Chapter-4

LC-MS-based Dereplication and Cytotoxic Evaluation of *A. cunninghamii* Gum-resin

4 LC-MS based dereplication and cytotoxic evaluation of *A. cunninghamii* gum-resin

4.1 Introduction

As discussed, the early identification is known as dereplication this strategy used to explore the natural products for the drug discovery. Liquid chromatography-mass spectrometry (LC-MS) is the most widely used system in chemistry in which HPLC is tethered with MS detector, and the separated components from HPLC column enter MS detector and are fragmented. Based on the fragments, the components can be identified. The mass spectra obtained provide the indispensable information about the separated molecule. Even if the molecule is known, its structure can be reconstructed by analyzing the mass spectra. The important advantage of MS analysis is that minute quantity of sample is required for complete analysis. Generally, soft ionization techniques are used in MS, which produces molecular ion peaks. However, the tandem mass spectrometry (MS-MS) produces collision-induced fragments of the produced molecular ion. Hyphenated techniques such as LC-UV and LC-MS have been extensively used in combination with biological screening for the discovery of natural products.

A. cunninghamii (Araucariaceae) commonly known as 'Hoop pine', is an evergreen coniferous tree, native to the Himalayan region of India and Pakistan. Coniferous trees are known to contain a wide variety of phytochemicals including alkaloids, flavonoids, steroids, terpenoids, saponins, tannins, terpenoids, and polyphenolic compounds. Many phytoconstituents have been isolated from different organs of LC-MS Mudie that

mainly of biflavones, lignans, diterpenes (labdane and abietane) [106-108]. In a series of phytochemical investigations of oleo-resins of Araucaria genus, LC-MS oleo-resin yielded many labdane diterpenes [108].

4.1.1 Labdane and abietane diterpenoids.

Labdane and abietane are major components of the secondary metabolites of various pines. These diterpenes found in nature, are made up of four isoprene units. Labdanes are bicyclic in nature characterized with a fused decalin system (C1–10) and a branched six-carbon side chain (C11–16, with C16 connected to C13) at carbon 9 make up the basic skeleton structure. The decalin system has the remaining four carbons (C17–20) attached at C8, C4 (where C18 and C19 are attached), and C10. The stereochemistry present at the ring intersection, labdane skeletons are defined according to the configuration of the ring junction and the substituents at C9 and C10. As a result, the terms, '*normal*-', '*ent*-', '*syn*-', and '*syn-ent*-', are used [135]. Andrographolides are classical examples of labdane [136].

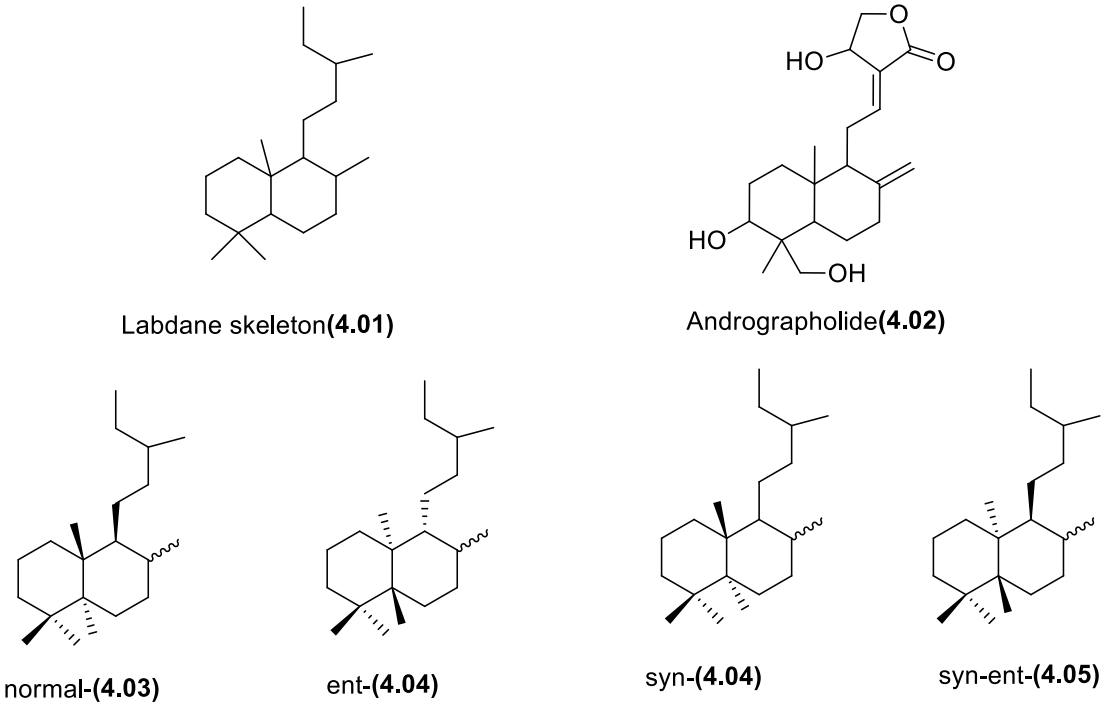
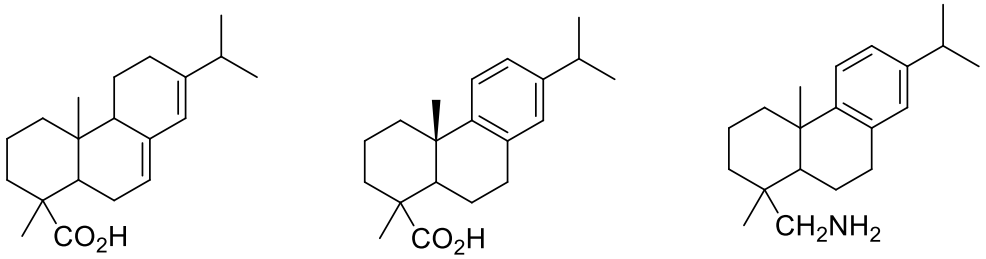


Figure 4. 1 Chemistry of labdane diterpenes

Abietanes that are diterpenoids have the distinctive carbon skeleton, as demonstrated in the the resin acid. Abietic acid and dehydroabietic acid are the main components of pine resin obtained from Pinus sp. Naturally occurring tricyclic diterpenoids that have many potential medicinal uses. Abietic acid is a colorless, crystalline, tricyclic acid with the two double bonds. It's a major component of rosin, the solid part of coniferous tree oleoresin, and is produced by pines and other conifers. Abietic acid is a conjugated diene and carboxylic acid, and a member of the abietane diterpene group of organic compound [138]. Numerous biological activities, such as antimicrobial, anti-plasmodial, antifungal, antitumor, cytotoxic, antiviral, antiulcer, cardiovascular, antioxidant, and anti-inflammatory activity, have been demonstrated by abietanes [138]. The structure-activity relationship mentioned above was provisional and needed further verification [139]. Abietane diterpenoids exhibit a variety of intriguing biological functions, such as antiviral capabilities. A number of research teams have

investigated the possibility of using derivatives of commercial starting materials, such as (-)-Abietic acid **(4.06)**, which can be transformed into (+)-dehydroabietic acid **(4.07)** and (+)-dehydroabietylamine (**4.08**, DHAA), also known as leelamine, as chemotherapeutic agents [140].



Abietic acid **(4.06)** Dehydroabietic acid **(4.07)** Dehydroabietylamine **(4.08)**

Figure 4. 2 Structure of abietane derivative

Liquid chromatography–mass spectrometry (LC–MS)-based dereplication has become a widely used analytical technique in the pharmaceutical and biotech industry. Amalgamation of mass spectrometry in the laboratory is highly relevant to student training, especially in the areas of metabolomics and metabolite profiling. As with other "-omics" approaches such as genomics and proteomics, which deal with the collective characterization of specific biological molecules, a metabolomics study focuses on the comprehensive analysis of metabolites in a biological specimen. Metabolite profiling is a focused approach to analyse and annotate molecules that are produced via a specific metabolic pathway or, in the case of medicinal plants, specific classes of molecules such as polyketides, terpenes, or shikimates. Additionally, liquid chromatography coupled with tandem mass spectrometry (MS/MS) allows for increased confidence in metabolite identification in untargeted analysis by comparing MS/MS fragmentation patterns of unknown analyses to those of authentic samples found in the literature or to

standards analysed independently by the researcher. LC-MS based dereplication strategies combined with database search can be very helpful in the identification of new metabolites to be targeted in the plant extracts.

4.2 Results and discussion

4.2.1 Phytochemical investigation of LC-MS gum-resins

After dissolving the gum resin in ethanol, the soluble and insoluble portions were separated through filtering. After that, the soluble material was concentrated, and its cytotoxicity was assessed. The ethyl acetate fraction was concentrated for further phytochemical analysis after it was further fractionated using ethyl acetate-water partitioning. Simultaneously, the ethyl acetate fraction was subjected to LC-MS analysis and chromatographic purification through silica gel column chromatography, sephadex LH-20, and preparative TLC by using hexane-ethyl-acetate respectively.

Total fourteen compound were isolated and characterised using 1D and 2D NMR spectroscopy and mass spectrometric analysis. Apart from one new compound **(4.17**) other known compounds viz. Pinifolic acid (**4.09**), dehydropinifolic acid (**4.10**, agathic acid), agathic acid methylester **(4.11**), cupressic acid (**4.12**), isocupressic acid (**4.13**), manool (**4.14**), imbricataloic acid (**4.15**), trans-communic acid (**4.16**), agatheol methylether **(4.17**), abietic acid (**4.18**), pimaric acid (**4.19**), 7-ketodehydroabietic acid **(4.20**), sandaracopimarinol-acetate (**4.21**), and sandaracopimarinol (**4.22**) (**Figure 4.3**) were identified by comparing the observed NMR spectra with the reported spectroscopic data**.**

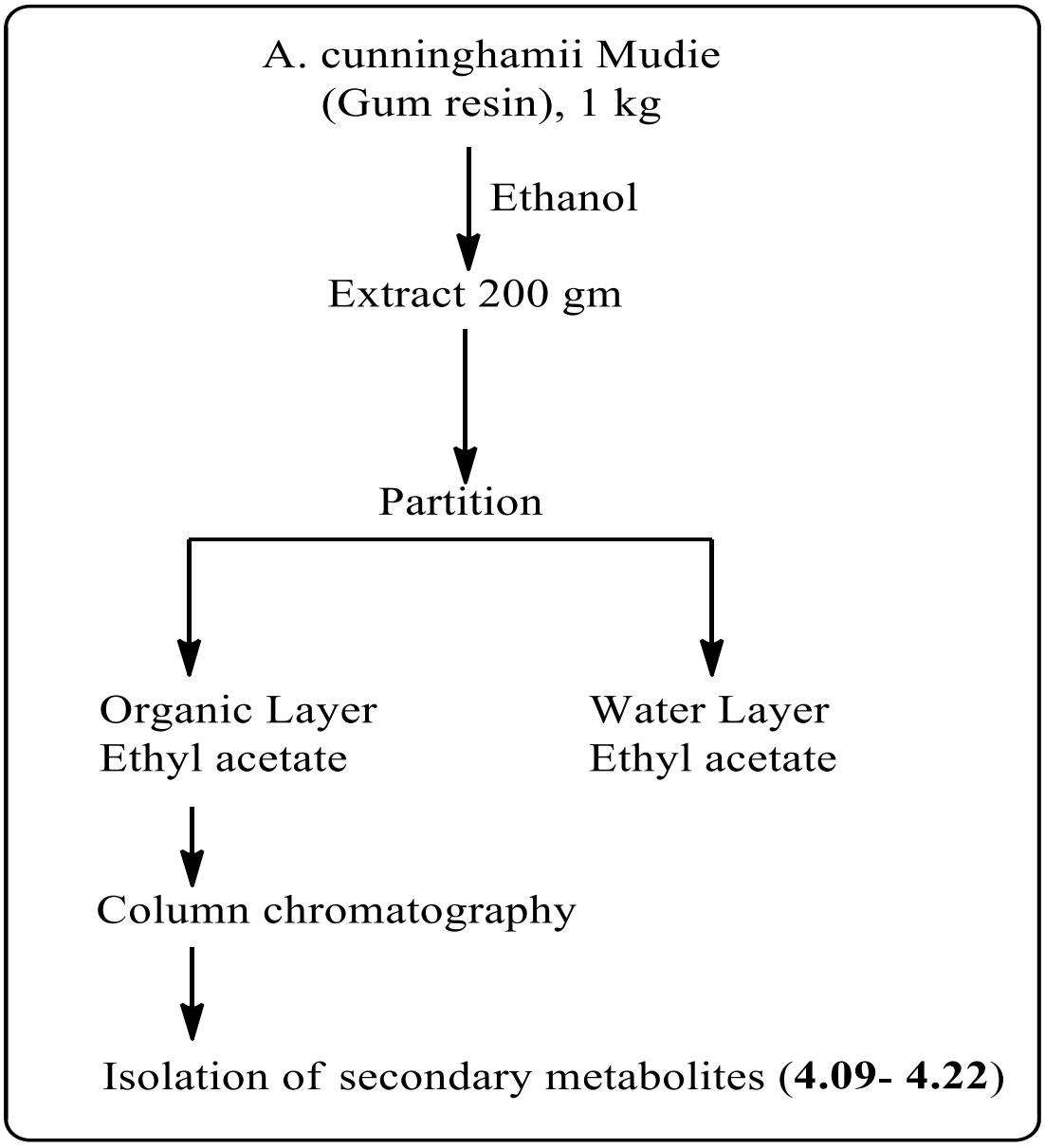


Figure 4.3 Schematic diagram of isolated compounds **4.09-4.22** from LC-MS Gum-resin

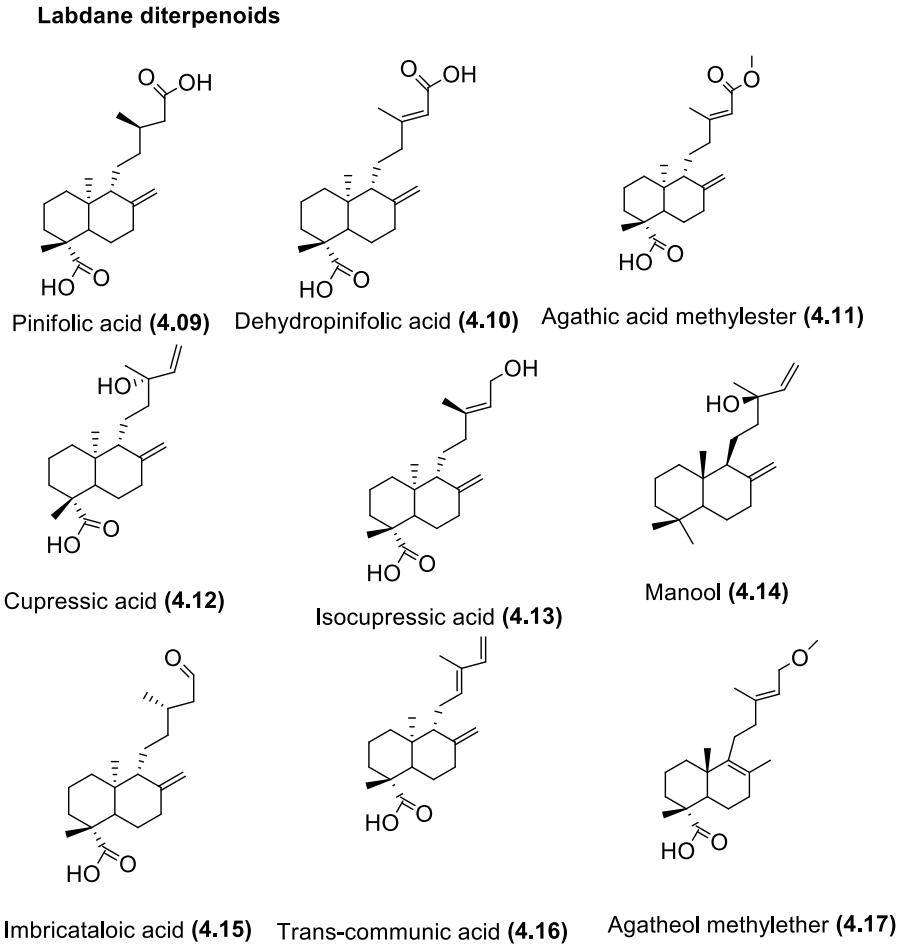


Figure 4. 4 Isolated labdane diterpenoids

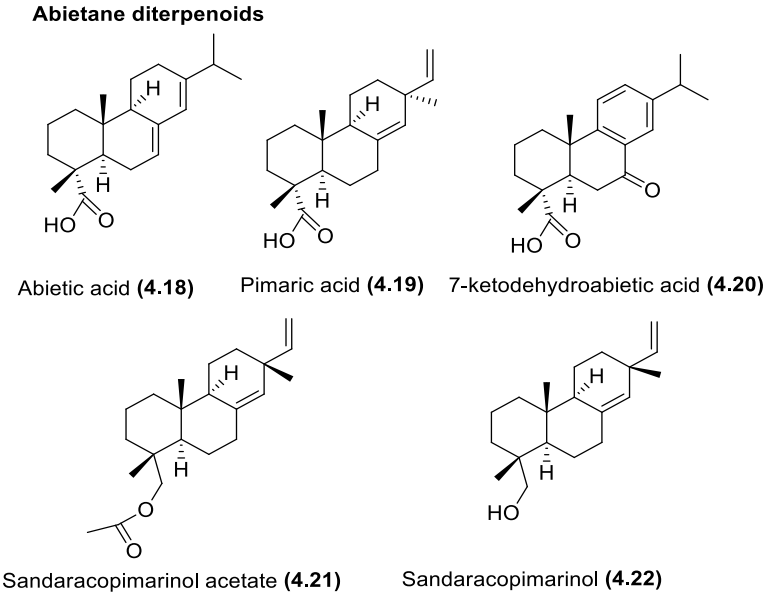


Figure 4. 5 Chemical structure of abietane from of LC-MS gum resin

4.2.2 Structure elucidation of compounds (2.09-2.22)

The repeated column chromatography yielded several pure and impure compounds, and their structures were established by NMR and mass spectrum studies. The structure of agathic acid methylester **(4.11)**, cupressic acid **(4.12)**, abietic acid **(4.18)**, pimaric acid **(4.19)**, 7-ketodehydroabietic acid **(4.20)**, and sandracopimarinol acetate **(4.21)** was confirmed by comparing the obtained NMR data with already reported data [141-145]. The structure of a new metabolite agatheol-methylether **(4.17)** was established by detailed 2D NMR spectral data. Cupressic acid **(4.12)** and pimaric acid **(4.19)** were isolated as closely associated. In LC-MS analysis, the mass *m/z* 335 $[M+H]^+$ was identified at a retention time different from agathic acid methylester, indicating the presence of another compound of mass 334 Da. Vigorous chromatographic isolation led to the purification of this compounds.

The new compound **4.17** was obtained as a gummy liquid and became solid white powder after triturating in acetone. The molecular formula of **4.17** was determined to be $C_{21}H_{34}O_3$ as confirmed by the extensive NMR (^1H, ^{13}C, DEPT-135, ^1H-^1H COSY, HSQC, and HMBC experiments) and HRESIMS at *m/z* 335.2579 $[M+H]^+$ (calculated 335.25807). The ^1H spectrum of **4.17** showed signals for four tertiary methyl singlets (δ_H 1.71, 0.8, 1.59, and 1.27) and an oxymethyl singlet (δ_H 3.35). The ^{13}C spectrum showed the presence of 21 carbon resonances, and with the help of DEPT-135 experiment, they were classified as six quaternary carbons including one carbonyl (δ_C 183.92), eight methylenes, two methine groups, and five methyl groups (δ_C 57.84, 16.52, 17.8, 19.7, 28.65). ^1H-^1H COSY correlations were observed between $H_1/H_2/H_3$, $H_5/H_6/H_7$, H_{11}/H_{12}, and H_{14}/H_{15}, as shown in **Table 4.1**. These data showed the presence of a labdane type diterpenoid. The structure of the compound was further analyzed by 2D NMR spectra. In the HMBC spectrum, the methyl group at δ_H 1.59 showed cross-

peaks to C_7, C_8, and C_9 at δ_C 34.2, 126.9, and 138.8, respectively. The presence of a double bond between C_8 and C_9. another methyl at δ_H 1.27 was correlated with C_3, C_4, C_5, and the carbonyl carbon at δ_C 37.4, 43.78, 53.54, and 183.92, respectively with C_4 being quaternary carbon indicated the presence of a methyl group and a carboxylic group at C_4. Methyl group at δ_H 0.86 was correlated to C_1, C_5, C_{10}, and C_9 at δ_C 37.1, 53.54, 39.8, and 138.8, respectively with C_{10} being quaternary carbon showing the attachment of methyl group to C_{10}. The fourth methyl group at δ_H 1.71 showed correlations to C_{12}, C_{13}, and C_{14} at δ_C 40.12, 141.1, and 120.03, respectively implying the presence of a double bond between C_{13} and C_{14}. Oxymethyl group at δ_H 3.35 was correlated to C_{15} at δ_C 69.01 showing the presence of an -OCH_3 attached to C_{15}. These HMBC correlations defined the structure of the compound as illustrated in **Table 4.1**. The absolute configuration of **4.17** was determined by the ECD analysis and NOESY spectrum. The NOESY correlations between H3-17/H3-18 indicated that H3-17 and H3-18 were in the same plane. While, no any correlations of H-5 with H3-17 and H3-18 were observed, indicating that the junction between ring A and B had the trans-configuration **(Table 4.1)**. Thus, the structure of 2.37 was deduced to be (4R, 5R, 10S) 15-methoxy-labda-8, 13E-dien-19-oic acid and named agatheol methylether (**4.17**).

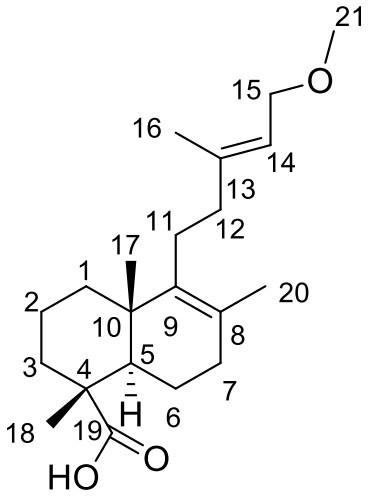
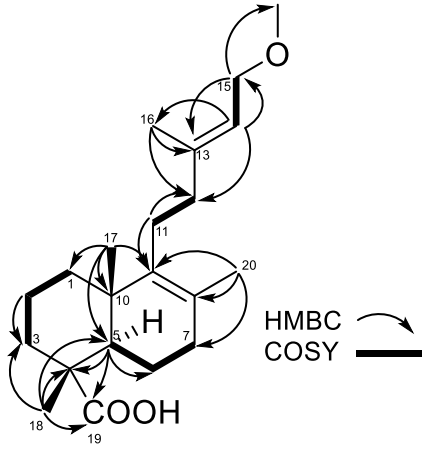
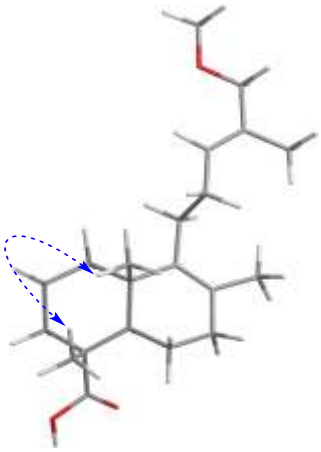


Figure 4. 6 Chemical structure of new compound, agatheol methylether (**4.17**)

Table 4. 1 ^{1}H and ^{13}C NMR (500 MHz; $CDCl_3$) spectroscopic data for new compound (**4.17**)





position	δ_H (ppm)	mult (J in Hz)	δ_C (ppm)	HMBC
1	1.14	m	37.1 (CH_2)	C-17
	1.88	s		
2	1.51	dd (8.7, 4.7 Hz)	19.51 (CH_2)	
	1.86	s		
3	1.00	dd (8.9, 4.3 Hz)	37.4 (CH_2)	C-2, C-18
	2.17	s		
4			43.78 (C)	C-5, C-18
5	1.33	d (11.4 Hz)	53.54 (CH)	C-17, C-18
6	1.80	m	20.7 (CH_2)	C-5
	1.93	s		
7	1.96	d (5.6 Hz)	34.2 (CH_2)	C-20
8			126.9 (C)	C-20
9			138.8 (C)	C-20, C-11, C-17
10			39.8 (C)	C-17
11	1.94	s	26.98 (CH_2)	
	2.11	m		
12	2.05	dd (16.7, 4.8 Hz)	40.12 (CH_2)	C-11, C-14, C-16
13			141.1 (C)	C-15, C-16
14	5.36	dt (22.9, 7.2 Hz)	120.03 (CH)	
15	3.95	dd (11.6, 6.9 Hz)	69.01 (CH_2)	C-14
15-OCH_3	3.35	m	57.84 (CH_3)	C-15
16	1.71	d (16.3 Hz)	16.52 (CH_3)	C-14
17	0.86	d (4.5 Hz)	17.8 (CH_3)	
18	1.27	d (4.0 Hz)	28.65 (CH_3)	
19			183.92 (C=O)	C-5, C-18
20	**1.59**	**d (12.1 Hz)**	**19.7 (CH_3)**	

Detailed spectroscopic data (NMR, MS, and FT-IR) of agatheol methylether 4.17)

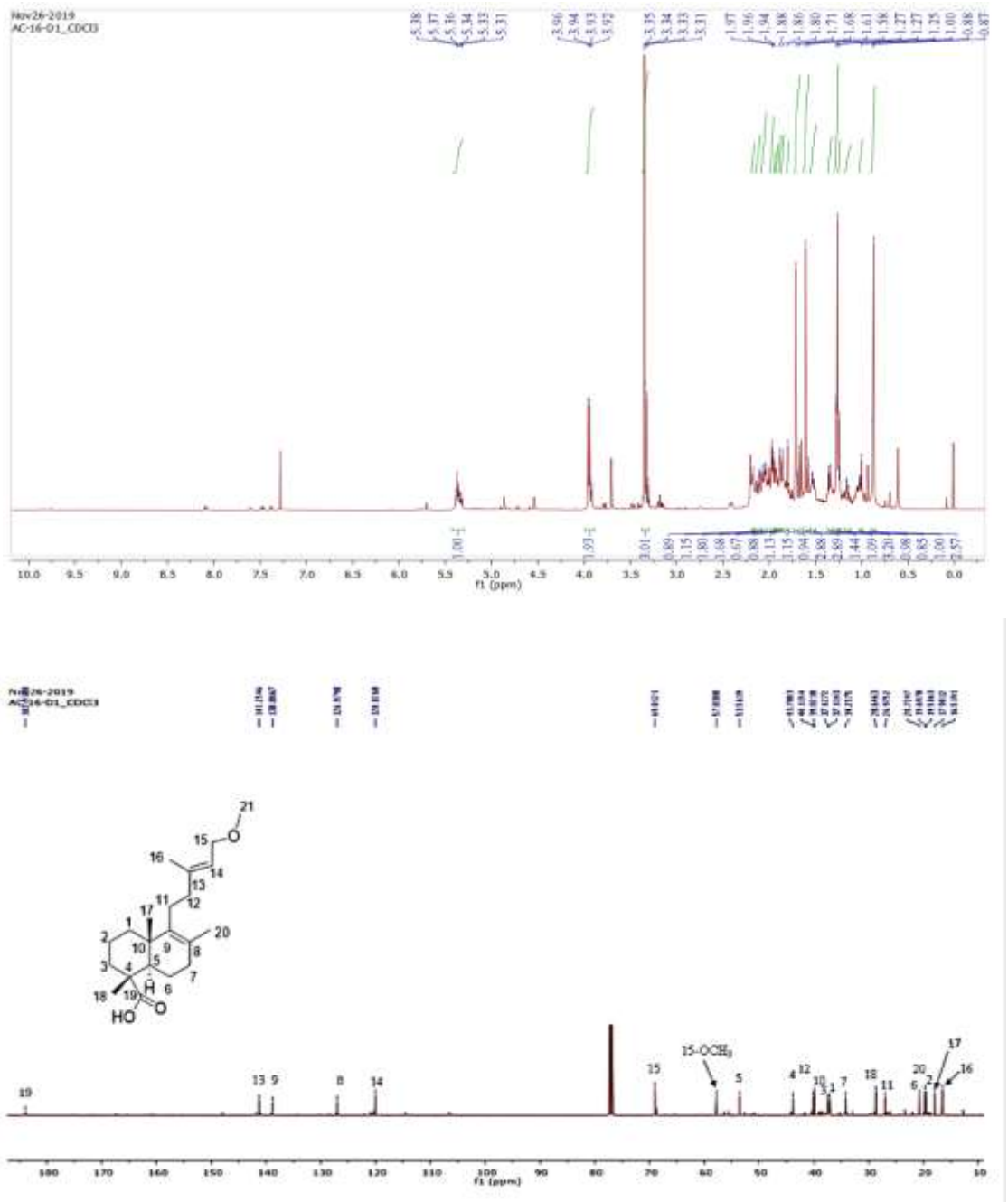


Figure 4. 7 ^{1}H and ^{13}C-NMR (500 MHz; $CDCl_3$) spectrum of agatheol methylether **(4.17)**

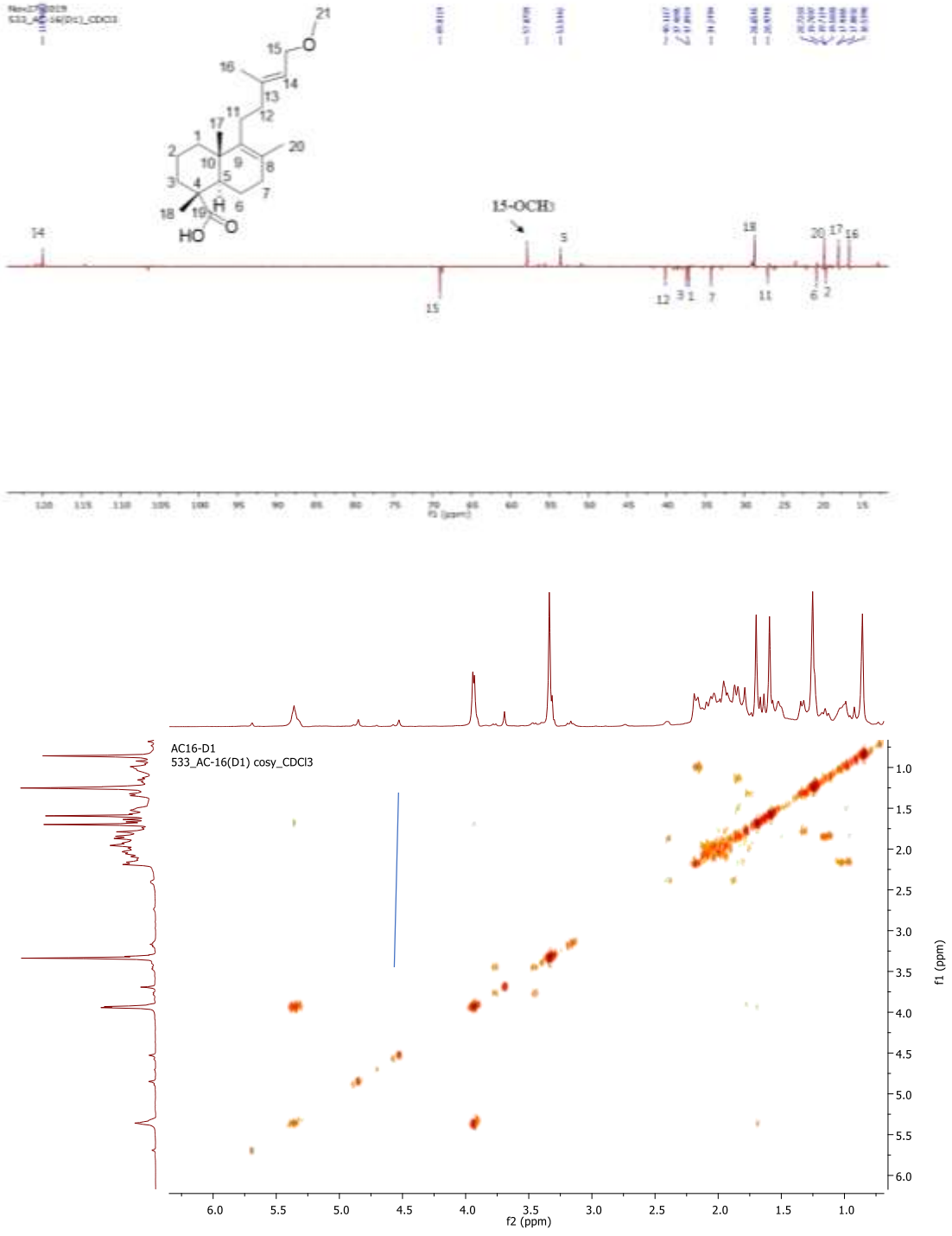


Figure 4. 8 DEPT-135 AND ^{1}H-^{1}H COSY NMR (500 MHz; $CDCl_3$) spectrum of agatheol methylether **(4.17)**

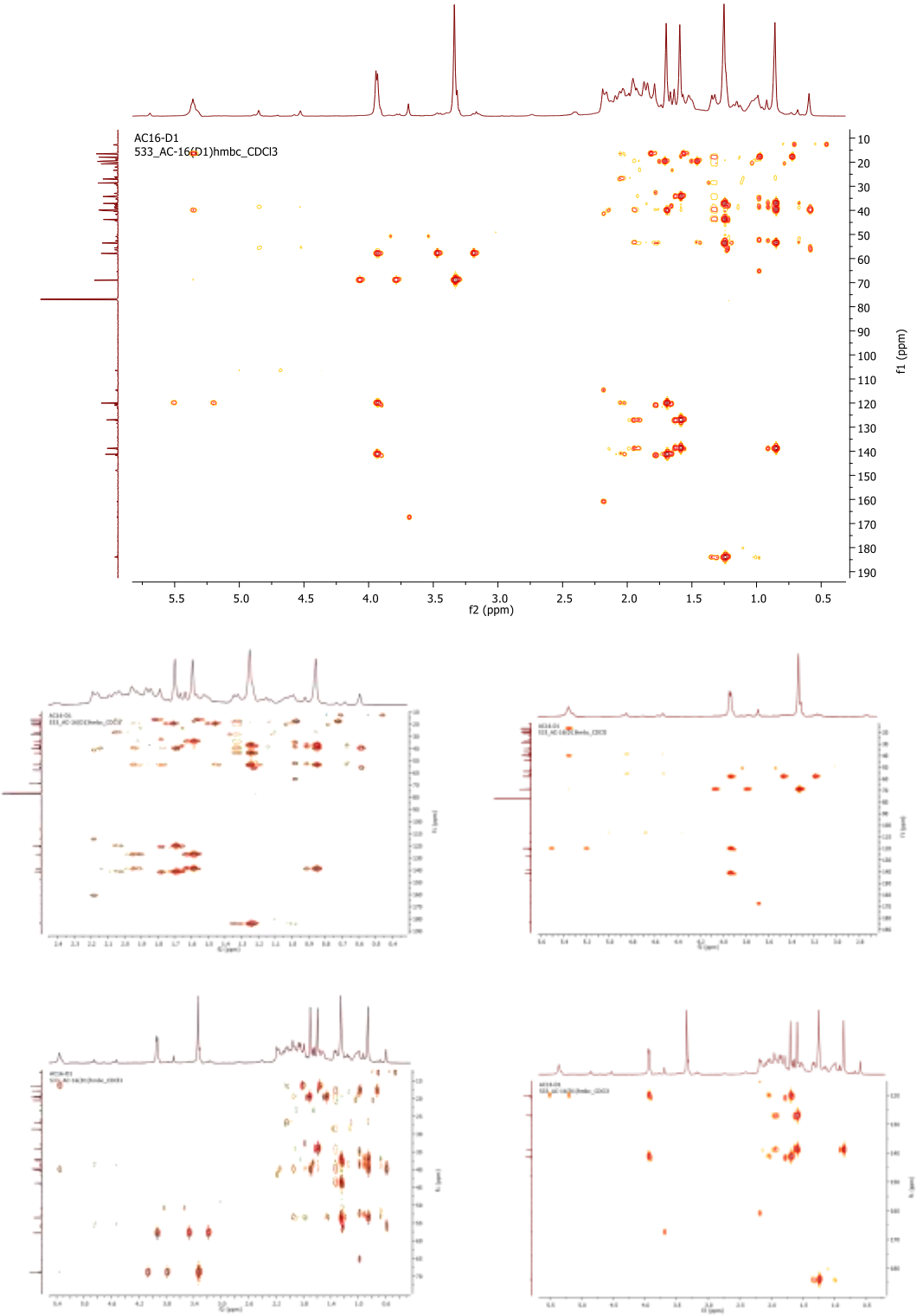


Figure 4. 9 HMBC (500 MHz; $CDCl_3$) spectrum and its expansions of agatheol methylether **(4.17)**

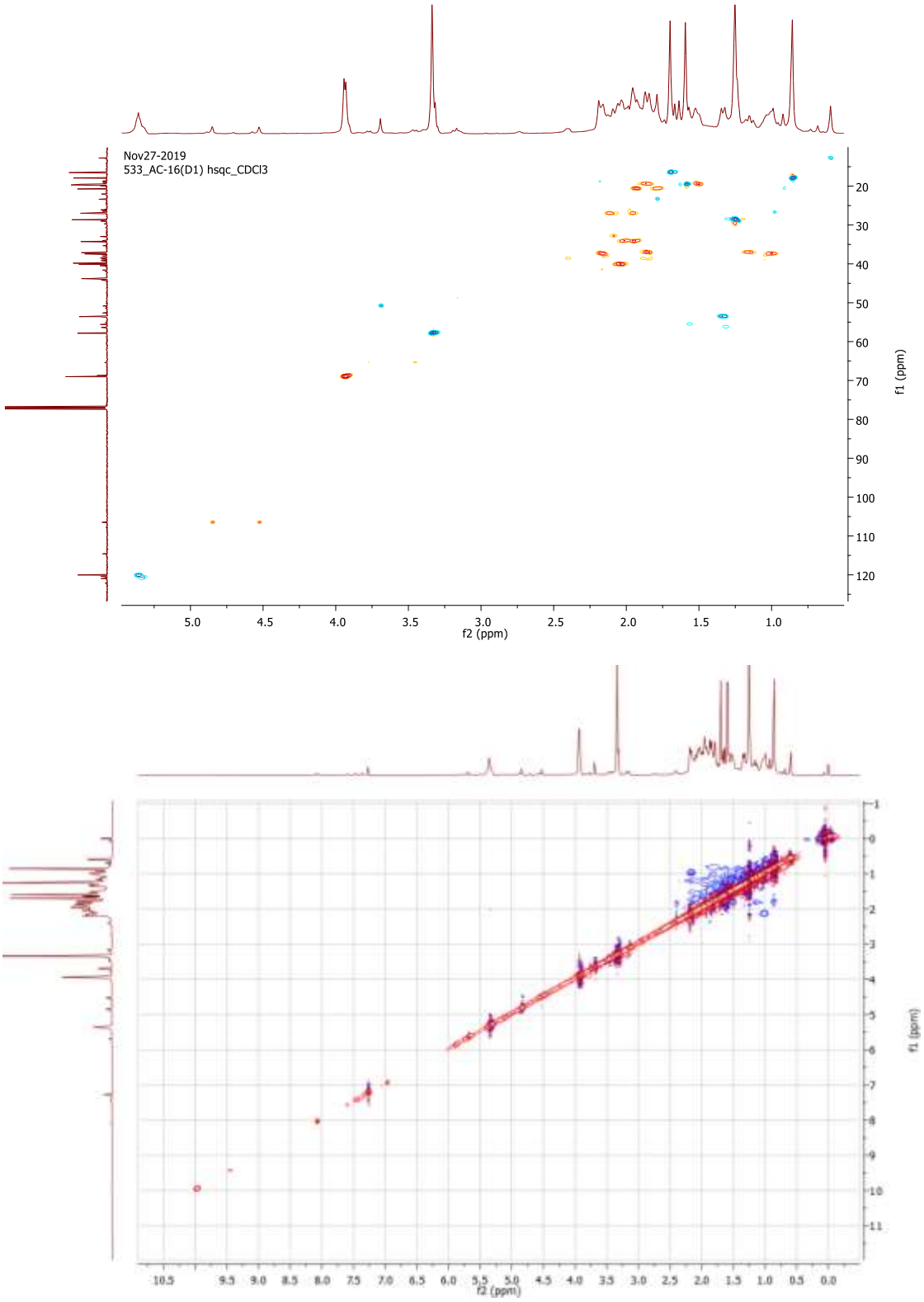


Figure 4. 10 HSQC and NOESY (500 MHz; $CDCl_3$) spectrum of agatheol methylether (**4.17**)

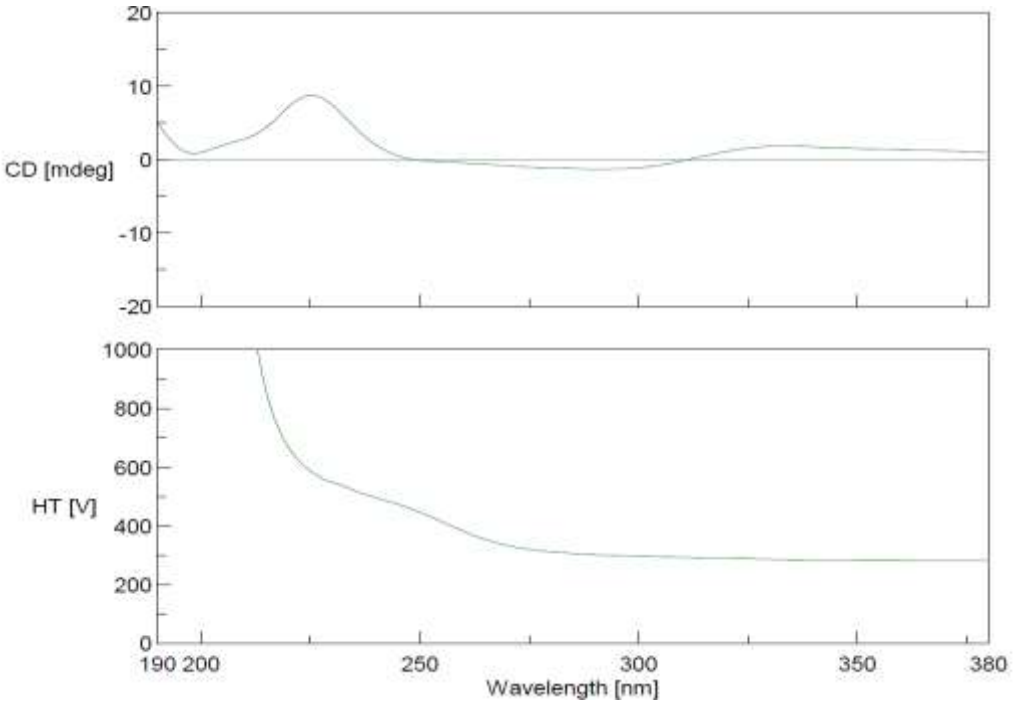


Figure 4. 11 CD spectra of agatheol methylether **(4.17)**

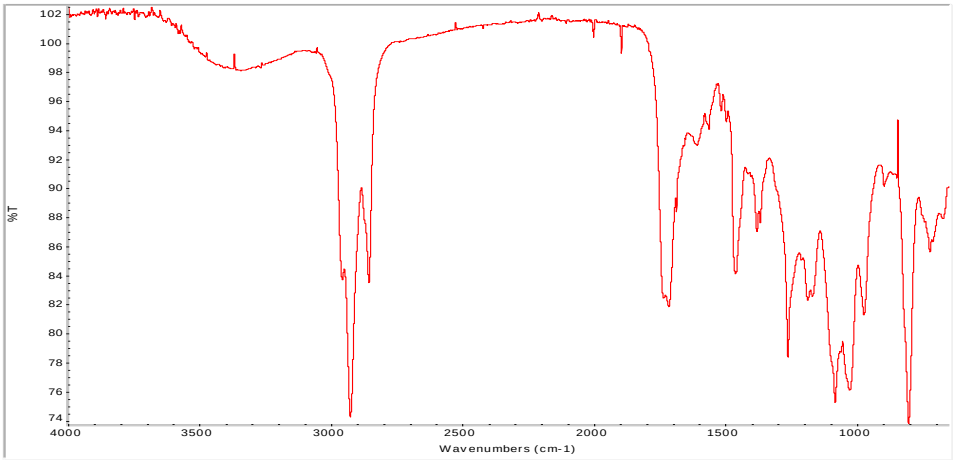


Figure 4. 12 FT-IR spectrum of agatheol methylether **(4.17)**

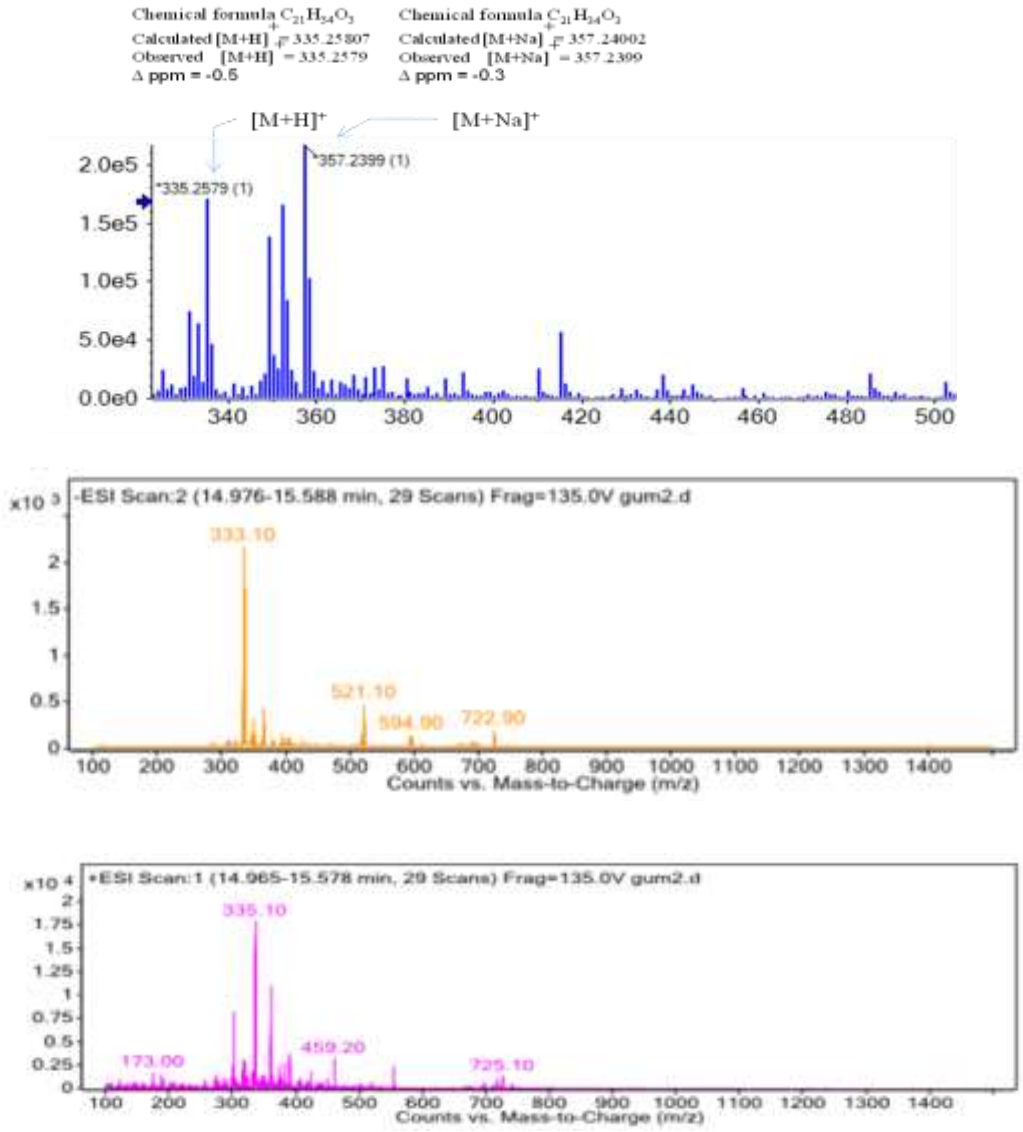


Figure 4. 13 Mass spectrum of agatheol methylether **(4.17)**

As reported earlier, abietane and labdane diterpenoids have shown their potential to be developed as anticancer agents, notably for andrographolide [146-148]. Also, based on the activity of fraction, we were prompted to screen the isolated compounds against human cancer cell lines and further investigate the possible molecular interactions via *in-silico* studies

4.2.3 NMR and LC-MS data of mixture of cupressic acid (4.12) and pimaric acid (4.19)

Both the components of the mixture were characterized as a total of 40 peaks in ^{13}C NMR that hints about the mixture of two diterpenoids. Peak integration of protons in a 2:1 ratio in ^{1}H-NMR corresponds to the ratio of two components present in the mixture. It was also supported by the presence of two different types of peak intensities of carbon signals in ^{13}C NMR. One set of 20 carbon signals was distinguishable from another set of carbon signals. For the first component in ^{1}H-NMR; two broad singlets at 4.5 and 4.8 ppm corresponds to two protons of exocyclic CH_2 exactly similar to agathic acid methylester **(4.11)**. Nine CH_2, 6 CH and/or CH_3, and 5 quaternary carbons were deduced by ^{13}C and DEPT-135 NMR. The two CH_2 appeared as negative signals in DEPT-135 in the aromatic region at 106.6 and 111.6 ppm are the characteristic of an exocyclic douale bond. The rest seven methylenes appeared as negative signals and resembled the reported values of methylenes of agathic acid confirming its labdane skeleton. All the 20Cs of the first set could be used to fix the structure as cupressic acid (**4.12**) [149]. By analysing the ^{13}C and DEPT-135 NMR, 8 -CH_2, 7 -CH and/or -CH_3, and 5 quaternary carbons were confirmed. In the ^{1}H NMR spectrum, four aromatic proton signals appeared in three multiples at 5.18 (2H of terminal CH_2), 5.21 (1H), and 5.89 (1H). Based on the above data, the second compound was identified as pimaric acid (**4.19**), which is in full agreement with the reported spectral data [145]. The mixture was further confirmed by LC-MS data, which showed two peaks corresponding to cupressic acid (**4.12**) at 11.9 minute and the other as pimaric acid (**4.19**) at 12.2 minute.

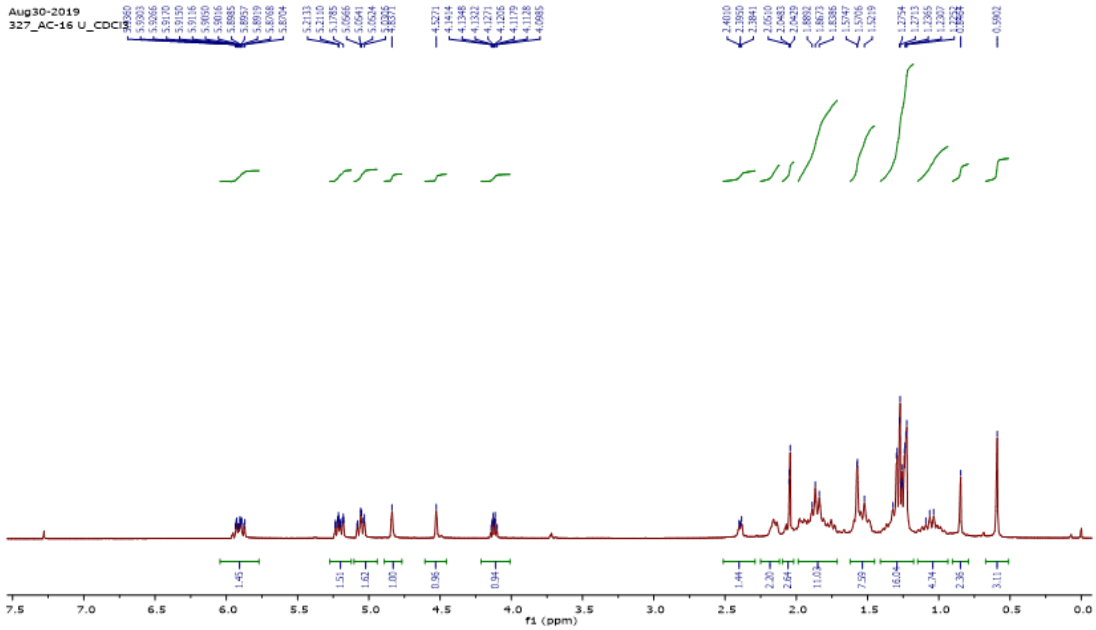


Figure 4. 14 ^{1}H NMR (500 MHz; $CDCl_3$) spectrum of mixture of cupressic acid (**4.12**) and pimaric acid (**4.19**)

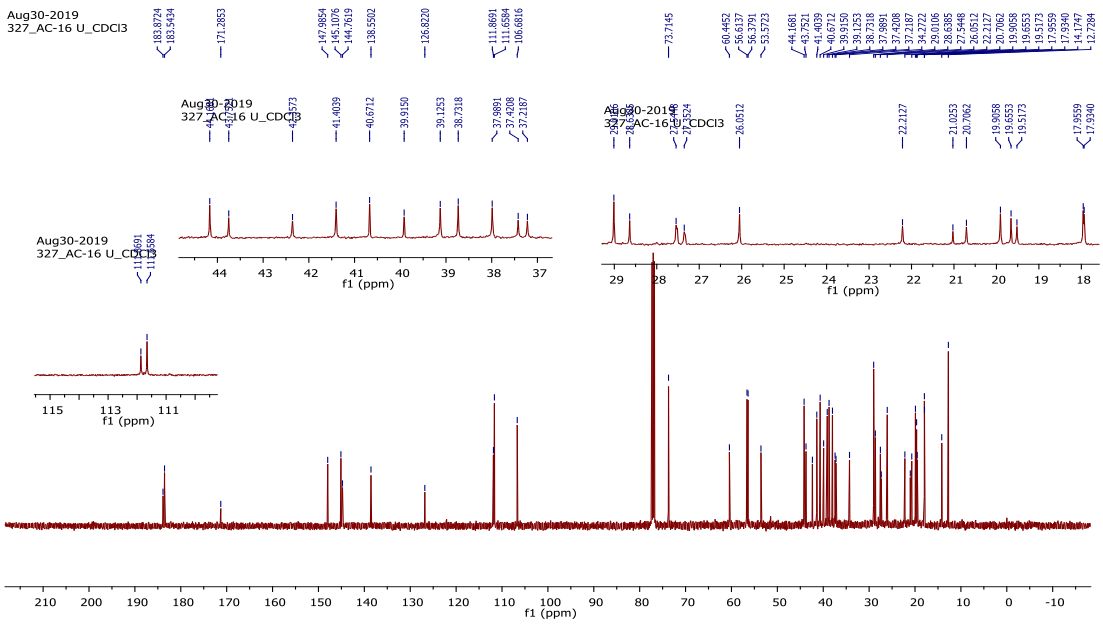


Figure 4. 15 ^{13}C NMR (500 MH; $CDCl_3$) spectrum of mixture of cupressic acid (**4.12**) and pimaric acid (**4.19**)

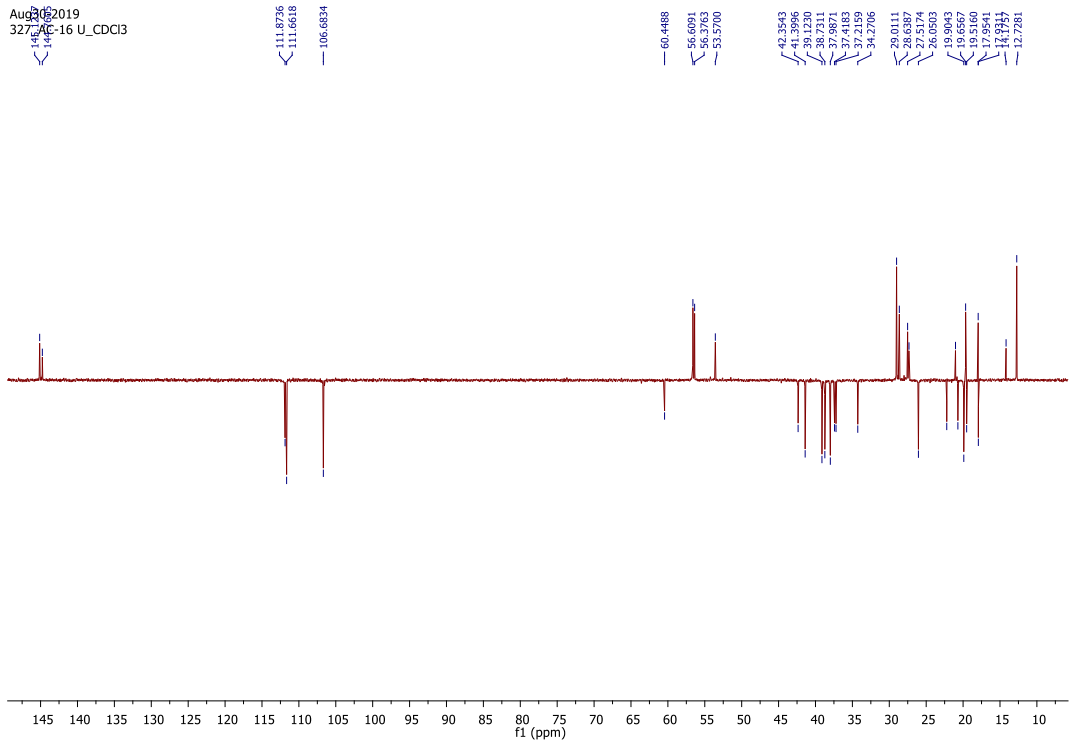


Figure 4. 16 DEPT-135 NMR (500 MHz; $CDCl_3$) spectrum of mixture of cupressic acid **(4.12**) and pimaric acid (**4.19)**

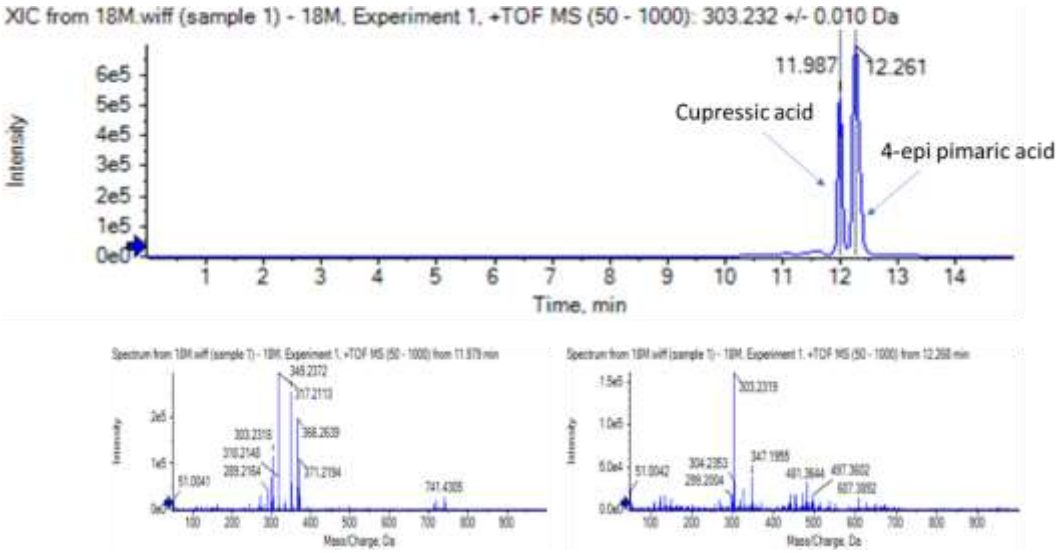


Figure 4. 17 MS-MS spectrums of mixture of cupressic acid (**4.12)** and pimaric acid **(4.19**)

Table 4. 2 LC/MS data of identified compounds **(4.09-4.22)**

Cpds	Name (Mass)	ESI +ve	ESI -ve	HRESIMS	Confirmation by
4.09	Pinifolic acid (336)	319[M-H_2O+H]$^+$ 303 [M-H_2O+H]$^+$	335[M-H]$^-$	[M2OH] = Cal:303.2324 Obs:303.2332 -2.63 ppm	MS/MS
4.10	Dehydropinifolic acid or Agathic acid (334	317[M-H_2O +H]$^+$	333 [M-H]$^-$	[M+H] = Cal: 335.2222 Obs: 335.2240 -5.36 ppm	MS/MS
4.11	Agathic acid methylester (348)	349[M+H]$^+$ 371[M+Na]$^+$	347 [M-H]$^-$	M+H] = Cal: 349.2379 Obs: 349.2390 -3.4 ppm	MS/MS NMR
4.12	Cupressic acid (320)	303[M- H_2O +H]$^+$	319[M-H]$^-$	[M+H] = Cal: 321.2430 Obs: 321.2441 -3.4 ppm	NMR
4.13	Isocupressic acid (320)	303[M-H_2O+H]$^+$	319[M-H]$^-$	[M+H] = Cal: 321.2430 Obs: 321.2441 -3.4 ppm	MS/MS
4.14	Manool (290)	291 [M+H]$^+$	289 [M-H]$^-$	[M-OH] Cal: 274.2650 Obs: 274.2758 -3.9 ppm	MS/MS
4.15	Imbricataloic acid (320)	303 [M-H_2O +H]$^+$	319[M-H]$^-$	[M+H] = Cal: 321.2430 Obs: 321.2441 -3.4 ppm	MS/MS
4.16	Trans communic acid (302)	286 [M+H-OH]	301 [M-H]$^-$	[M+H] Cal: 303.2324 Obs: 303.2334 -3.2 ppm	MS/MS
4.17	Agatheol methylether (334)	335[M+H]$^+$		[M+H] Cal:335.2580 Obs:335.2579 0.5 ppm	NMR
4.18	Abietic acid (302)	303 [M+H]$^+$	301 [M-H]$^-$	Cal:303.2324 Obs:303.2333 -2.73 ppm	NMR
4.19	Pimaric acid (302)	303 [M+H]$^+$	301 [M-H]$^-$	Cal:303.2324 Obs:303.2331	NMR

				-2.60 ppm	
4.20	Ketodehydroabietic acid (314)	315 $[M+H]^+$		Cal:315.1960 Obs:315.1975 4.7 ppm	NMR
4.21	Sandaracopimarinol-acetate (330)	331$[M+H]^+$		Cal:330.2599 Obs:330.2604 1.51ppm	NMR
4.22	Sandaracopimarinol (288)	301 $[M+Na]^+$	287 $[M-H]^-$	Cal: 288.2453 Obs:288.2459 2.08ppm	MS

4.2.4 Biological activity of compounds

4.2.4.1 *In-vitro* cytotoxicity screening of compounds

The crude extract was evaluated for cytotoxicity at a concentration of 10 µg/ml and was found to be active **(Table 4.3)**. Thus, we proceeded further with the fractionation and isolation of pure compounds in search of anticancer leads. The isolated compounds **4.11**, **4.17**, **4.18**, and mixture of **4.12** and **4.19** were subjected to *in-vitro* cytotoxicity study against seven human cancer cell lines *viz.* A549 (lung carcinoma), SCC9 (tongue squamous cell carcinoma), MDA-MB-231 (breast adenocarcinoma), Hs 578T (breast carcinoma), FaDU (pharynx squamous cell carcinoma), MOLT-4 (acute lymphoalastic leukemia), and MCF7 (breast adenocarcinoma), as well as *fR2* (breast epithelial) *cell line*, by employing MTT assay [150, 151]. Paclitaxel was used as a positive control. **Table 4.3** shows that all isolated compounds inhibited the growth of different cancer cell lines to different extents. Compound **4.17** seemed to be most active, as it inhibited the growth by 24-54 % at the dose of 10 µg/mL, with IC_{50} values of 9 µg/mL against Hs 578T and MOLT-4 cells, showing that it can be a potential anticancer lead molecule. At the given concentration, there was no toxicity shown in normal cells (breast epithelial). So, the compound **4.17** was found to be selectively toxic for cancer cells while non-toxic towards the normal cells.

Table 4. 3 % Growth Inhibition of human cancer cell lines at the concentration

Compound	A549	SCC9	MDA-MB-231	Hs 578T	FaDU	MOLT-4	MCF7	Normal Cell (fR2)
Extract	0	51±0.999	25±0.23	42±0.387	43±0.389	26±0.1997	29±0.217	0
4.17	35±0.765	24±0.221	40±0.421	54±0.4987	25±0.212	52±0.478	34±0.312	0
4.11	42±0.412	46±0.432	24±0.198	18±0.0879	30±0.299	47±0.423	25±0.278	0
4.18	15±0.0143	40±0.3998	10±0.098	75±1.123	45±0.412	78±1.213	60±1.012	0
4.12+4.19	42±0.421	46±0.4	24±0.194	18±0.076	30±0.278	ND	25±0.298	0
Paclitaxel[a]	0.158±0.011	ND	0.023±0.001	0.113±0.01	ND	0.225±0.021	0.113±0.011	0.675±0.0571

[a]IC_{50} values of paclitaxel (control) against cancer cells and normal cells. ND: not determined

4.2.4.2 *In-silico* investigation of compound against EGFR and α, β-tubulin

Molecular docking studies of compounds with EGFR and *α, β*-tubulin

Rational for *In-silico* studies are an essential part of modern drug discovery, offering a cost-effective and time-efficient approach to understanding molecular interactions and predicting biological activities. A molecular docking study was performed to explore the mechanism of action of the studied compounds, whether these compounds modulate the activity of known human anticancer targets, *α, β*-tubulin and EGFR (epidermal growth factor receptor). These targets were selected because andrographolide derivatives have shown good binding affinity in earlier studies [152]. The binding energy of compounds ranged from -6.06 kcal mol^{-1} to -7.22 kcal mol^{-1} against *α, β*-tubulin dimer protein, and from -8.8 Kcal mol^{-1} to -7.11 Kcal mol^{-1} against EGFR protein, respectively. The binding energy, ligand efficiency, and interactions of ligands with *α, β*-tubulin protein (PDF: 1JFF) and EGFR (PDB ID: 6DUK) are presented in the **Table 4.4** and T**able 4.5** respectively.

Figure 4. 18 Superimposition of docked structure (green) with co-crystallized ligand (purple)

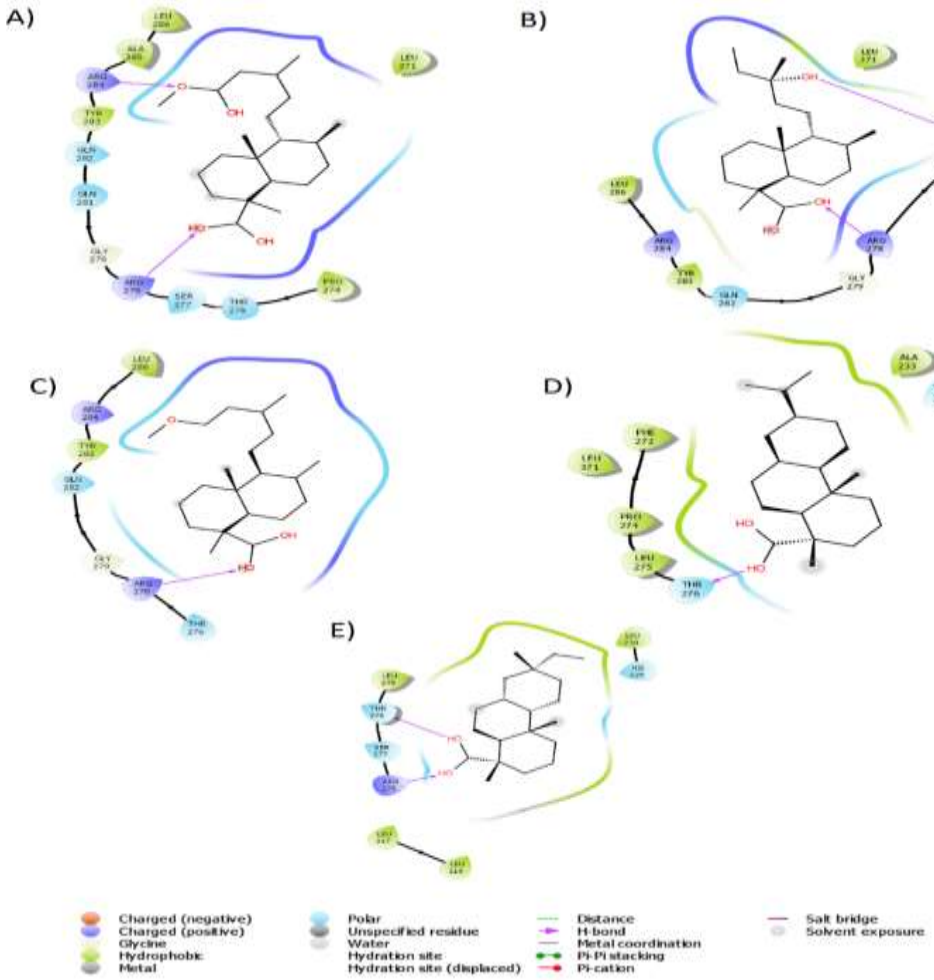


Figure 4. 19 The protein-ligand interaction diagram of compounds A) compound **4.11**, B) compound **4.12,** C) compound **4.17**, D) compound **4.18**, and E) compound **4.19** against α, β-tubulin protein (PDB ID: 1JFF).

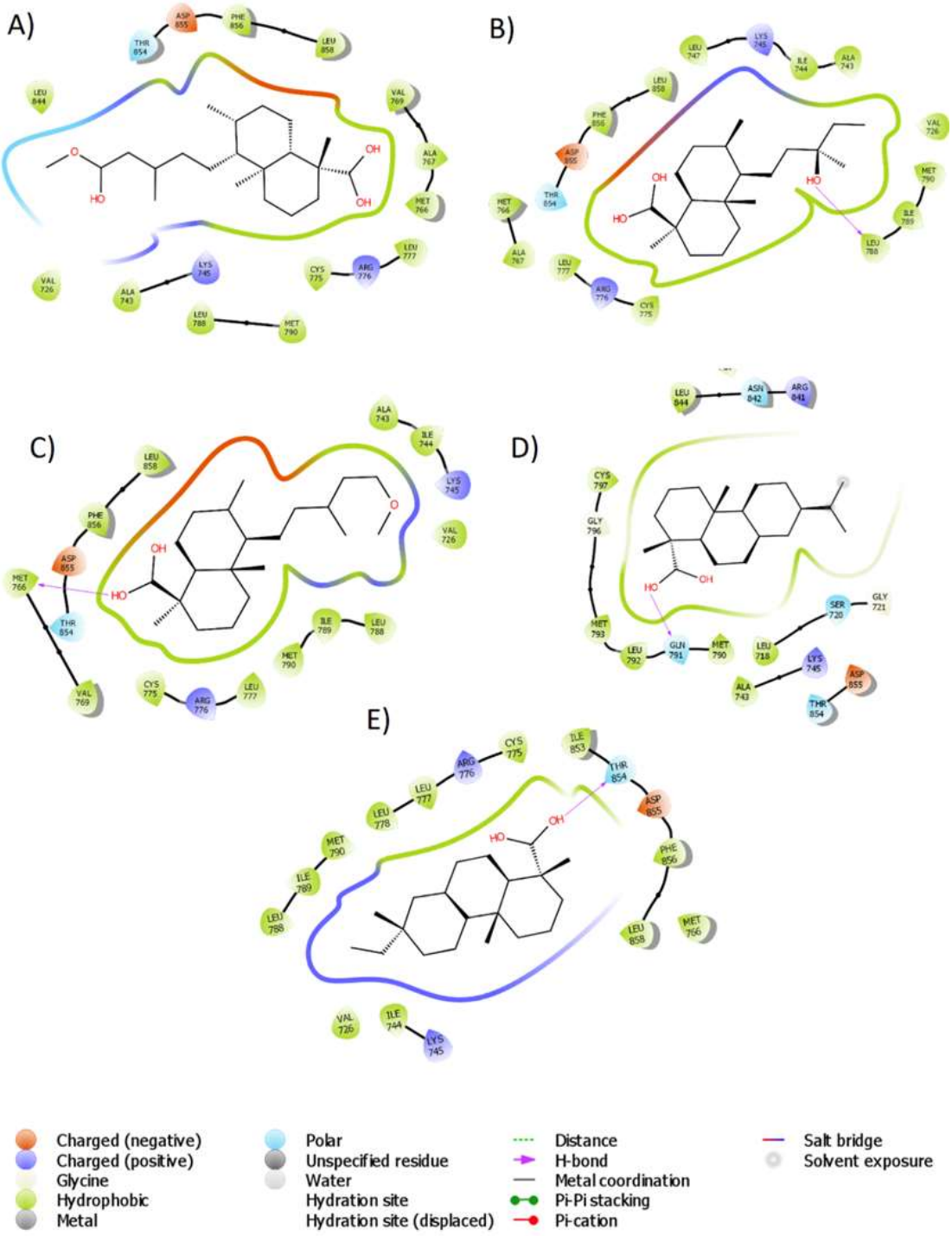


Figure 4. 20 The protein-ligand interaction diagram of compounds A) compound **4.11**, B) compound **4.12**, C) compound **4.17**, D) compound **4.18**, and E) compound **4.19** against EGFR protein (PDB ID: 6DUK).

Table 4. 4 Molecular docking interactions of α, β-tubulin protein (PDB ID: 1JFF) with isolated compounds

Compound	Binding energy (Kcal mol^{-1})	Ligand efficiency (Kcal mol^{-1})	Ligand interactions (PDB id. 1JFF)
4.11	-6.84	-0.274	Arg 284 (H-bond), Arg 278 (H-bond), Ala- 285(Alkyl), Pro 274 (Alkyl)
4.12	-6.27	-0.273	Thr 276 (H-bond), Arg 278 (H-bond), Tyr- 283 (Alkyl), Leu 371(Alkyl)
4.17	-6.58	-0.274	Arg 278 (H-Bond), Tyr 283 (Alkyl), Leu- 286 (Alkyl)
4.18	-7.22	-0.328	Thr 276 (H-Bond), Phe 272 (Alkyl), Pro- 274 (Alkyl), Ala 233 (Alkyl)
4.19	-6.06	-0.275	Thr 276 (H-bond), Arg 278 (H-bond), Pro- 274 (Alkyl), Leu 371 (Alkyl)

Table 4.5 Molecular docking interactions of **EGFR** (PDB ID: 6DUK**)** protein with isolated compound

Compound	Binding energy (in Kcal mol^{-1})	Ligand efficiency (Kcal mol^{-1})	Ligand interactions (PDB ID. 6DUK)
4.11	-8.14	-0.326	Thr 854 (Alkyl), Arg 776 (Charged), Lys 745 (Charged), Met 766 (Alkyl), Ala 743 (Alkyl)
4.12	-7.11	-0.309	Leu 788 (H-Bond), Met 790 (Alkyl), Thr 854- (Polar), Phe 856 (Alkyl)
4.17	-7.33	-0.305	Met 766 (H-Bond), Val 726 (Alkyl), Ile -744(Alkyl), Cys 775 (Alkyl)
4.18	-7.28	-0.331	Gln 791 (H-bond), Cys 797 (Alkyl), Met 793- (Alkyl), Leu 718 (Alkyl)
4.19	-8.8	-0.4	Thr 854 (H-bond), Ile 744 (Alkyl), Asp 855- (Charged)

In the present study, 14 labdane and abietane types of diterpenoids were identified *via* LC-MS-DNP based dereplication study that prompted us for isolation of several compounds viz. **4.11**, **4.12**, **4.17**, **4.18**, **4.19**, **4.10**, and **4.21** from LC-MS Mudie gum-resin. Out of the isolated compounds, **4.17** was found to be a novel compound and was named agatheol methylether **(4.17)**. Few of the isolated compounds i.e. **4.11**, **4.17**, **4.18**,

and a mixture of **4.12** and **4.19** along with extract were evaluated for cytotoxicity against several human cancer cell lines and were found to be active when compared with paclitaxel, the positive control. Moreover, these compounds did not display any cytotoxicity when tested against normal breast epithelial cells. To further deepen the understanding of molecular interactions of these compounds with prominent human cancer target proteins i.e. α, β-tubulin and EGFR proteins, molecular docking studies were also performed. Still, thorough research is required for these molecules to be carried to further the anticancer drug development process.

4.3 Experimental section

4.3.1 General Experimental Procedures

The chemicals were obtained from Sigma-Aldrich and used as received. Optical rotation was measured on JASCO P-2000 polarimeter. The UV spectrum was recorded on Cary 60 UV-Vis spectrophotometer (Agilent Technologies). IR spectra were recorded on Perkin-Elmer IR spectrophotometer. NMR spectra were recorded on Bruker Avance III HD 500 MHz instrument using tetramethylsilane (TMS) as the internal standard and are referenced to the residual proton/carbon in the NMR solvent ($CDCl_3$, 7.26/77.1 ppm; DMSO-d_6, 2.50/39.5 ppm). ESI-MS and HRMS spectra were recorded on Agilent 1100 LC-QTOF and HRMS-6540-UHD machines. LC-ESI-MS/MS analysis was carried out on the Agilent Triple-Quad LC−MS/MS system (model 6410). Analytical HPLC was carried out on the Agilent 1260 infinity equipped Photodiode array (PDA) detector and Merck Chromolith column RP18e (4.6 ×100 mm).

4.3.2 Plant material

The plant material, *A. cunninghamii*, was collected and authenticated by, Dr. Bikarma Singh, Scientist on 07 August 2019, from the Jammu region of India. with specimen sample (accession number, 24176) was preserved in Janaki Ammal Herbarium at the CSIR-Indian Institute of Integrative Medicine, Jammu, India. Presently he is working as Senior scientist CSIR-National Botanical Research Institute, Lucknow, India.

4.3.3 Extraction and isolations

The gum resin (1000 gm) was extracted with 1.0 L absolute ethanol (three times at room temperature, by overhead stirring for 30 min for each cycle). It was filtered to get a clear organic layer and the solvent was evaporated to get viscous crude extract weighing 200 gm (60 % extractive value of dry mass). The dry extract was fractionated by ethyl acetate-water partitioning. The organic soluble fraction was filtered to get a clear solution, which gave 150 gm extract after concentration over the rotary evaporator and 50 gm residual which remained insoluble. Simultaneously, soluble fraction was subjected to LC-MS analysis and chromatographic purification.

4.3.4 LC-MS analysis

The separation was performed at a flow rate of 0.5 mL min^{-1} under a gradient elution of eluent-A (0.1 % formic acid in water) and eluent-B (acetonitrile) as follows 10-90 % of B in 0 to 20 min, 90-95 % of B in 20 to 25 min, 95 % to 10 % of B in 25 to 31 min and hold for 3 min, in a total run time of 34 min. The MS acquiring parameters were followed; the capillary voltage of 3KV in positive mode with gas temperature 300 °C, drying gas 12 L min^{-1}, and nebulizer pressure was 35 psi. The scan source parameter,

skimmer fragmentor, and octapole RF peak were 65 V, 175 V, and 700V respectively. The data acquired with a mass scan range 100-2000 m/z were measured in (+) and (-) ESI modes. MS/MS acquisition of molecular ion peaks was operated in the same parameter using fix collision energy at 30 eV in positive mode for target compounds.

4.3.5 Chromatographic purification

Ethyl acetate fraction (200 gm) was subjected to silica gel VLC (vacuum liquid chromatography) in a sintered glass funnel with a gradient elution of hexane-EtOAc (100:0, 75:25, 50:50, 25:75, 0:100) to get 5 fractions (1-5). The fraction 2 (75 gm) was further subjected to silica gel column chromatography with hexane-EtOAc gradient elution. The pure and semi-pure compounds were obtained by repeated silica and size exclusion column chromatography as discussed in the following section. The 75 gm of fraction 2 was again charged with silica and loaded over the regular #60-120 silica glass column. The adsorbed material was eluted with hexane and ethyl acetate in a gradient manner, increasing from 100% distilled hexane to 100% ethyl acetate. Total 20 fractions were collected (each 50 mL) at the gradient elution of increasing ethyl acetate in hexane. All the fractions were monitored based on the TLC profile and the common fractions were pooled together to obtain 8 fractions. A fraction eluted with 5-10% ethyl acetate in hexane (10 gm) was further loaded to the silica gel column and its isocratic elution with 5% ethyl acetate in hexane yielded a mixture of two compounds (500 mg). The mixture was subjected to repeated Sephadex LH-20 based size exclusion chromatography in methanol. Compound **4.31** was obtained as a single spot-on TLC (350 mg), which was characterized as agathic acid methylether **(4.11)** by comparing its NMR values with reported ones. Compound **4.17** was obtained in a pure form by

repeated chromatography on LH-20 but in a low yield (25 mg). It was identified as a novel compound and characterized as agatheol methylether **(4.17)** with the aid of 2D NMR data along with IR, UV, and HRMS data. A fraction obtained at 15-20 % ethyl acetate (40 gm) showed a major compound along with minor on TLC when visualized under UV 254. A part of this fraction (10 gm) was subjected to silica gel column chromatography and fractionated in isocratic elution using 10% ethyl acetate in hexane. Compound **4.18** (4 gm) was obtained as a white powder and appeared as a single spot-on TLC and was characterized as abietic acid **(4.18)** by comparing its spectroscopic data with the reported data. The rest of the fractions of this column were pooled together and subjected to repeated size exclusion chromatography on LH20 in methanol to get pure compound **4.19** (40 mg), **4.20** (55 mg), and an inseparable mixture of two compounds (**4.12-4.19)** (32 mg). Compounds **4.20** and **4.21** were characterized as 7-ketodehydroabeitic acid and sandaracopimarinol acetate, respectively via comparison of reported spectral data with the obtained data. The inseparable mixture was found to contain cupressic acid **(4.12)** and 4-epipimeric acid **(4.19)** when analyzed by LC-MS and NMR.

Agathic acid methylester (4.11)

White solid; 350 mg; ^{1}H NMR (500 MHz, $CDCl_3$): δ (ppm) 5.66 (s, 1H), 4.89 (s, 1H), 4.53 (s, 1H), 3.71 (s, 3H), 2.44-2.42 (m, 1H), 2.34-2.29 (m, 1H), 2.17 (s, 3H), 2.02-1.98 (m, 2H), 1.92-1.88 (m, 2H), 1.87-1.86 (m, 1H), 1.85-1.83 (m, 1H), 1.74-1.68 (m, 2H), 1.60-1.58 (m, 1H), 1.56-1.51 (m, 2H), 1.36-1.33 (m, 1H), 1.26 (s, 3H), 1.10 (m, 1H), 1.08-1.03 (m, 2H), 0.62 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ (ppm) 184.33, 167.33, 160.99, 147.62, 115.00, 106.57, 56.25, 55.34, 50.80, 44.20, 40.47, 39.76, 39.06, 38.66,

37.86, 28.99, 26.04, 21.67, 19.87, 18.89, 12.78; ESIMS *m/z* 349 $[M+H]^+$, 347 $[M-H]^-$; HRESIMS *m/z* 349.2390 $[M+H]^+$ (calculated for $C_{21}H_{33}O_4$, 349.2379).

Agatheol methylether (4.17)

White solid; 25 mg; Melting point, 152-156 °C (uncorrected); $[\alpha]_D^{21}$ +18.8 (*c* 0.70, $CHCl_3$); UV (MeOH) λ_{max} 227 nm; IR (KBr) ν_{max} 2920, 2840, 1700, 1490, 1420, 1260, 1090, 1025, 800 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) and ^{13}C NMR (125 MHz, $CDCl_3$) data, see **Table 4.1;** ESIMS *m/z* 335 $[M+H]^+$; HRMS m/z 335.2579 $[M+H]^+$ (calculated for $C_{21}H_{35}O_3$, 335.25807).

Abietic acid (4.18)

White powder; 4 gm; ^1H NMR (500 MHz, $CDCl_3$): δ (ppm) 5.78 (s, 1H), 5.37 (m, 1H), 2.21 (p, 1H), 2.11 (s, 1H), 2.07 (m, 2H), 1.92 (d, 1H), 1.89 (d, 2H), 1.84 (m, 2H), 1.69 (m, 2H), 1.61 (m, 2H), 1.27 (s, 3H), 1.23 (2H), 1.03 (d, 3H), 1.02 (d, 3H), 0.84 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ (ppm) 184.96, 145.28, 135.57, 122.39, 120.49, 50.94, 46.34, 44.90, 38.27, 37.18, 34.88, 34.45, 27.45, 25.61, 22.47, 21.42, 20.86, 18.05, 16.71, 14.03; ESIMS *m/z* 303 $[M+H]^+$, 301 $[M-H]^-$; HRESIMS *m/z* 303.2333 $[M+H]^+$ (calculated for $C_{20}H_{31}O_2$, 303.2324).

***7*-keto-dehydroabietic acid (4.20)**

White solid; 40 mg; ^1H NMR (500 MHz, $CDCl_3$): δ (ppm) 7.89 (d, 1H), 7.42 (dd, 1H), 7.31 (d, 1H), 2.94 (p, 1H), 2.75 (m, 2H), 2.50 (d, 1H), 2.38 (m, 2H), 1.37 (s, 3H), 1.28 (s, 3H), 1.27 (d, 3H), 1.25 (d, 3H), 1.02-1.05 (m, 1H), 0.94-0.95 (m, 1H), 0.89-0.91 (m, 1H), 0.86 (m, 1H); ^{13}C NMR (125 MHz, $CDCl_3$): δ (ppm) 198.11, 153.01,146.92, 132.69, 130.64, 125.11, 123.49, 46.36, 43.59, 37.78, 37.24, 37.08, 36.52, 33.59, 23.81,

23.74, 23.64, 18.14, 16.18; ESIMS *m/z* 315 $[M+H]^+$; HRESIMS *m/z* 315.1975 $[M+H]^+$ (calculated for $C_{20}H_{27}O_3$, 315.1960).

Sandaracopimarinol acetate (4.21)

White solid; 55 mg; ^{1}H NMR (500 MHz, $CDCl_3$): δ (ppm) δ 5.77 (dd, J = 17.4, 10.5 Hz, 1H), 5.22 (s, 1H), 4.94 – 4.86 (m, 2H), 3.87 (d, J = 10.9 Hz, 1H), 3.66 (d, J = 11.0 Hz, 1H), 2.27 – 2.21 (m, 1H), 2.07 (s, 3H), 2.06 – 2.04 (m, 1H), 1.74 (dt, J = 12.8, 8.0 Hz, 2H), 1.55 – 1.43 (m, 6H), 1.36 – 1.34 (m, 1H), 1.33 (d, J = 4.1 Hz, 1H), 1.32 – 1.24 (m, 3H), 1.04 (s, 3H), 1.02 (s, 1H), 0.88 (s, 3H), 0.84 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ (ppm) 171.37, 149.05, 136.76, 128.88, 110.06, 73.10, 50.62, 48.68, 38.83, 38.11, 37.40, 36.63, 35.98, 35.66, 34.54, 25.94, 22.53, 21.04, 18.76, 18.20, 17.87,15.55; ESIMS *m/z* 331 $[M+H]^+$.

4.3.6 Cytotoxic activity (Cell viability assay)

To study the cytotoxic effect of isolated compounds, cell viability was analysed using MTT colorimetric assay against different human cancer cell lines (A549, SCC09, MDAMB231, HS578T, FaDU, Molt4, and MCF7) and the breast epithelial *cell line (fR2)*. In this assay, cells were incubated with a fixed concentration (10 μg) of compounds for 72 hours, followed by the addition of MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, a tetrazole) is reduced to purple formazan in living cells and further incubation for 2 hours. The purple formazan formed was dissolved using a solubilisation solvent (usually either dimethyl sulfoxide, an acidified ethanol solution; or a solution of the detergent sodium dodecyl sulfate in diluted hydrochloric acid). The absorbance of this coloured solution was recorded at a certain wavelength (usually between 500 and 600 nm) by a UV spectrophotometer, to measure

the % growth inhibition. The absorption maximum is dependent on the solvent employed [150, 151, 153].

4.3.7 Molecular docking

Molecular docking studies were performed using AutoDock 4.2 to deepen the understanding of the molecular interactions of active compounds and known human anticancer targets *α*,*β*-tubulin (PDB ID: **1JFF**) and EGFR (PDB ID: **6DUK**) [154]. The crystal structures of these proteins were obtained from the protein data bank [155]. The correct protonation state to the residues was assigned using the pdb2pqr web server. All the water molecules, ligands, and ions were removed. Non-polar hydrogen atoms were removed, and gasteiger charges were added using M.G.L Tools 1.5.6. AutoDock employs Autogrid4 to compute maps. The active site of the protein was determined using the PLIP web server [156]. The docking study was performed using the Lamarckian Genetic Algorithm (LGA). The docking was performed with 100 runs, 150 population size, 27,000 number of generations, and 2,500,000 number of energy evaluation. It employs a 'semiempirical free energy force field' to evaluate conformations at the time of docking simulation. The docking protocol validation was done by re-docking the co-crystallized ligand and calculating RMSD between docked pose and co-crystallized ligand. The docked pose was visualized by Discovery studio 2019 for studying interactions. Upon successful completion of docking simulation, the best confirmation was selected with the best binding energy in the largest cluster of 2.0 Å