
 TRW 437 GPTLKAHHEFVKASQVRDNPPGDFSKMYRHISKGAWTF SIQDHGWQV SDCTAEGLKVSL
 OEW 435 GPTLMKAHNEFVKASQVQENPSGDFENEMYRHISKGAWTF SMQDHGWQV SDCTAEGLKAAALL
 PEN5 439 RSTLIKGYDFIKK SQISENPGDHLKMRDITKGGWTFQDREOGLPI SDGTAESTIECCIH
 PEN6 444 RSTLIKGYSEFNK SQITONPPGDH RKM LKDI AKGGWTFSDQDQGWV SDCTAESLECCLV
 PEN4 443 KTTLVKCYDYIKK SQITENPRGDHFKMFRHITKGGWTFSDQDQGWV SDCTAESLECCLF
 PEN1 443 RSTLVKCYDYIKK SQVTENPPSDHIKMRHISKGWTFSDKQDQGWV SDCTAEGLKCCLL
 RsM2 439 GPTLKKGYDFIKK SQAKDNPPGDFKRM YRHF GKGAWAFSSQDYGVIAL DCTAESLMCCLLH
 LcIMS1 434 GSVLNSGHDFIKK SQVRNPPG DYKSMRYMSKGSWTFSDCDHGWQV SDCTAENLKCCLL
 MiFRS 436 APTLVKAYDFIKK SQVRDNPSGDFNEFKMRHISKGSWTFSDQDHGWQL SDCTAEALKCCLL
 LUP2 438 DDVLRKGHSFIKKSQVRENPSGDFKSMYRHISKGAWT LSDRDHGWQV SDCTAEALKCCML
 LUP1 435 DDALKRGHNYIKASQVRENPSGDFRSMYRHISKGAWTFSDRDHGWQV SDCTAEALKCCLL
 RcLUS 436 GPTLKQGHVFIKKSQATENPSGDFRKMFRHISKGAWTFSDKQDQGWQV SDCTAESLKCCLL
 KcMS 436 APTLEKGHNF IKN SQVTENPSGDFRMRHISKGSWTFSDKDHGWQV SDCTAESLKCCLL
 BgLUS 436 GPTLEKGHNF IKN SQVTQNP SGDFRMRHISKGSWTFSDKDHGWQV SDCTAESLKCCLL
 SS 436 APTLKKGHEFIKASQVRDNPSGDFKSMYRHISKGSWTFSDQDHGWQL SDCTTIGLTCCLL
 KdFRS 437 RTTLAKAHD CIKKSQVKDNPSGDFKSMYRHISKGAWTFSDQDHGWQL SDCTAEGLKCCLL
 KdLUS 437 GITLAKGHDF IKN SQVKDNPSGDFKSMYRHISKGAWTFSDQDHGWQV SDCTAEGLKCCLL
 KdGLS 437 GTTLAKAHYCIKESQVKDNPSGDFRSMYRHISKGSWTFSDQDHGWQL SDCTAEGLKCCLL
 EtAS 437 GQVLKKGHDF IKN SQVKENPSGDFKSMHRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL
 RsM1 437 GGVLRRGHDF IKN SQVDNPSGDFKSMYRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL
 PNY1 438 GPTLMKGHDF IKN SQVKDNPSGDFKSMYRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLI
 PNY2 436 RPTLMKGHDF IKN SQVKENPSGDFKSMHRHISKGSWTFSDQDHGWQV SDCTAEALKCCLL
 KdTAS 456 GETLAKGHDF IKN SQVKDNPSGDFKSMHRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL
 BPY 437 GPTLARGHDF IKN SQVKDNPSGDFESMHRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL
 PSM 437 GPTLTKAHEFIKKSQVRDNPSGDFKSMYRHISKGSWTFSDKDHGWQV SDCTAESLKCCLL
 AMY2 437 GFALAKGHDF IKN SQVKDNPSGDFKSMHRHISKGAWTFSDQDHGWQV SDCTAEGFKCCLL
 AMY1 437 EFALAKGHDF IKN SQVTENPSGDFKSMHRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL
 PSY 437 KFALAKGHDF IKN SQVTENPSGDFKSMHRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL
 OSC1 437 GPTLAKGHDF IKN SQVRDNPSGDFKSMHRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL
 GgbAS1 437 APTLAKGHDF IKN SQVRDNPSGDFKSMYRHISKGSWTFSDQDHGWQV SDCTAEGLKCCLL

ScLAS 468 VKNSEVFSEVHHMISSEERLEFEGIDVLLNLQNIQSFYEGSFATYEKIKAPLAMETLNPAEV
 1W6K 467 LQEKCPH--VTEHLPRELRCDAVAVLLNMRN----PDGGFATYETKRGGHLELELLNPSEV
 BPX1 505 LSQFPSE--TVGKSVQVVKRLYDAVHVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 KdCAS 496 LSQFPSE--TVGAPLSAKLVYNAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 CAS1 495 LSKVPKA--TVGEPIDAKRLYEAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PNX 495 LSKLPSE--TVGEPIDAKRLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PSX 495 LSKLAPSE--TVGEPIDSKRLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 GgCAS1 495 LSKLAPSE--TVGEPIDAKRLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 OSC5 495 LSKLAPSE--TVGEPIDAKRLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 BPW 494 FSQMPPE--TVGKLEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 GgLUS1 494 LSEMPSE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 OSC3 494 LSEMSDE--TVGAKMTEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 TRW 497 YSQMNEK--TVGKVEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 OEW 495 FSQMPTE--TVGAELEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PEN5 499 FHRMPESE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PEN6 504 FGSMPSE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PEN4 503 FESMPSE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PEN1 503 FERMPSE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 RsM2 499 FSMMPPE--TVGKLEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 LcIMS1 494 LSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 MiFRS 496 AATMPPE--TVGKLEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 LUP2 498 LSMMPAE--TVGKIDPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 LUP1 495 LSMMSAD--TVGKIDPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 RcLUS 496 FSMMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 KcMS 496 FSMMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 BgLUS 496 FSMMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 SS 496 LSTMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 KdFRS 497 FSLMQPE--TVGEMPPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 KdLUS 497 FSMMPPE--TVGEMPPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 KdGLS 497 FSLMQPE--TVGEMPPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 EtAS 497 FSMMPPE--TVGEMPPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 RsM1 497 FSMMPPE--TVGEMPPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PNY1 498 FSTMPEE--TVGKLEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PNY2 496 FSRMPTE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 KdTAS 516 FSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 BPY 497 FSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PSM 497 LSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 AMY2 497 LSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 AMY1 497 LSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 PSY 497 LSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 OSC1 497 LSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET
 GgbAS1 497 LSLMPPE--TVGKMEPERLYDAVNVLLSLQNIQSFYEGSFATYELTRSYHWLELNPAAET

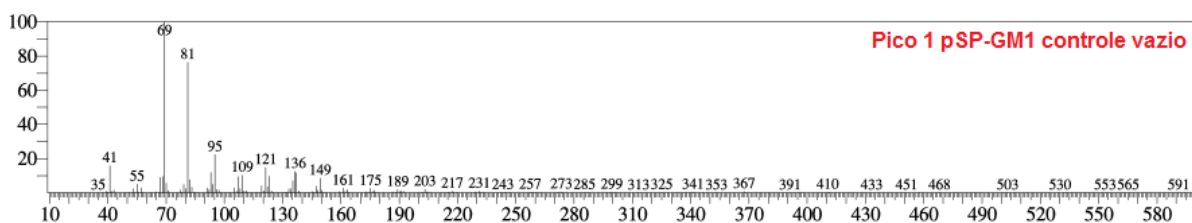
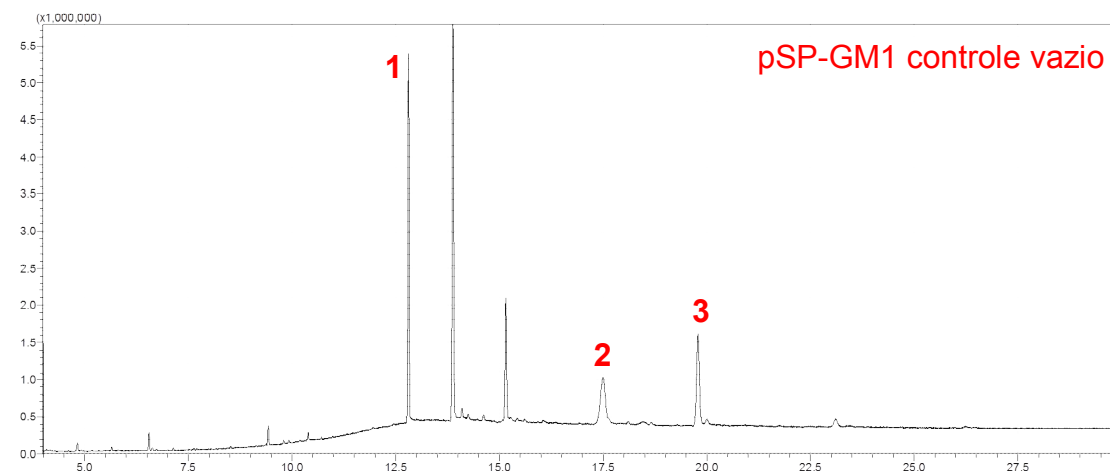
ScLAS 528 FGNIMVEYFYVECTDSSVLGLTYEHHKYF-DYRKEETRTRIRIAIEFIKKSQLPDGSWYGS
 1W6K 521 FGDIMIDYTYVECTSAVMQALKYFHKRFPEHRAAEIRETLTOGLEFCRROQRADGSWEGS
 BPX1 560 FGDIVIDYPYVECTSAAIQALTLFKKLHPGHRREIEIENCIAKAAEFIENIQASDGSWYGS
 KdCAS 551 FGDIVIDYPYVECTSAAIQSLVLFKKLHPGHRKEEVELCKKAAAFIEKIQESDGSWYGS
 CAS1 550 FGDIVIDYPYVECTSAAIQALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGS
 PNX 550 FGDIVIDYPYVECTSAAIQALTAFFKLLPGHRREIQHSIEKAALFIEKIQSSDGSWYGS
 PSX 550 FGDIVIDCPYVECTSAAIQALATFGKLYPGHRREIQCCIEKAVAFIEKIQASDGSWYGS
 GgCAS1 550 FGDIVIDYPYVECTSAAIQALTSFKKLYPGHRREIQCCIEKAASFIEKIQASDGSWYGS
 OSC5 550 FGDIVIDYPYVECTSAAIQALTSFRKLYPGHRREIQHSIEKAAAFIEKIQSSDGSWYGS
 BPW 549 FEDTLIEREYVECTSPAHHGLALFRKLYPRHRGTEIDSSLYRGTQYIEDVQEPDGSWYGH
 GgLUS1 549 FEDTMIEREYVECTGSAMQGLALFRKQEPCHRSKEIDRCIAKAIRYIENMONPDGSWYGC
 OSC3 549 FEETLIEREYVECTGSAMQALALFRKLYPKHRRKEIDRCISKAIRYIENTONPDGSWYGC
 TRW 552 FEDVLIEREYVECTSSAIQGLTLFKKLHPGHRRTKEIEHCISRASKYVEDTQESDGSWYGC
 OEW 550 FEDVLIEREYVECTSSAIQALKLFLKLPGHRREIASCISKAIQYIEATONPDGSWDGS
 PEN5 554 VENTVVEQEYLECTGSAIAGLVCFKKEFPDHRPKEIEKLIKGLKYIEDIQMPDGSWYGN
 PEN6 559 MEDTIVEHEHYVECTGSAIVALARLKEFPPEHRRREVEKFIKNAVKYIESFQMPDGSWYGN
 PEN4 555 -----MIFRYVECTGSAIAALTOFNKQFPGYKNVEKRFITKAQYIEDMQTVDGSWYGN
 PEN1 558 VQDITVEIEHYVECTGSAIVALTOFSKQEPERKKEVERFITNGVKYIEDIQMKDGSWCGN
 RsM2 554 FENIVVEHEHYVEVTASAINALVMFKKRYPGYREKEIEHFISKAVHYLIQOTQFPNGPWYGV
 LcIMS1 549 LEDLIEHQHVECTSSAIQALTLFRKQYPGHRREINNFINKAVQFLQDIQLPDGSWYGN
 MiFRS 552 LENIVIEHTYLECTGSSIIAFITLKKLPGHRTKDIDNFIVNAIRYLEDQYPDGSWYGN
 LUP2 553 FTCVMAEREYVECTSAVIQALVLFKQLYDPDRHTKEIKSIEKGVQFIESKQTPDGSWHGN
 LUP1 550 MANTMVEREYVECTSSVIQALDLFRKLYPDRHKEINRSIEKAVQFIQDNOTPDGSWYGN
 RcLUS 551 MEDIVVEHEHYVECTSSAIQALVLFKLYPRHRNKEIENCIINAAQFIENIQEPDGSWYGN
 KcMS 551 LEDIVIEHEHYVECTSSSVQALVLFKLYPEHRRKEIENFIVNAVRFIEEIQKPDGSWYGN
 BgLUS 551 LADIVIEHEHYVECTSSSIQALVLFKLYPEHRRKEIEIFILNAVRFTEEIQQPDGSWYGN
 SS 551 IEDIVIEHEHYVECTGTGMEALVLFKLYPKHRTKEVESFLTNAARYLDNTQMPDGSWYGE
 KdFRS 552 FENIVIEHEHYVECTSSAVQALVLFKLYPLHRRKEVERFITNGAKYLEDIQMPDGSWYGN
 KdLUS 552 FENIVIEHEHYVECTSSAVQALVLFKLYPLHRRKEVERFITNGAKYLEDIQMPDGSWYGN
 KdGLS 552 FENIVIEHEHYVECTSSAVQALVLFKLYHPGHRREIENFIVNAVRFIEEIQMPDGAWYGN
 EtAS 552 FADIVIEHEHYVECTASAIHALIMFKKLYPGHRKKEIENFIVNAVRFIEEIQMPDGAWYGN
 RsM1 552 FADIVIEHEHYVECTSSAIHALVLFKLYPGHRKKEIEDFIAKSVRFLESIQTSDGTWYGN
 PNY1 553 FADIVIEHEHYVECTSSAIQALVLFKLYPGHRKKEIDNFITNAVRYLEDQMPDGSWYGN
 PNY2 551 FEDIVIEHEHYVECTSSAIQALVMFKKLYPGHRKKEIEVSTITNAVRYLEDQMPDGSWYGN
 KdTAS 571 FADIVIEHEHYVECTASAIQALVLFKLYPGHRKKEIETFIKGAAQYIEDRQMPDGSWYGS
 BPY 552 FADIVIEHEHYVECTASAMQTLVLFKLYPGHRKKEIENFIKNAQFLQVIQMPDGSWYGN
 PSM 552 LEHIVVEREYVECTSSAIQALVLFKLYPEHRRKEIENFIANAVRFLEYKQTSDGSWYGN
 AMY2 552 FEDIVIEHEHYVECTGSAIICALVLFKNHYPEHRRKEIEDCIANAVRYFEDIQTADGSWYGN
 AMY1 552 FADIVVEHEHYVECTGSAIQALVLFKLYPGHRKKEIENFISEAVRFIEDIQATADGSWYGN
 PSY 552 FADIVVEHEHYVECTGSAIQALVLFKLYPGHRKKEIENFIFNAVRFLEDTQTEDGSWYGN
 OSC1 552 FADIVVEHEHYVECTGSAIICALVLFKLYPGHRKKEIENFISEAVRFLEDTQATADGSWYGN
 GgbAS1 552 FADIVVEHEHYVECTGSAIQALVLFKLYPGHRKKEIENFIANAVRFLEDTQATADGSWYGN

ScLAS	587	WGICFTYAGMFALEALHTVGETYE---NSSTVRKGCDFLVSKOMKDGGWGESMKSSSELHS
1W6K	581	WGVCFYGTWFGLEAFACMGQTYRDGTACAEVSRACDFLLSRQADGGWGEDFESCEERR
BPX1	620	WGVCFYAGWFGIKGLVAAGRTYK---NCSSIHKACDYLLSKELASGGWGESYLSQDKV
KdCAS	611	WAVCFYGTWFGVIGLVAAGRNFK---NSPSTRKACDFLLSKQLASGGWGESYLSQDKV
CAS1	610	WAVCFYGTWFGVIGLVAAGKTLK---NSPHVAKACEFLLSKQOPSGGWGESYLSQDKV
PNX	610	WGVCFYGTWFGIKGLVITAGRTFS---SCASIRKACDFLLSKQVASGGWGESYLSQDKV
PSX	610	WGVCFYGTWFGIKGLIAAGKNTFS---NCLSTRKACEFLLSKQLPSGGWGESYLSQDKV
GgCAS1	610	WGVCFYGTWFGVIGLVAAGKSEFN---NCSSIRKACEFLLSKQLPSGGWGESYLSQDKV
OSC5	610	WGVCFYGTWFGVIGLVAAGKSEFS---NCSSIRKACEFLLSKQLPSGGWGESYLSQDKV
BPW	609	WGICYTYGTWFVAVGALAAAGRNFK---NCPALRKSCDFLLSKQLPNDGGWGESYLSQDKV
GgLUS1	609	WGICYTYGTWFVAVEGLTACGKNCH---NLSLRKACDFLLSKQLPNDGGWGESYLSQDKV
OSC3	609	WGICYTYGTWFVAVEGLTACGKNHQ---NSVTLRRACKFLLSKQLPNDGGWGESYLSQDKV
TRW	612	WGICYTYGTWFVAVDALVACGKNYH---NCPALQKACKFLLSKQLPDGGWGESYLSQDKV
OEW	610	WGVCFYGTWFVAVEGLVACGKNYH---NSPTLRRACEFLLSKQLPDGGWGESYLSQDKV
PEN5	614	WGVCFYGTTFVAVRGLAAAGKTEG---NSEAIRRAVQFILNTQNAEGGWGESYLSQDKV
PEN6	619	WGVCFMYGTTFVAVRGLVAAGKTYQ---NCEPTRKAVQFILETONVEGGWGESYLSQDKV
PEN4	610	WGVCFIYGTTFVAVRGLVAAGKTYFS---NCEPTRKAVRFLDTONPEGGWGESYLSQDKV
PEN1	618	WGVCFIYGTTFVAVRGLVAAGKTEH---NCEPTRKAVRFLDTONPEGGWGESYLSQDKV
RsM2	614	WGVCFMYGTTFALKGLAAAGNTYA---NCPALPKAVDFLLKTCQDGGWGESYLSQDKV
LcIMS1	609	WGICYTYGTWFALKALSMAGKTYE---NCEAVRKGANFLRKIQNPEGGWGESYLSQDKV
MiFRS	612	WGICFTYSTMFALGGLAAAGRTYK---NCQAVRRGVDFLLINQSDGGWGESYLSQDKV
LUP2	613	WGICFTIYATWFALGGLAAAGKTYK---SCLAVRKGVDLFLAIQEEEDGGWGESYLSQDKV
LUP1	610	WGVCFIYATWFALGGLAAAGRTYN---DCLAMRNGVHFLLTQORDGGWGESYLSQDKV
RcLUS	611	WGICFSYGTWFALKGLAAAGRTYE---NCSAIRKGVDFLLKSCRDGGWGESYLSQDKV
KcMS	611	WGICFLFGTWFGIKGLATAGKTYY---NCTAVRKGVEFLLRTQREDSGGWGESYLSQDKV
BgLUS	611	WGICFLSCTWFGIKGLAAAGKTYY---NCTAVRKGVEFLLQTCRDGGWGESYLSQDKV
SS	611	WGICFTYGTYYALGGLAAAGKTYE---NCQSTRKAVRFLDKTCGEDGGWGESYLSQDKV
KdFRS	612	WGVCFYTCAWFALEGLSAAGKTYN---NCAAVRKGVDLFLNIQLEDGGWGESYLSQDKV
KdLUS	612	WGVCFYTCAWFALEGLSAAGKTYN---NCAAVRKGVDLFLNIQLEDGGWGESYLSQDKV
KdGLS	612	WGVCFYTCAWFALGGLAAAGKTYN---NCAAVRKGVDLFLRIQLEDGGWGESYLSQDKV
EtAS	612	WGVCFYGTWFVAVGGLAAAGKTYN---NCAAMRKAVDFLLRTQKQDGGWGESYLSQDKV
RsM1	612	WGVCFYGTWFVAVGGLAAAGKTYN---SCLAMRKAVDFLLRIQKDDGGWGESYLSQDKV
PNY1	613	WGVCFYGSWFALGGLAAAGKTYY---NCAAVRKAVEFLLKSCMDGGWGESYLSQDKV
PNY2	611	WGVCFYGTWFVAVGGLTAAGKTYN---NCQTLHKAVDFLLKSCRSDDGGWGESYLSQDKV
KdTAS	631	WGVCFYGTWFVAVGGLAAAGKTYD---NCAAIRKCTEFLNTOCENGGWGESYLSQDKV
BPY	612	WGVCFYGTWFVAVGGLAAAGKTYN---NCLAVRRAVDLFLRAQRDNGGWGESYLSQDKV
PSM	612	WGICFTYGSWFALNGLVAAGKTYD---NCAAIRKGVDFLLITQREDDGGWGESYLSQDKV
AMY2	612	AGICFTIYGTWFALGGLAAAGKTYA---NCAAIRKGVDFLLITQSKDGGWGESYLSQDKV
AMY1	612	WGVCFYGSWFALGGLAAAGKTYT---NCAAIRKAVKDFLLITQREDDGGWGESYLSQDKV
PSY	612	WGVCFYGSWFALGGLAAAGKTYT---NCAAIRKGVDFLLITQREDDGGWGESYLSQDKV
OSC1	612	WGVCFYGSWFALGGLAAAGKTYA---NCAAIRKAVKDFLLITQREDDGGWGESYLSQDKV
GgbAS1	612	WGVCFYGSWFALGGLAAAGKTYA---NCAAIRKAVKDFLLITQREDDGGWGESYLSQDKV

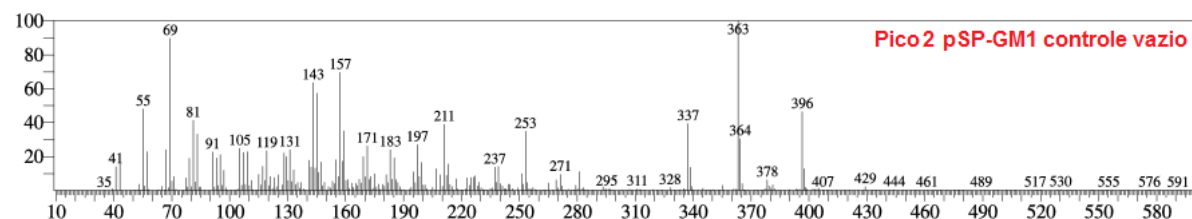
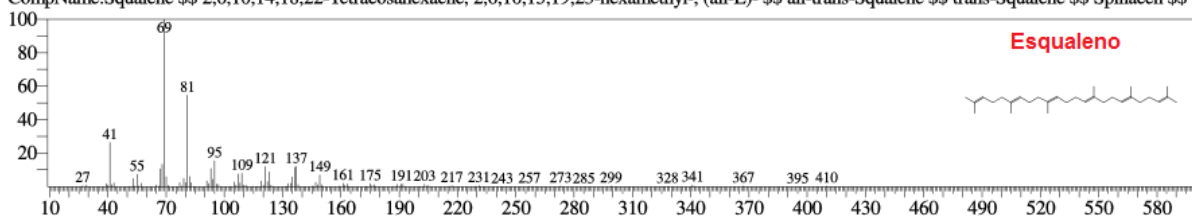
ScLAS 644 YVD--SEKSLV VOTAWALIALIFAEYPNK--EVD RGDILLKNRQEESEGEWKFEVVEGVF
 1W6K 641 YLQ--SAQSQIHNTCWAMMGLMAVRHPDI--EAQERGVRCLEKQLPNGDWPOENIAGVF
 BPX1 677 YTNLKDNRPHTVNTGWAMLALIDAGO AERDPTPLHRAARILINSQ MENGDFFPQEEIMGVF
 KdCAS 668 YTNLPGGRSHVVNTGWAMLALIGAGO AERDPTPLHRAAKFLIESOLENGDFFPQEEIMGVF
 CAS1 667 YSNLDGNRSHVVNTAWAMLALIGAGO AERDPTPLHRAARYLINAQ MENGDFFPQEEIMGVF
 PNX 667 YTNLEGNRSHVVNTGWAMLALIDAGO AERDPTPLHRAAKLLINSQ MENGDFFPQEEIMGVF
 PSX 667 YSNLEGNRSHVVNTGWAMLALIEAEQA KRDP TPLHRAAVCLINSOLENGDFFPQEEIMGVF
 GgCAS1 667 YSNVEGNRSHVVNTGWAMLALIDAEQA KRDP TPLHRAAVYLINSQ MENGDFFPQEEIMGVF
 OSC5 667 YSNLEGNRPHAVNTGWAMLALIEAEQA KRDP TPLHRAALYLINSQ MENGDFFPQEEIMGVF
 BPW 666 WTNLEGNRANLVOTAWALLSLIDARQAEIDPTPIH RGVRLINSQ MEDGDFFPQEEITGVF
 GgLUS1 666 YTNLEGNRANLVQSSWALLSLTHAGQAEIDPTPIH RGMKLLINSQ MEDGDFFPQEEITGVF
 OSC3 666 YTNLEGNRANLVQSSWALLSLMRAGQAEIDPTPIH RGI RLLINSQ MDDGDFFPQEEITGVF
 TRW 669 YTNLEGNRSLNVHTSWALLSLIKAGQAEIDPTPIH RGVRLINSQ MEEGDFFPQEEITGVF
 OEW 667 YTNLEGNRSLNVHTSWALLSLIKAGQAEIDPTPIH RGIKLLINSQ MEDGDFFPQEEITGAF
 PEN5 671 YIPSKGNVTNVNTGQAMVLIIGQMERDPTPVHRAAKV LINSQ LDI GDFFPQOERRGIY
 PEN6 676 YTTLEGNRINNVNTGQALMVLIMGGQMERDPTPVHRAAKV LINSQ LNDGDFFPQOEEIMGVF
 PEN4 667 YTPLKGNSTNVVOTAQALMVLIMGGQMERDPTPVHRAAKV LINSQ LNDGDFFPQOEEIMGT F
 PEN1 675 YTPLAGNKTINVSTGQALMVLIMGGQMERDPTPVHRAAKV V INLQ LNDGDFFPQOEEIMGVF
 RsM2 671 YTPLEGNRSLNVOTAWALMGLIHSQAERDPTPLHRS AKLLINSQ TSDGDFFPQODSTGLL
 LcIMS1 666 YTPLDGKRSNLVOTAWGMMGLICAGQAEVDPTPIH RAAKLLINSQ TEDGDFFPQOEEITGEF
 MiFRS 669 YTPLEGRRSNVVOTAWAMGLIYAGQAEIDPTPLH RGA KLLINSQ MEEGGYPQOEEITGVF
 LUP2 670 YTPLEGNRSLNVOTAWAMGLIHAGQAEIDPTPLH RAAKLLINSOLENGDFFPQOEEILGVF
 LUP1 667 YIPSEGNRSLNVHTSWAMMALIHTGQAEIDPTPLH RAAKLLINSOLENGDFFPQOEEIVGAF
 RcLUS 668 YVPLEGNRSLNVOTAWAMGLIYGGQAEIDPTPLH RAAKLLINSQ TELGDFFPQOEEITGAF
 KcMS 668 YVPLEGNRSLNVHTALAMGLILSGQAEIDPTPLH RAAKLLINSQ TELGDFFPQOEEISGCF
 BgLUS 668 YVPLEGNRSLNVOTALAMGLILGGQAEIDPTPLH RAAKLLINSQ TELGDFFPQOEEISGCF
 SS 668 YTPLDGNRSTVVHTAWAMGLIHSQAERDPTPLH RAAKLLINSQ MENGDFFPQODTITGAF
 KdFRS 669 YVPLEDNRSNLVHTSWALMGLIYAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 KdLUS 669 YVPLEDNRSNLVHTSWALMGLIYAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 KdGLS 669 YVPLEDNRSNLVHTSWALMGLICSGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 EtAS 669 YVPLEDNRSNLVHTSWALMGLISAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 RsM1 669 YVPLEANHSNLVHTAWAMMALVHAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 PNY1 670 YVPLEGNRSLNVHTSWALMGLIHSQAERDPTPLH RAAKLLINSQ MEDGDFFPQOEEISGVF
 PNY2 668 YTPLEGNRSLNVHTSWAMGLIHSQAERDPTPLH RAAKLLINSQ MESGDFFPQOEEITGVF
 KdTAS 688 YVPLEENKRSNLVHTAWALMGLIHSQAERDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 BPY 669 YVPLEGNRSLNVHTAWAMGLIHAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 PSM 669 YVPLERSQSNIVOTSWALMGLIHAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 AMY2 669 YVPLEGNRSLNVVOTAWALMGLIHAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 AMY1 669 YVPLEGSRSNVVHTAWALMGLIHAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 PSY 669 YVPLEGNRSLNVHTAWALMGLIHSQAERDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 OSC1 669 YVPLEGNRSLNVHTAWALMGLIHSQAERDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF
 GgbAS1 669 YVPLEGSRSNVVHTAWALMGLIHAGQAEIDPTPLH RAAKLLINSQ LEDGDFFPQOEEITGVF

ScLAS 700 NHSCALEYPSYRELFPIKALCMYSEAYETHTL-----
 1W6K 697 NKSCALSYTSYRNIFPIWALGRFSQLYPERALAGHP-----
 BPX1 737 NKNCMISYSAYRNIFPIWALGEYRCRVLKAL-----
 KdCAS 728 NKNCMISYAAYRNIFPIWALGEYRCRVLNASRGQMKT-----
 CAS1 727 NRNCMTIYAAYRNIFPIWALGEYRCOVLLQOGE-----
 PNX 727 DKNCMITYAAYRNIFPIWALGEYRCRVLQGPS-----
 PSX 727 NKNCMITYAAYRCIFPIWALGEYRRVLOAC-----
 GgCAS1 727 NKNCMITYAAYRNIFPIWALGEYRHRVLQSQ-----
 OSC5 727 NKNCMITYAAYRSIFPIWALGEYRCRVLQAR-----
 BPW 726 MRNCTLNYSSYRNIFPIWALGEYRRRVLFA-----
 GgLUS1 726 MRNCTLNYSSYRNIFPIWAMGEYRRQVLCCHSY-----
 OSC3 726 MRNCTLNYSSYRNIFPIWALGEYRRRVLCA-----
 TRW 729 MKNCNLYSSFRNIFPIWALGEYRRIVONI-----
 OEW 727 MRNCTLNYSSYRNIFPIWALGEYRRRILHAQT-----
 PEN5 731 -MNMILHYPTYRNIFSLWALALYTNALRLVLS-----
 PEN6 736 KMNVMHYATYRNIFPIWALTYTTRALRVPLC-----
 PEN4 727 MFTVMLHPTYRNIFSLWALHTYTHALRLLP-----
 PEN1 735 NMNVLLHYPTYRNIFSLWALTYTQALRRLQP-----
 RsM2 731 KGSCAMHYAAYRNIFPIWALAAAYRTHVLGLTSKAHSSAVME-----
 LcIMS1 726 FKNCMLHFAAAREVFPVWALGEYCNKVPVPSKKK-----
 MiFRS 729 KMNCMLHYPTYRNAFPIWALGEYRKRVPVPSKGNMAMKINSA-----
 LUP2 730 MNTCMLHYATYRNIFPIWALAEYRKAFAFATHQDL-----
 LUP1 727 MNTCMLHYATYRNIFPIWALAEYRKHVVFVN-----
 RcLUS 728 MRNCMLHYALERNFPIWALAEYRRHVLFPASAGFGFGFTNNL-----
 KcMS 728 MRNCMLHYSAYRDIFFPWALAEYCKLFPVPSKND-----
 BgLUS 728 MRNCMLHYSEYRDIFFPWALAEYCKLFPVPSKND-----
 SS 728 KKNCLLHYPMYRNIFPIWALAQYRKKVLRQPTGI-----
 KdFRS 729 KMNCMLHFAAYRNIFPIWALAVYRRCFNPNSEAISKPSK-----
 KdLUS 729 QRNCMLHYAAYRNIFPIWALAEYRRQVQLHSEATKMV-----
 KdGLS 729 KMNCMLHFAAYRSIFPWALAEYRRCNLSSEAISKPSK-----
 EtAS 729 MKNCLHFAAYRNIFPIWALAEYRNRVPLPSTTL-----
 RsM1 729 NRNCMLHYAAYRNIFPIWALAEYCRRVPLPS-----
 PNY1 730 MKNCLHYAAYRNIFPIWALAEYRRRVPLPSLGT-----
 PNY2 728 MKNCLHYAASRNIFPIWALAEYRKNVRLPSKSV-----
 KdTAS 748 MKNCMQHYAAYRNIFPIWGLAEYRQVPLPLR-----
 BPY 729 MKNCLHYAAYRNIFPIWALAEYRKHVPLPLGKNLNQVVNCIGQSLYKKYK
 PSM 729 MKNCLQYAMYRDIFFPWALAEYRRRILASPVAI-----
 AMY2 729 VKNCMLHYPMYRNIFPIWALAEYRRRVPLPSIAV-----
 AMY1 729 MKNCLHYPMYRDIFFPWALAEYRRRVPLPSTAV-----
 PSY 729 MKNCLHYPMYRDIFFPWALAEYRRRVPLP-----
 OSC1 729 MKNCLHYPMYRDIFFPWALAEYRRRVPLPSTAV-----
 GgbAS1 729 MKNCLHYPMYRDIFFPWALAEYRRRVPLPSTPVCLT-----

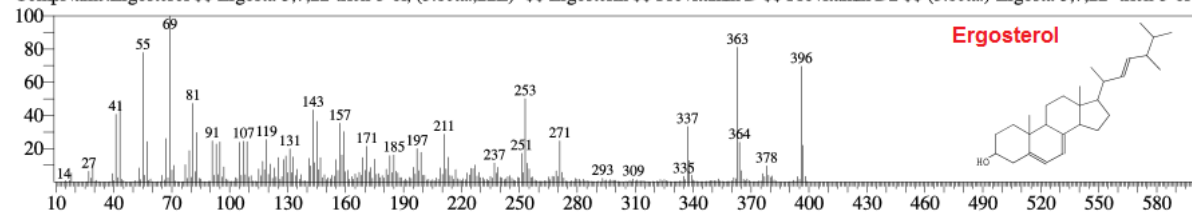
APÊNDICE C - Cromatogramas das frações apolares extraídas de *S. cerevisiae* e espectro de massas dos picos identificados no cromatograma e em sequência de cada pico sua identificação segundo a biblioteca NIST11.

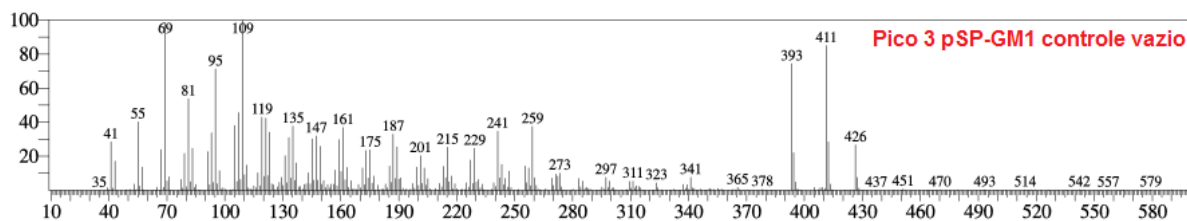


Entry:186078 Library:NIST11.lib
 SI:95 Formula:C30H50 CAS:111-02-4 MolWeight:410 RetIndex:2914
 CompName:Squalene \$\$ 2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)- \$\$ all-trans-Squalene \$\$ trans-Squalene \$\$ Spinacen \$\$ S

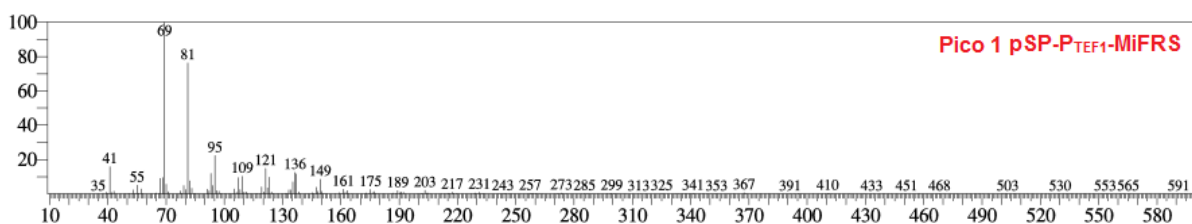
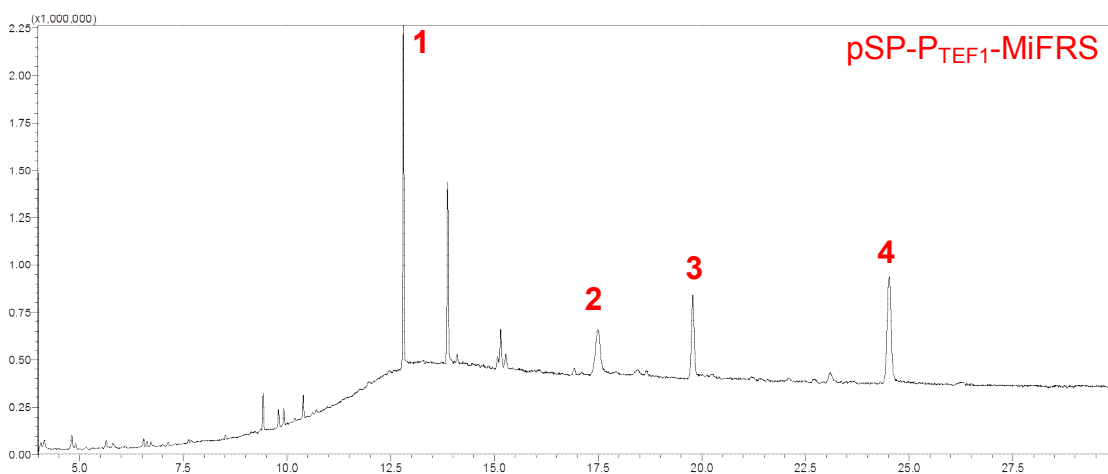
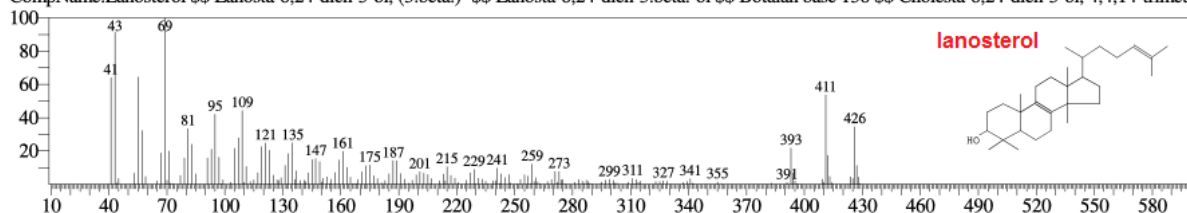


Entry:180874 Library:NIST11.lib
 SI:89 Formula:C28H44O CAS:57-87-4 MolWeight:396 RetIndex:2650
 CompName:Ergosterol \$\$ Ergosta-5,7,22-trien-3-ol, (3.beta.,22E)- \$\$ Ergosterin \$\$ Provitamin D \$\$ Provitamin D2 \$\$ (3.beta.)-Ergosta-5,7,22-trien-3-ol

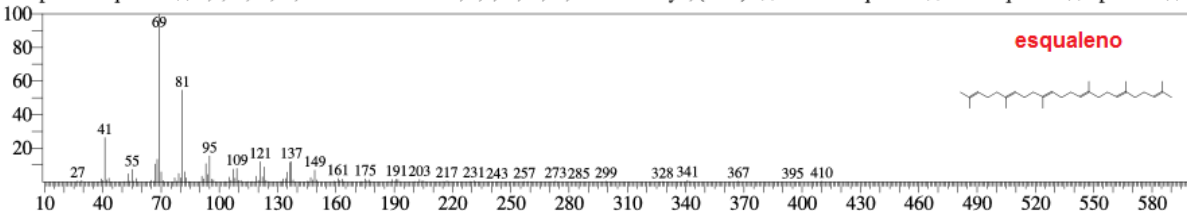


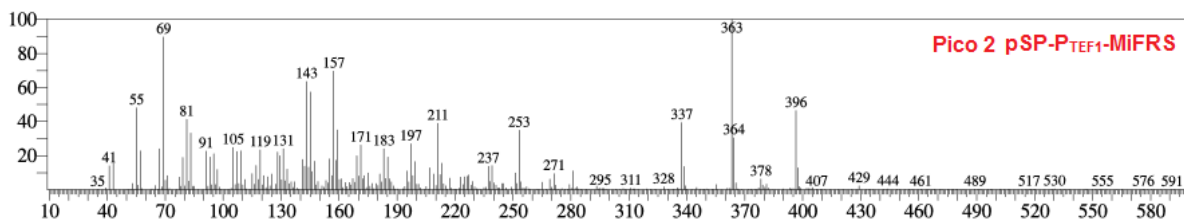


Entry:191107 Library:NIST11.lib
 SI:80 Formula:C₃₀H₅₀O CAS:79-63-0 MolWeight:426 RetIndex:2882
 CompName:Lanosterol \$\$ Lanosta-8,24-dien-3-ol, (3.beta.)- \$\$ Lanosta-8,24-dien-3.beta.-ol \$\$ Botolan base 138 \$\$ Cholesta-8,24-dien-3-ol, 4,4,14-trimet



Hit#:1 Entry:186078 Library:NIST11.lib
 SI:95 Formula:C₃₀H₅₀ CAS:111-02-4 MolWeight:410 RetIndex:2914
 CompName:Squalene \$\$ 2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)- \$\$ all-trans-Squalene \$\$ trans-Squalene \$\$ Spinacen \$\$ S

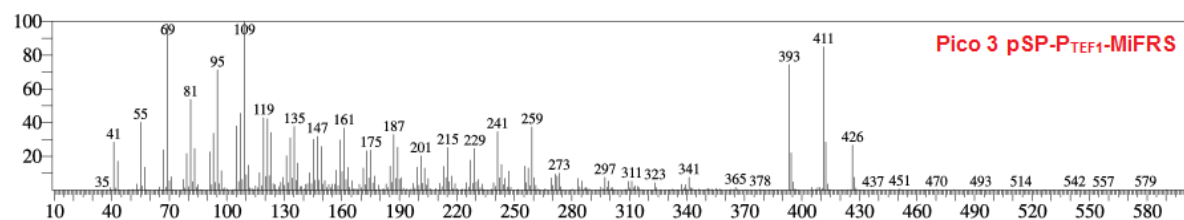
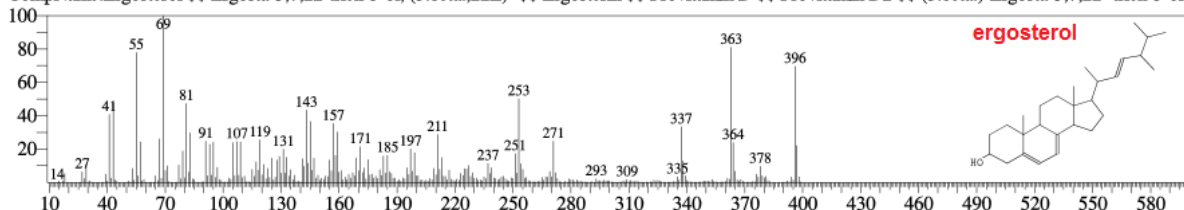




Hit#:1 Entry:180874 Library:NIST11.lib

SI:89 Formula:C₂₈H₄₄O CAS:57-87-4 MolWeight:396 RetIndex:2650

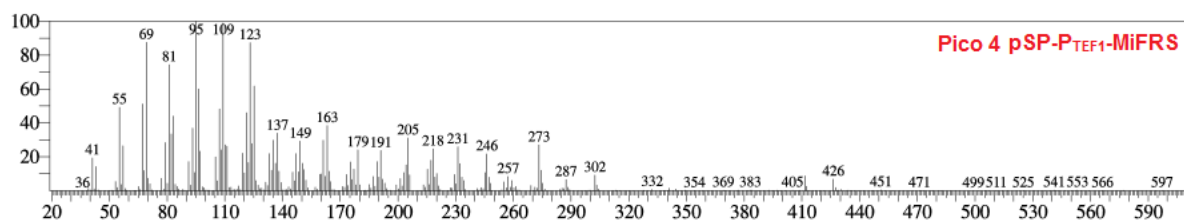
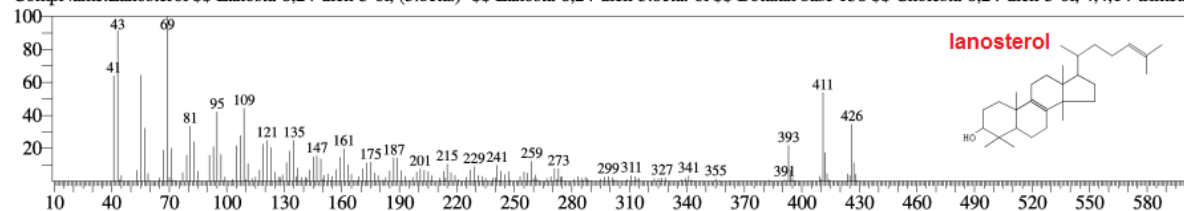
CompName:Ergosterol \$\$ Ergosta-5,7,22-trien-3-ol, (3.beta.,22E)- \$\$ Ergosterin \$\$ Provitamin D \$\$ Provitamin D2 \$\$ (3.beta.)-Ergosta-5,7,22-trien-3-ol



Hit#:1 Entry:191107 Library:NIST11.lib

SI:80 Formula:C₃₀H₅₀O CAS:79-63-0 MolWeight:426 RetIndex:2882

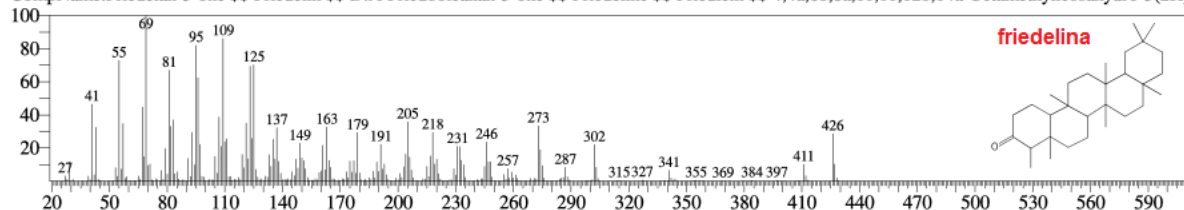
CompName:Lanosterol \$\$ Lanosta-8,24-dien-3-ol, (3.beta.)- \$\$ Lanosta-8,24-dien-3.beta.-ol \$\$ Botalan base 138 \$\$ Cholesta-8,24-dien-3-ol, 4,4,14-trimethyl-

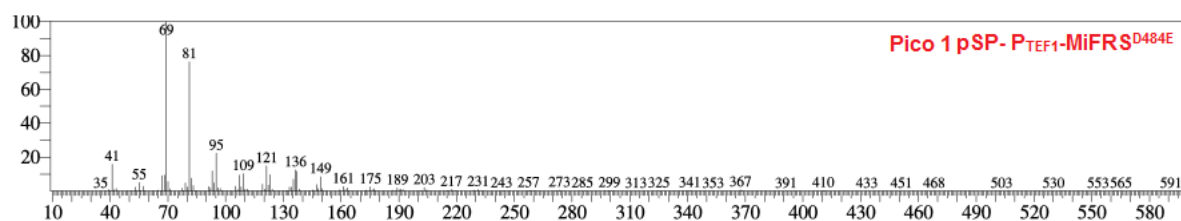
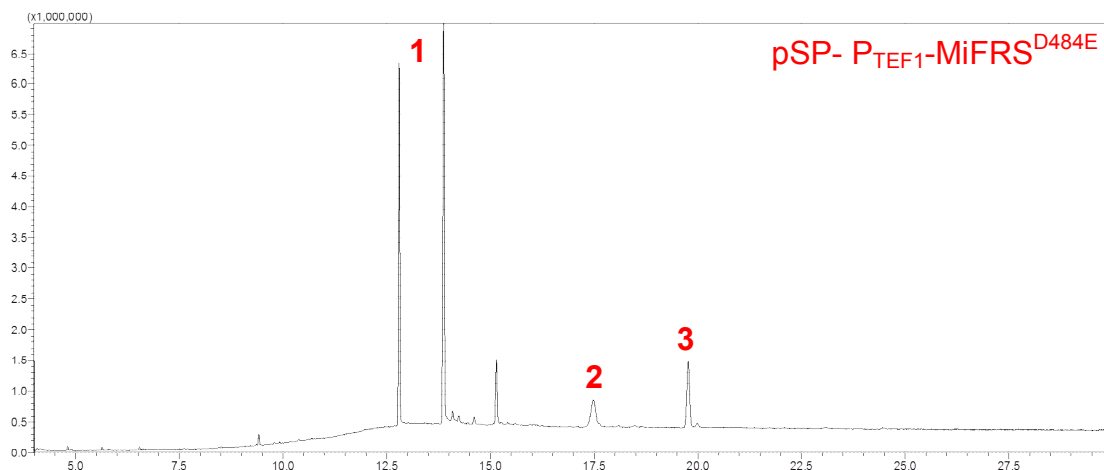


Hit#:1 Entry:191113 Library:NIST11.lib

SI:90 Formula:C₃₀H₅₀O CAS:559-74-0 MolWeight:426 RetIndex:2858

CompName:Friedelin-3-one \$\$ Friedelin \$\$ D:A-Friedooleanan-3-one \$\$ Friedeline \$\$ Friedlein \$\$ 4,4a,6b,8a,11,11,12b,14a-Octamethylcosahydro-3(2H)

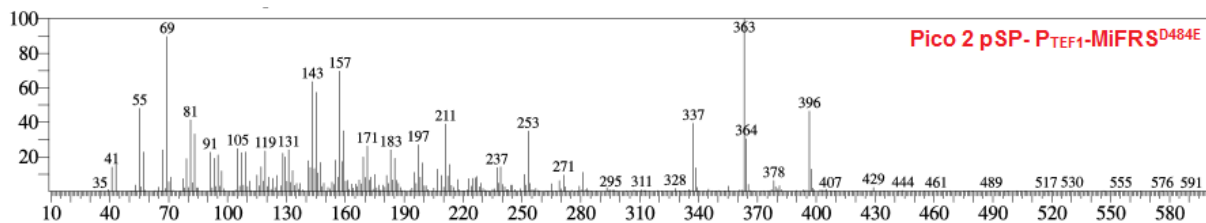
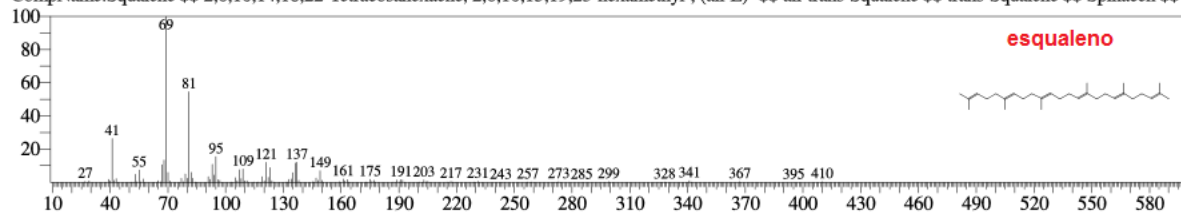




Hit#1 Entry:186078 Library:NIST11.lib

SI:95 Formula:C₃₀H₅₀ CAS:111-02-4 MolWeight:410 RetIndex:2914

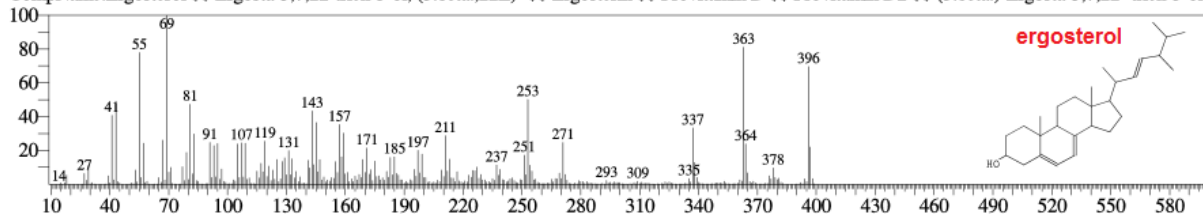
CompName:Squalene \$\$ 2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)- \$\$ all-trans-Squalene \$\$ trans-Squalene \$\$ Spinacen \$\$ S

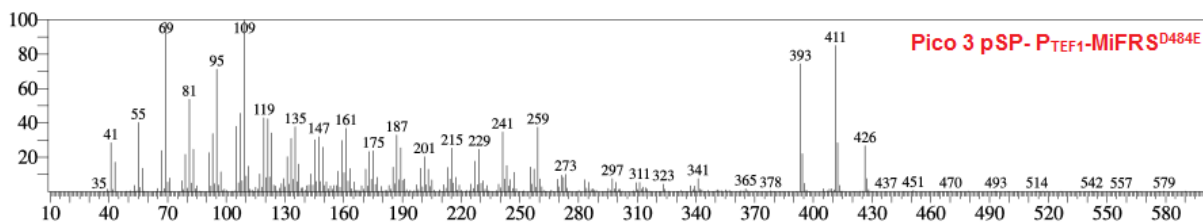


Hit#1 Entry:180874 Library:NIST11.lib

SI:89 Formula:C₂₈H₄₄O CAS:57-87-4 MolWeight:396 RetIndex:2650

CompName:Ergosterol \$\$ Ergosta-5,7,22-trien-3-ol, (3.beta.,22E)- \$\$ Ergosterin \$\$ Provitamin D \$\$ Provitamin D2 \$\$ (3.beta.)-Ergosta-5,7,22-trien-3-ol

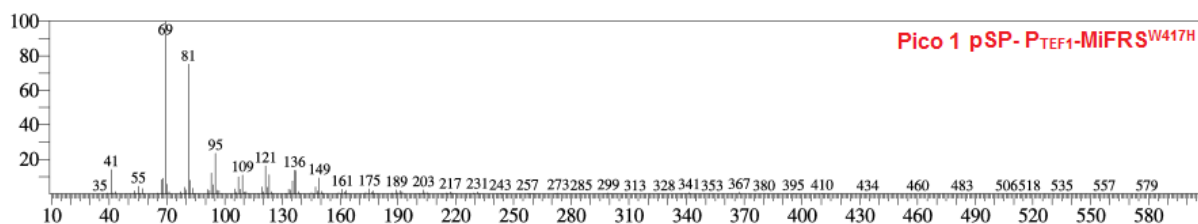
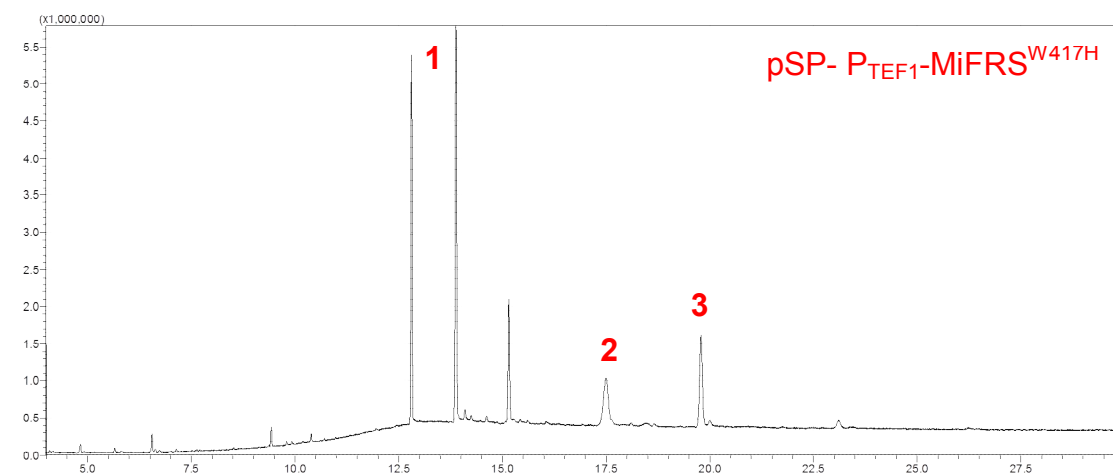
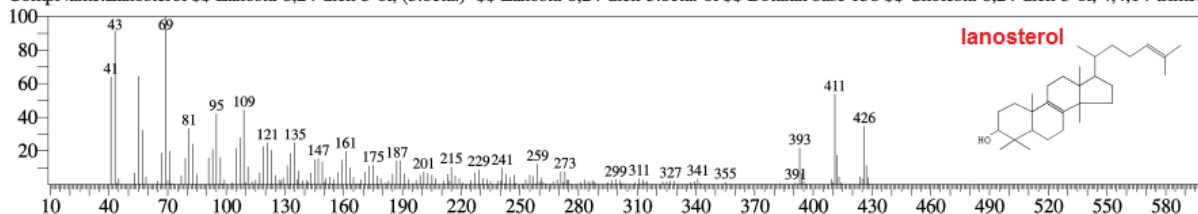




Hit#:1 Entry:191107 Library:NIST11.lib

SI:80 Formula:C₃₀H₅₀O CAS:79-63-0 MolWeight:426 RetIndex:2882

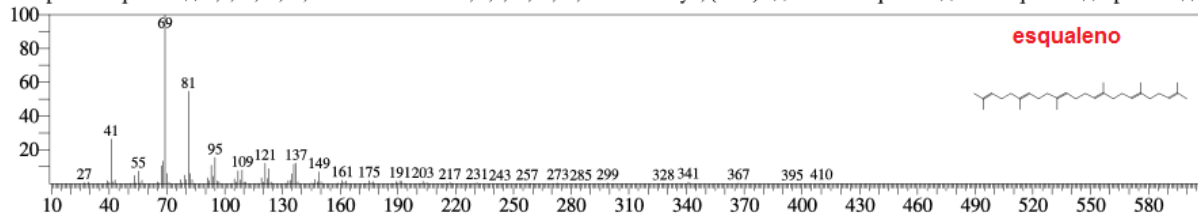
CompName:Lanosterol \$\$ Lanosta-8,24-dien-3-ol, (3.beta.)- \$\$ Lanosta-8,24-dien-3.beta.-ol \$\$ Botalan base 138 \$\$ Cholesta-8,24-dien-3-ol, 4,4,14-trimett

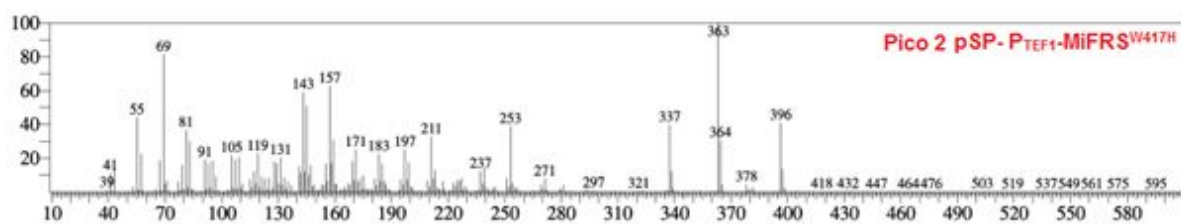


Hit#:1 Entry:186078 Library:NIST11.lib

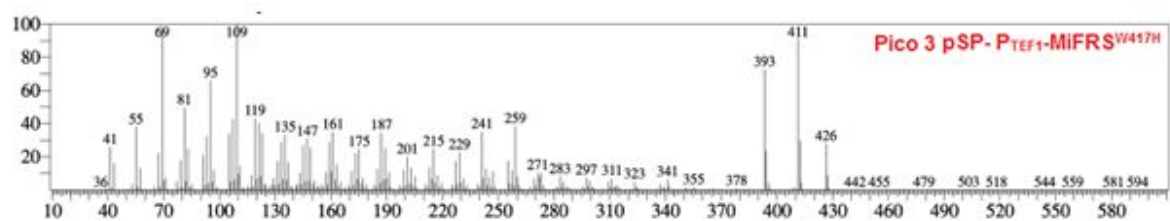
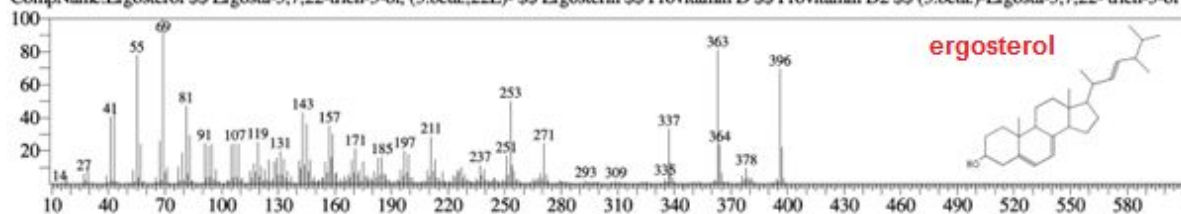
SI:94 Formula:C₃₀H₅₀ CAS:111-02-4 MolWeight:410 RetIndex:2914

CompName:Squalene \$\$ 2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)- \$\$ all-trans-Squalene \$\$ trans-Squalene \$\$ Spinacen \$\$ S

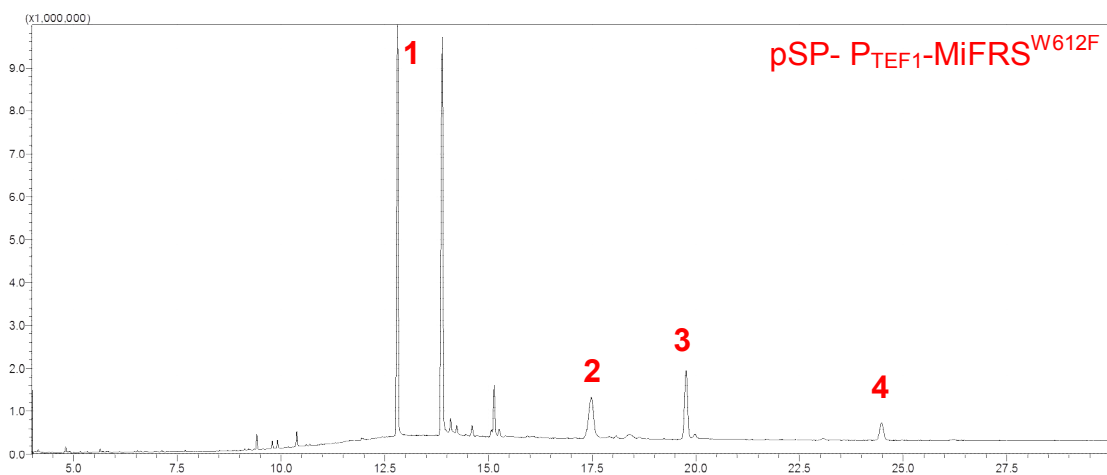
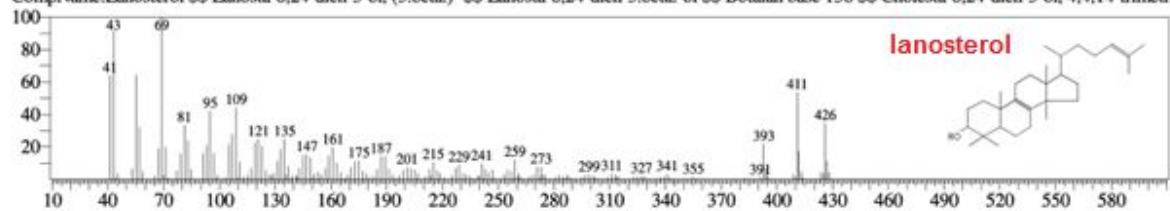


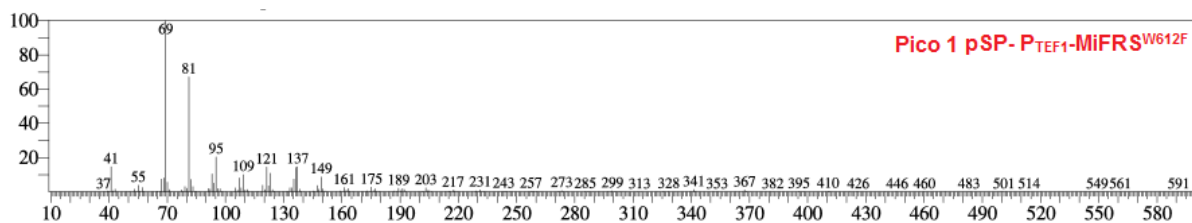


Hit#:1 Entry:180874 Library:NIST11.lib
 SI:89 Formula:C₂₈H₄₄O CAS:57-87-4 MolWeight:396 RetIndex:2650
 CompName:Ergosterol \$\$ Ergosta-5,7,22-trien-3-ol, (3.beta.,22E)- \$\$ Ergosterin \$\$ Provitamin D \$\$ Provitamin D2 \$\$ (3.beta.)-Ergosta-5,7,22- trien-3-ol !



Hit#:1 Entry:191107 Library:NIST11.lib
 SI:79 Formula:C₃₀H₅₀O CAS:79-63-0 MolWeight:426 RetIndex:2882
 CompName:Lanosterol \$\$ Lanosta-8,24-dien-3-ol, (3.beta.)- \$\$ Lanosta-8,24-dien-3.beta.-ol \$\$ Botolan base 138 \$\$ Cholesta-8,24-dien-3-ol, 4,4,14-trimett

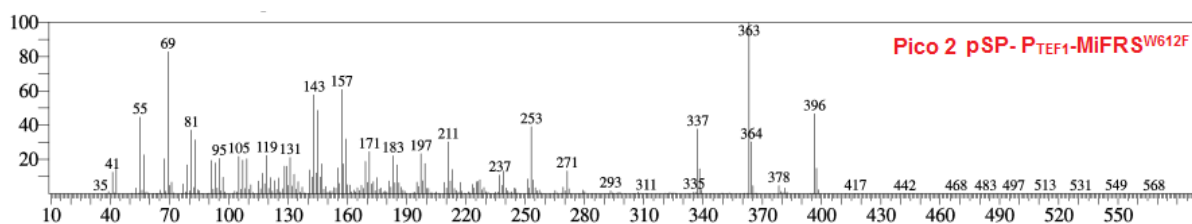
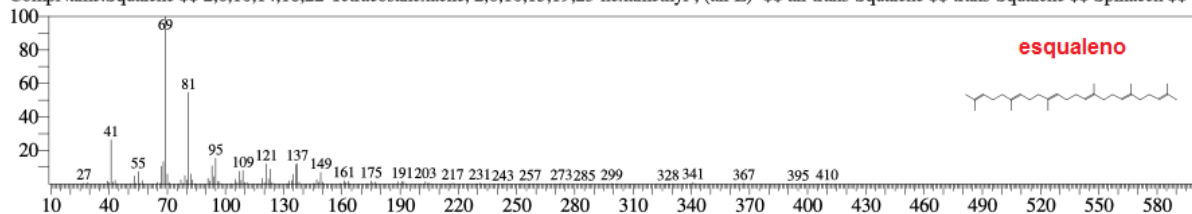




Hit#:1 Entry:186078 Library:NIST11.lib

SI:94 Formula:C₃₀H₅₀ CAS:111-02-4 MolWeight:410 RetIndex:2914

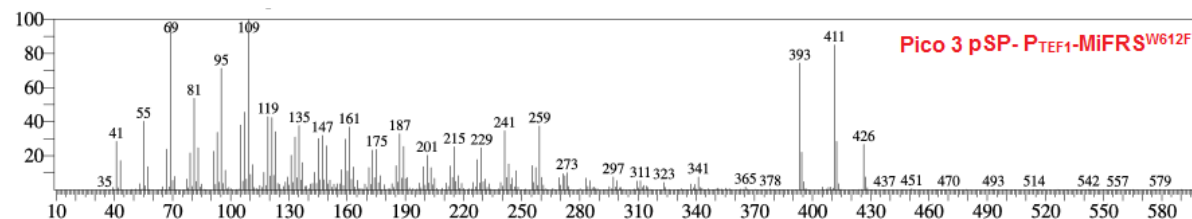
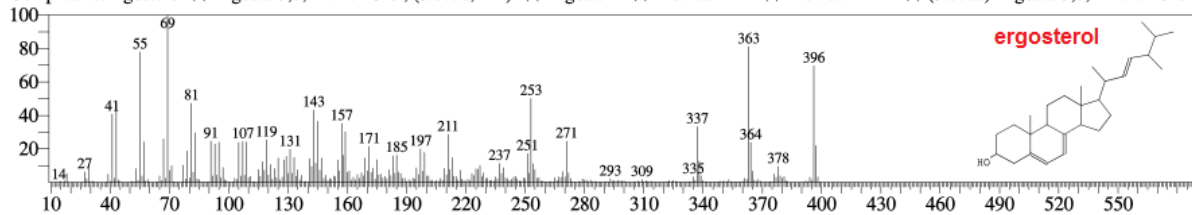
CompName:Squalene \$\$ 2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)- \$\$ all-trans-Squalene \$\$ trans-Squalene \$\$ Spinacen \$\$ S



Hit#:1 Entry:180874 Library:NIST11.lib

SI:91 Formula:C₂₈H₄₄O CAS:57-87-4 MolWeight:396 RetIndex:2650

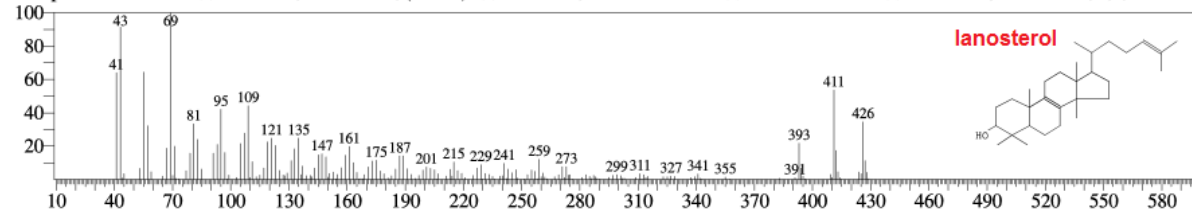
CompName:Ergosterol \$\$ Ergosta-5,7,22-trien-3-ol, (3.beta.,22E)- \$\$ Ergosterin \$\$ Provitamin D \$\$ Provitamin D2 \$\$ (3.beta.)-Ergosta-5,7,22-trien-3-ol

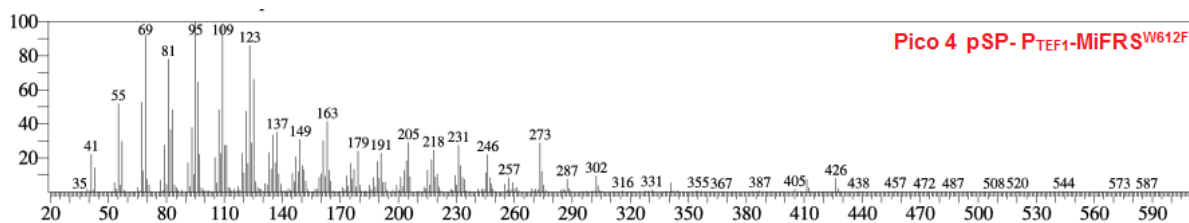


Hit#:1 Entry:191107 Library:NIST11.lib

SI:80 Formula:C₃₀H₅₀O CAS:79-63-0 MolWeight:426 RetIndex:2882

CompName:Lanosterol \$\$ Lanosta-8,24-dien-3-ol, (3.beta.)- \$\$ Lanosta-8,24-dien-3.beta.-ol \$\$ Botalan base 138 \$\$ Cholesta-8,24-dien-3-ol, 4,4,14-trimethyl-

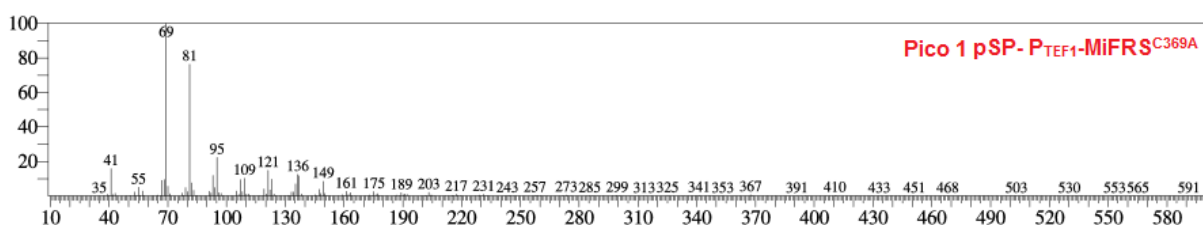
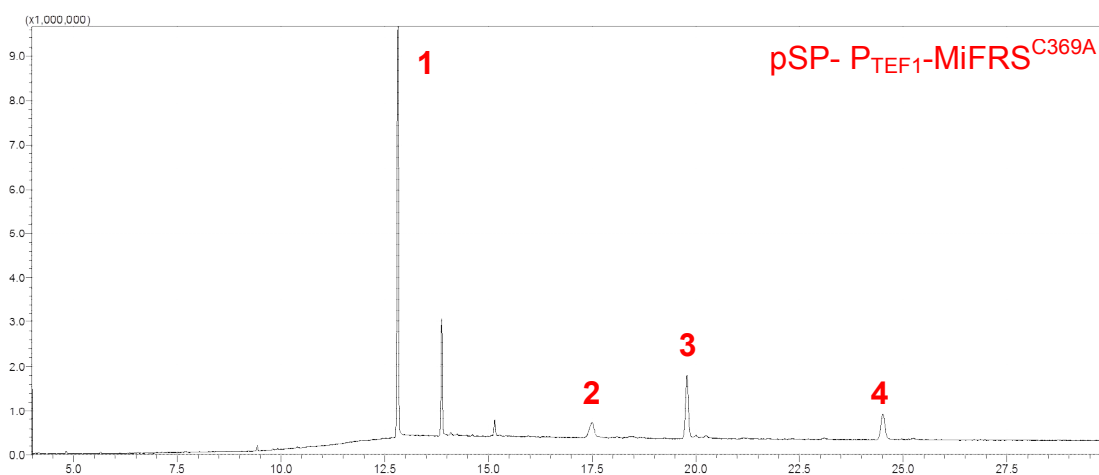
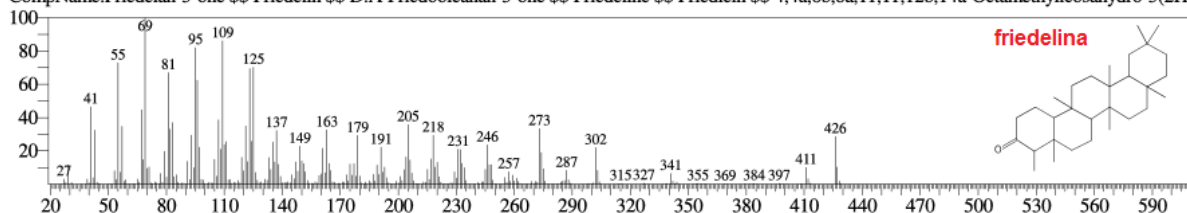




Hit#:1 Entry:191113 Library:NIST11.lib

SI:90 Formula:C₃₀H₅₀O CAS:559-74-0 MolWeight:426 RetIndex:2858

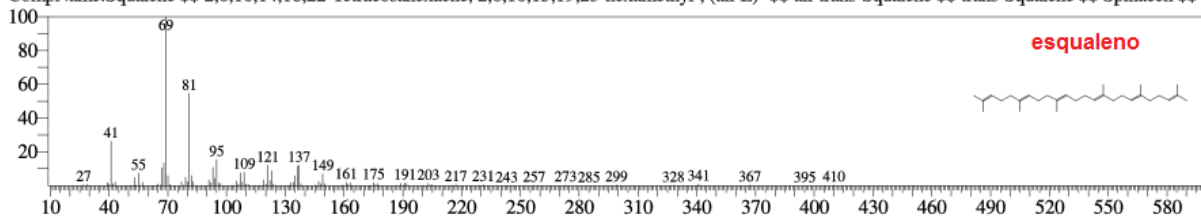
CompName:Friedelan-3-one \$\$ Friedelin \$\$ D:A-Friedoolean-3-one \$\$ Friedeline \$\$ Friedlein \$\$ 4,4a,6b,8a,11,11,12b,14a-Octamethylsahydro-3(2H)

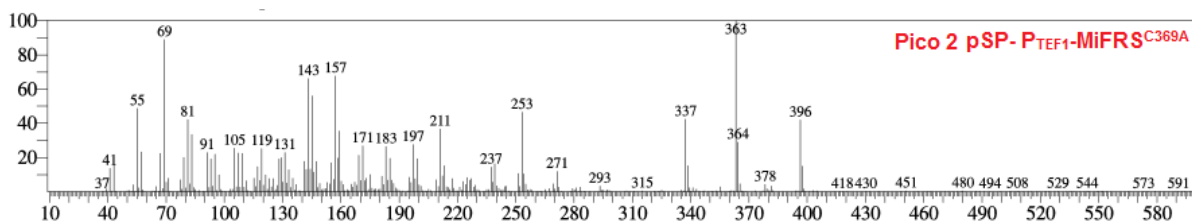


Hit#:1 Entry:186078 Library:NIST11.lib

SI:95 Formula:C₃₀H₅₀ CAS:111-02-4 MolWeight:410 RetIndex:2914

CompName:Squalene \$\$ 2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)- \$\$ all-trans-Squalene \$\$ trans-Squalene \$\$ Spinacen \$\$ S

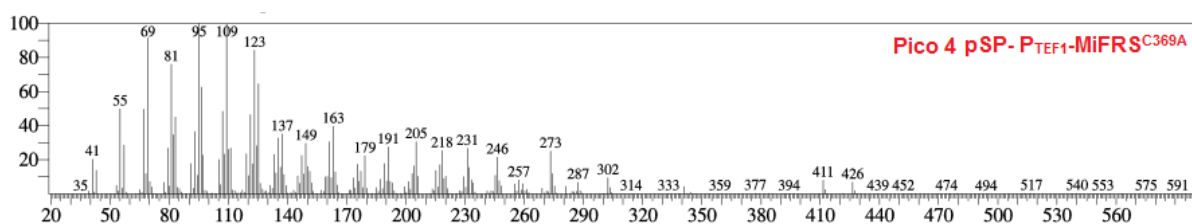
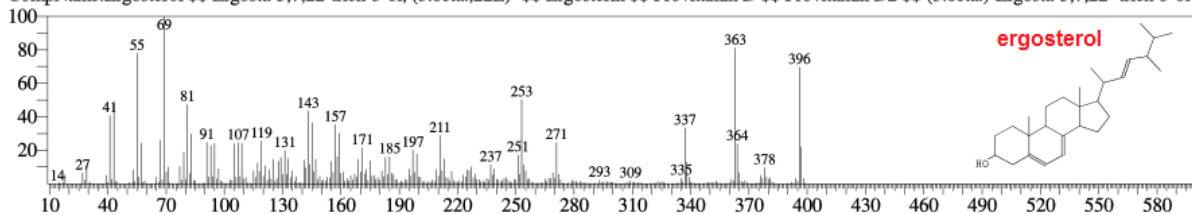




Hit#:1 Entry:180874 Library:NIST11.lib

SI:90 Formula:C₂₈H₄₄O CAS:57-87-4 MolWeight:396 RetIndex:2650

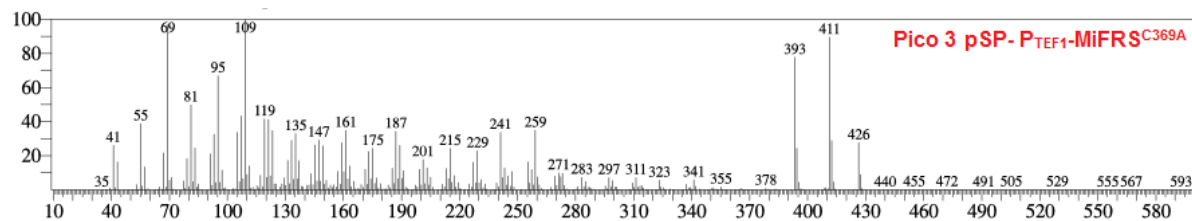
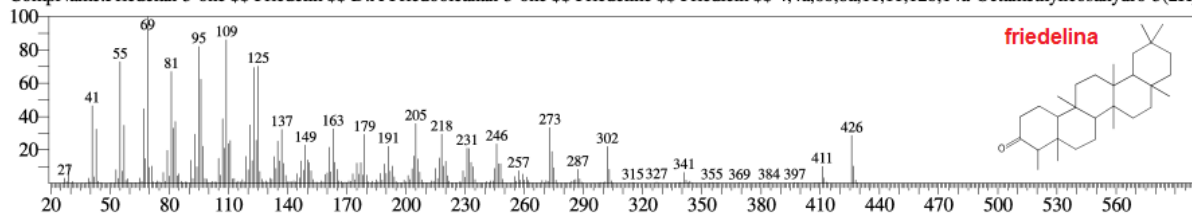
CompName:Ergosterol \$\$ Ergosta-5,7,22-trien-3-ol, (3.β.,22E)- \$\$ Ergosterin \$\$ Provitamin D \$\$ Provitamin D2 \$\$ (3.β.,)-Ergosta-5,7,22-trien-3-ol



Hit#:1 Entry:191113 Library:NIST11.lib

SI:90 Formula:C₃₀H₅₀O CAS:559-74-0 MolWeight:426 RetIndex:2858

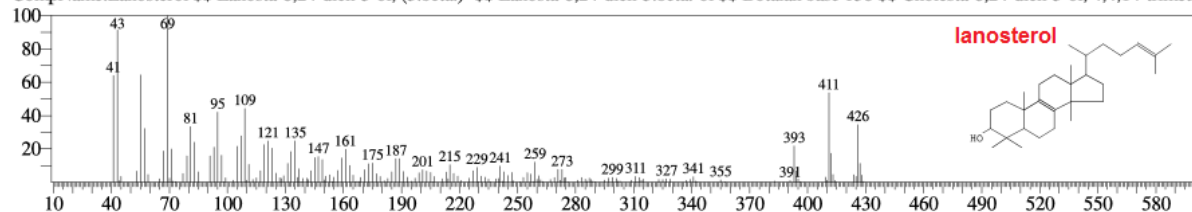
CompName:Friedelan-3-one \$\$ Friedelin \$\$ D:A-Friedooleanan-3-one \$\$ Friedeline \$\$ Friedlein \$\$ 4,4a,6b,8a,11,11,12b,14a-Octamethylcosahydro-3(2H)

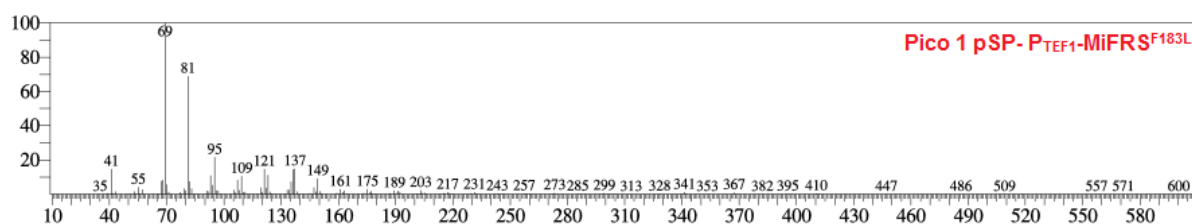
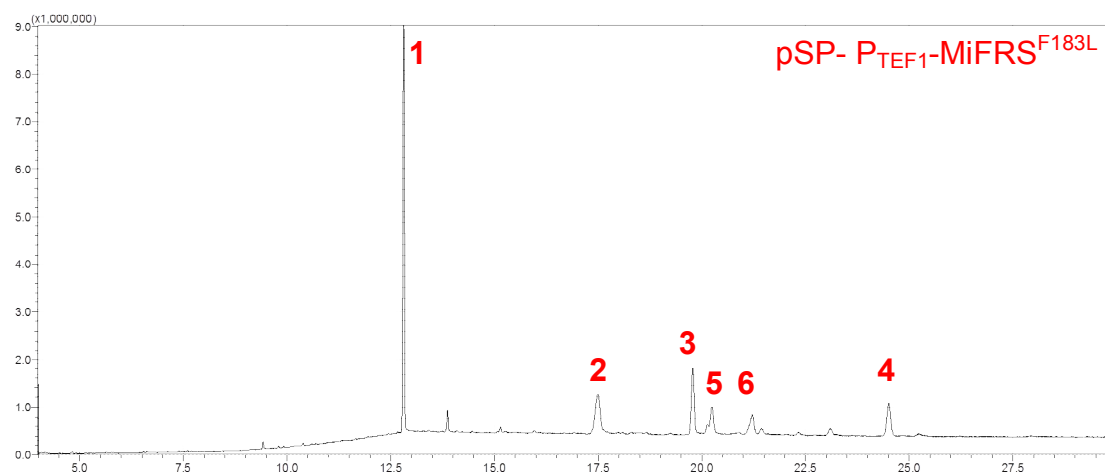


Hit#:1 Entry:191107 Library:NIST11.lib

SI:80 Formula:C₃₀H₅₀O CAS:79-63-0 MolWeight:426 RetIndex:2882

CompName:Lanosterol \$\$ Lanosta-8,24-dien-3-ol, (3.β.,)- \$\$ Lanosta-8,24-dien-3.β.,-ol \$\$ Botalan base 138 \$\$ Cholesta-8,24-dien-3-ol, 4,4,14-trimethyl

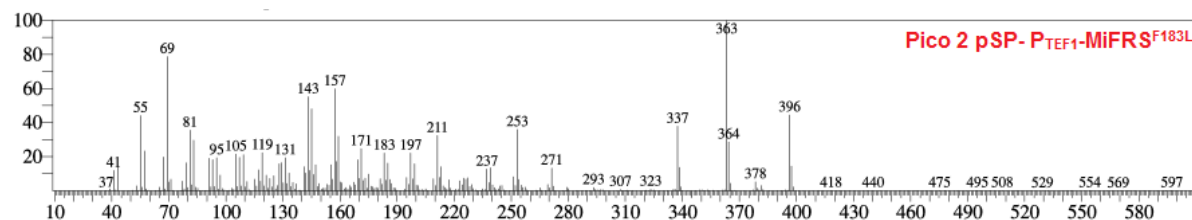
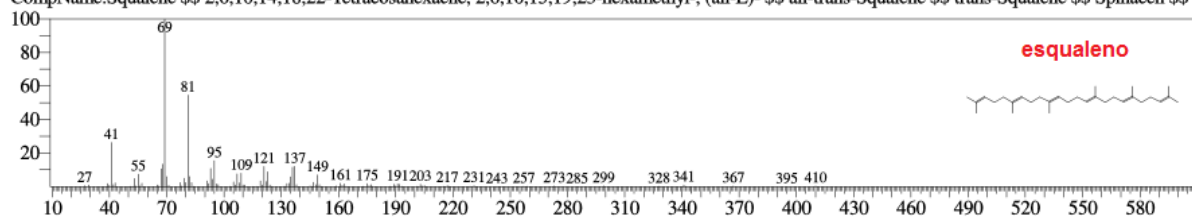




Hit#:1 Entry:186078 Library:NIST11.lib

SI:94 Formula:C₃₀H₅₀ CAS:111-02-4 MolWeight:410 RetIndex:2914

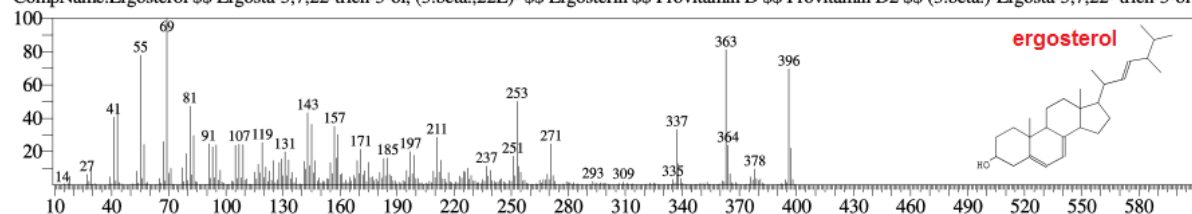
CompName:Squalene \$\$ 2,6,10,14,18,22-Tetracosahexaene, 2,6,10,15,19,23-hexamethyl-, (all-E)- \$\$ all-trans-Squalene \$\$ trans-Squalene \$\$ Spinacen \$\$ S

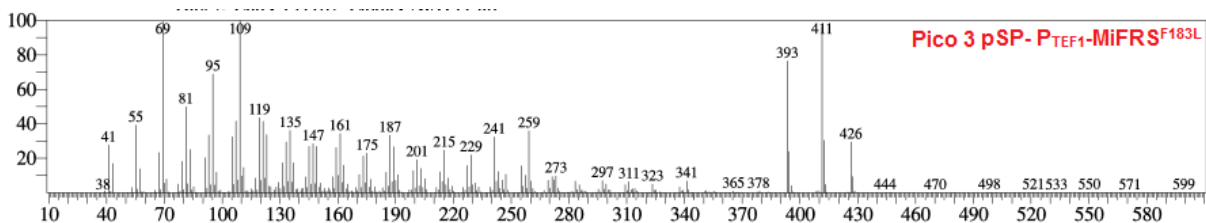


Hit#:1 Entry:180874 Library:NIST11.lib

SI:92 Formula:C₂₈H₄₄O CAS:57-87-4 MolWeight:396 RetIndex:2650

CompName:Ergosterol \$\$ Ergosta-5,7,22-trien-3-ol, (3.beta.,22E)- \$\$ Ergosterin \$\$ Provitamin D \$\$ Provitamin D2 \$\$ (3.beta.)-Ergosta-5,7,22-trien-3-ol

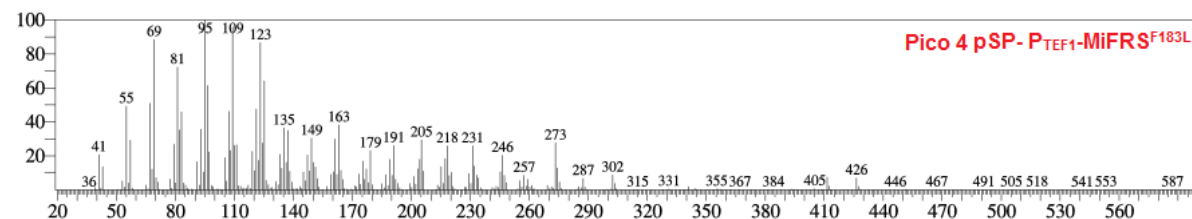
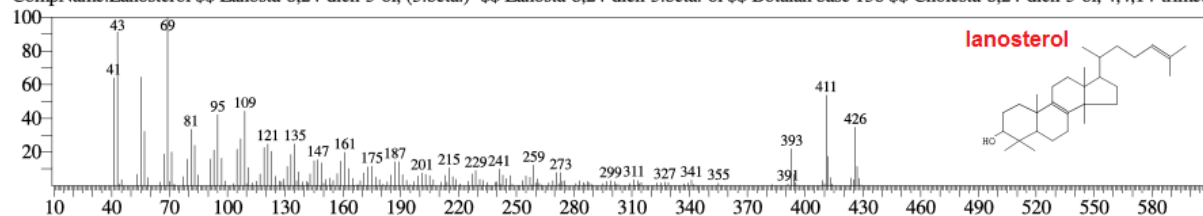




Hit#:2 Entry:191107 Library:NIST11.lib

SI:81 Formula:C₃₀H₅₀O CAS:79-63-0 MolWeight:426 RetIndex:2882

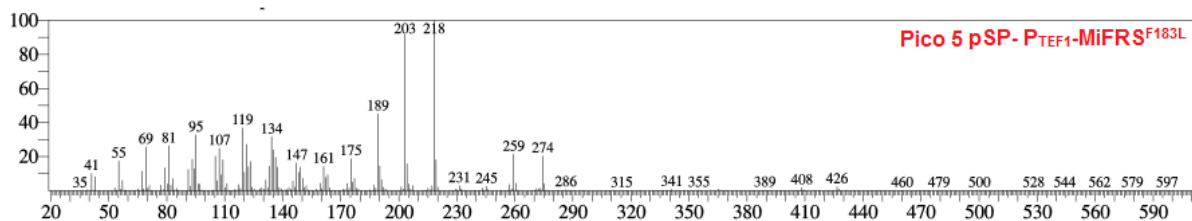
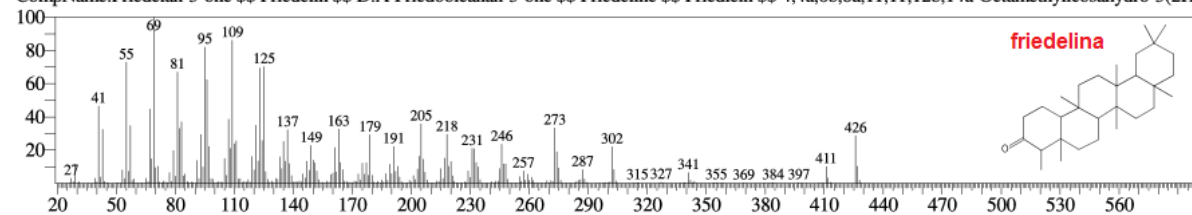
CompName:Lanosterol \$\$ Lanosta-8,24-dien-3-ol, (3.beta.)- \$\$ Lanosta-8,24-dien-3.beta.-ol \$\$ Botalan base 138 \$\$ Cholesta-8,24-dien-3-ol, 4,4,14-trimet



Hit#:1 Entry:191113 Library:NIST11.lib

SI:90 Formula:C₃₀H₅₀O CAS:559-74-0 MolWeight:426 RetIndex:2858

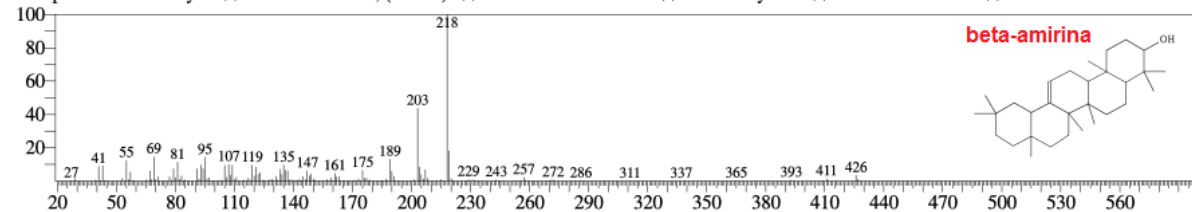
CompName:Friedelan-3-one \$\$ Friedelin \$\$ D:A-Friedoolean-3-one \$\$ Friedeline \$\$ Friedlein \$\$ 4,4a,6b,8a,11,11,12b,14a-Octamethylcosahydro-3(2H)

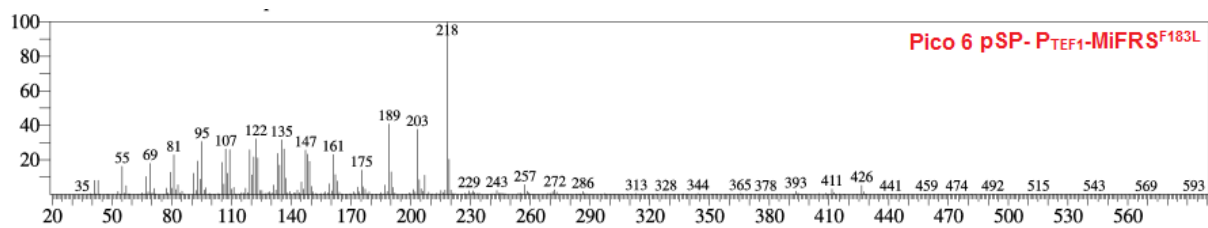


Hit#:2 Entry:191127 Library:NIST11.lib

SI:81 Formula:C₃₀H₅₀O CAS:559-70-6 MolWeight:426 RetIndex:2886

CompName:.beta.-Amyrin \$\$ Olean-12-en-3-ol, (3.beta.)- \$\$ Olean-12-en-3.beta.-ol \$\$.beta.-Amyrenol \$\$ Olean-12-en-3-ol # \$\$

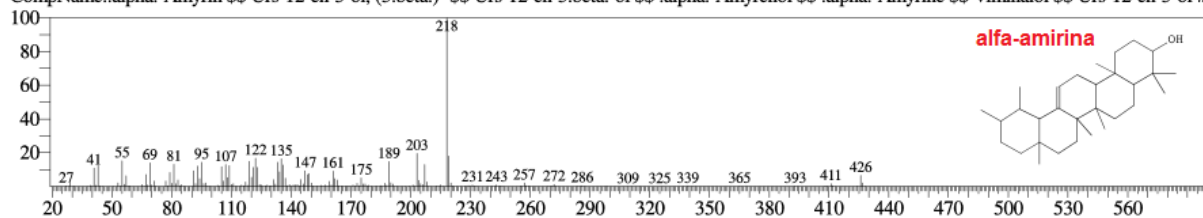




Hit#:1 Entry:191128 Library:NIST11.lib

SI:85 Formula:C₃₀H₅₀O CAS:638-95-9 MolWeight:426 RetIndex:2873

CompName:.alpha.-Amyrin \$\$ Urs-12-en-3-ol, (3.beta.)- \$\$ Urs-12-en-3.beta.-ol \$\$.alpha.-Amyrenol \$\$.alpha.-Amyrine \$\$ Viminalol \$\$ Urs-12-en-3-ol #



ANEXO A - Matriz *BLOSUM62* de substituição de aminoácidos

A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4
*	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V