

Crispene E, a *cis*-clerodane diterpene inhibits Stat3 dimerization in breast cancer cells

Julia Mantaj^{a ‡}, S. M. Abdur Rahman^{*b ‡}, Bishwajit Bokshi^b, Choudhury M. Hasan^{c †}, Paul J. M. Jackson^a, Richard B. Parsons^a and Khondaker M. Rahman^{a*}.

^aInstitute of Pharmaceutical Science, King's College London, London SE1 9NH.

^bDepartment of Clinical Pharmacy and Pharmacology, Faculty of Pharmacy, University of Dhaka, Dhaka-1000, Bangladesh.

^cDepartment of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Dhaka, Dhaka-1000, Bangladesh.

Supporting Information

1.0 Expression and Purification of STAT3

1.1 Growth of *E. coli* cells containing the STAT3 β tc plasmid

An aliquot of the master cell bank (BL21 Rosetta cells transformed with the STAT3 β tc plasmid kindly provided by Müller and co-workers) was grown overnight at 37°C in the presence of ampicillin and chloramphenicol. One of the seed cultures was used to inoculate a 10L Electrolab fermenter in the presence of 100 μ g ampicillin and 20 μ g chloramphenicol. The seed was gently stirred in the fermenter at 250 rpm and 37 °C, with a suitable air flow (3 L/min). The cells were induced with 1 mM IPTG when an OD of 0.6 was reached. The temperature was then reduced to 21°C, and the bacteria grown overnight before being harvested by centrifugation at 3600 x g for 25 minutes at 4 °C.

1.2 Extraction of the unphosphorylated STAT3 β tc protein

To each gram of pellet, 10 mL of extraction buffer (20 mM Hepes-KOH, pH 7.6, 0.1 M KCl, 10% glycerol, 1mM EDTA, 10 mM MnCl₂, 20 mM DTT, 0.5 mM PMSF, protease inhibitor cocktail tablet) was added to re-suspend the pellet. The sample was placed into a chilled beaker of ice and sonicated for 5 mins at 15 mA, repeating for 15 seconds on followed by 15 seconds off, using a medium-sized sonicator probe (Camlab Trans-sonic TS701 H, Cambridge, UK). The solution was then centrifuged (Beckman Coulter Centrifuge, JA25.50 rotor, 4°C, 1 hour, 27,000 x g), and the supernatant collected and chilled prior to protein precipitation using ammonium sulphate. The mixture was then centrifuged (Beckman Coulter centrifuge, JA25.50 rotor, 4 °C, 1 hour, 27,000 xg), and the pellet stored at 4 °C until required.

1.3 Purification of the unphosphorylated STAT3 β tc protein

Salt-free buffer was first prepared (100 mM Tris pH 8.5, 1 mM EDTA, 2 mM DTT), and then ion exchange buffers were prepared from the salt-free buffer with increasing concentrations of NaCl (0.1 M, 0.2 M, 0.3 M, 0.4 M, 1.0 M) to elute the protein.

The pellet was dissolved in 5 mL salt-free buffer, and the solution was filtered through a 0.22 μ m filter. The filtrate was then diluted to 20 mL using double distilled H₂O. Approximately 5 mL of the dissolved protein was passed through an ion-exchange (IE) column (GE Healthcare, Little Chalfont Buckinghamshire, UK) and stored on ice during collection. The IE column was washed with 10 mL salt-free buffer before eluting the protein from the column with successive volumes (20 mL) of each of the ion exchange buffers, increasing the concentration of NaCl each time. Once the purification process was complete, the column was washed again with 20 mL salt-free buffer before being re-used.

The purified protein was dialyzed overnight in either 10 mM ammonium acetate (to collect ESI data) or 100 mM ammonium bicarbonate (for digestion purposes).

1.5 FP Assay Protocol

A 96 well plate CORNING black plate was thoroughly washed with distilled water and allowed to dry. 100 nM FAM-LPQTV working solution was prepared from 10 μ M FAM-LPQTV stock solution (in DMSO) using PBS pH 7.4 Buffer. 10 μ l FAM-pYLPQTV was added to 350 nM uSTAT3 protein, final concentration of protein in each well, in 90 μ l PBS buffer pH 7.4. The MP value of the surrogate dimer complex was measured using a fluorescent plate reader (Envision, Perkin Elmer, USA). This provided the base MP value. At this point the inhibitor solution was added to the well and the assay plate was left on the shaker for 5 minutes. The MP value for each well was again measured by the fluorescent plate reader. This represented a shift in fluorescent value due to the displacement of the fluorescently-labelled surrogate peptide by the inhibitors (Crispene E, STA-21 and pYLKTKF) and the average inhibition for each inhibitor was calculated as follows -

Analysis of results

Base MP value (X) for each well = MP value of Protein and Probe (350 nM STAT3 + 10 nM FAM-pYLPQTV) - MP value of free FAM-pYLPQTV

Inhibitor MP value (Y) for each well = MP value of Protein +Probe + Inhibitor (350 nM uSTAT3 + 10 nM FAM-pYLPQTV + 2 μ L Inhibitor) - MP value of free FAM-pYLPQTV with 2 μ L DMSO

% inhibition by any ligand/peptide = $(X-Y)/X*100$

For comparison purpose, the % inhibition produced by 100 μ M LKTKFI is considered as 100% and the inhibition produced by different ligands relative to 100 μ M LKTKFI is measured

% Inhibition produced by 100 μ M LKTKFI = A

% Inhibition produced by 100 μ M Ligand = B

Relative inhibition (%) = $B/A*100$

2.1 MTT Assay

The cells were grown in normal cell culture conditions at 37 °C under a 5% CO₂ humidified atmosphere using appropriate medium. The cell count was adjusted to 10⁵ cells/ml and 5,000-20,000 cells were added per well depending on the cell line. The cells were incubated for 24 hours and 1 µl of the appropriate inhibitor concentrations were added to the wells in triplicates. After 96 h of continuous exposure to each compound, the cytotoxicity was determined using the 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Lancaster Synthesis Ltd, UK) colorimetric assay.¹³⁴ Absorbance was quantified by spectrophotometry at $\lambda = 570$ nm (Envision Plate Reader, PerkinElmer, USA). IC₅₀ values were calculated by a dose-response analysis using the Prism GraphPad Prism® software.

Table S1: IC₅₀ values (µM) determined after 24 hours exposure for the Crispene E in a panel of human cancer cell lines (HeLa, MIA PaCa2 and NCI H1975).

Compound	IC ₅₀ (µM)		
	HeLa (Cervical)	MIA PaCa2 (Pancreatic)	NCI H1975 (Non small cell Lung)
1	10.5 ± 1.3	8.3 ± 2.3	11.8±2.6

2.2 Trypan Blue Assay

MDA-MB-231 cells were plated in 24 well plates overnight to achieve 80% confluency. The cells were then incubated with inhibitor for 24 hours. Unstained (viable) and stained (non-viable) cells were counted and calculated as a percentage of total cells using a haemocytometer.

3.1 General procedure for RT-PCR detection of *cyclin D1*, *NNMT*, *fascin* and *STAT3* mRNA

1x10⁶ cells were plated into 10% FBS media in 6-well plates and incubated until 70% confluence was achieved. LPS (500 µg/ml) was added for 24 h followed by addition of 100 µM Crispene E for another 24 h. Total RNA was extracted from whole cell extracts using the RNeasy® Plus Mini Kit (Quiagen, Hilden, Germany) according to the manufacturer's instructions. Isolated total RNA was reverse-transcribed using the High Capacity RNA-to-cDNA kit (Applied Biosystems, California, USA) as per manufacturer's instructions followed by cDNA amplification utilizing PCR with primers

for *cyclin D1*, *Bcl-2*, *NNMT*, *STAT1*, *STAT3*, *fascin* and *GAPDH* (listed in Table S2). Resulting PCR products were electrophoresed using a 1.5% agarose gel and visualised using the U-GENIUS Gel-DocIt system.

3.2 General procedure for Western Blots

1x10⁶ cells were plated into 10% FBS media in T-75 flasks and incubated until 70% confluence was achieved. IL-6 (20 ng/ml) was added for 24 h followed by addition of 5 µM Crispene E for another 24 h. Cells were lysed using RIPA buffer with protease and phosphatase inhibitors. Proteins were separated by SDS-PAGE electrophoresis (4-12% NuPAGE gel for 50 min at 200 V). Blots were carried out for fascin, pSTAT3, pSTAT1 and α tubulin (Abcam) and visualized using electrochemiluminescent (ECL) detection (GE Healthcare Life Sciences).

Table S2. Primers used during RT-PCR of MDA-MB-231 cell line

Gene	Sequence
<i>Cyclin D1</i>	Forward: 5'-ACA GAT CAT CCG CAA ACA CG-3' Reverse: 5'-CTC CTC CTC TTC CTC CTC CT-3'
<i>Bcl-2</i>	Forward: 5'-CAT CTT CTC CTC CCA GCC C-3' Reverse: 5'-CCG AAC TCA AAG AAG GCC AC-3'
<i>NNMT</i>	Forward: 5' -TGG CCC CAC TAT CTA TCA GC-3' Reverse: 5' -CCT CTT TCA CAG CAG CCT CT-3'
<i>STAT3</i>	Forward: 5'-CTT TGA GAC CGA GGT GTA TCA CC-3' Reverse: 5'-GGT CAG CAT GTT GTA CCA CAG G-3'
<i>Fascin</i>	Forward: 5'-ACT GGC TAC ACG CTG GAG TT-3' Reverse: 5'-GGA AGG CAC ACT TTT TGG TG-3'
<i>GAPDH</i>	Forward: 5'- AGC CAC ATC GCT CAG ACA C-3' Reverse: 5'- ACC CAA TAC GAC CAA ATC C-3'

4.0 Molecular Docking

4.1 Methodology

4.1.1 Receptor Preparation

The STAT3 homodimer structure was downloaded from the Protein Data Bank (PDB ID: 1BG1) and missing residues and loop structures were generated using the SwissModel online tool (<http://swissmodel.expasy.org/interactive>).

The STAT3 homodimer was then subjected to a number of steps of preparation in *Chimera*, including assignment of partial charges (AMBER ff98SB) and removal of hydrogens using the *DockPrep* module of AMBER, followed by writing to mol2 and pdb files. A molecular surface of the receptor was generated using *write dms*.

4.1.2 Ligand Preparation

Crispene E was constructed and energy minimised using ChemBioOffice and exported in pdb format and converted into the Sybyl format mol2 using *Chimera* (1). The *DockPrep* module in *Chimera* was then used to assign partial charges of Crispene E (*AMI-BCC* charges in this instance), and atom types were subsequently assigned *via* the AMBER GAFF force-field using ANTECHAMBER.

4.1.3 Docking

Docking experiments were undertaken using the DOCK6 software suite and the STAT3 monomeric crystal structure (PDB ID: 1BG1) was used in the study. In each case, homodimer A was used as the receptor and the PYLKTKFI peptide motif of homodimer B (created through unit cell generation in *Chimera*) was considered the ligand.

The receptor was prepared (outlined above), and a number of steps were undertaken in DOCK6 to isolate the binding pocket of interest. Firstly, spheres were generated around the surface of the molecule using *sphgen* and *Sphere_selector* was then used to filter results. Only spheres within 10 Å of the PYLKTKFI peptide were selected for further analysis. This resulted in the assessment of the full SH2 domain of the STAT3 molecule (residues 582 – 688) for potential binding of Crispene E.

A further identical study was conducted on the DNA binding domain (residues 321 – 465) using spheres within 10 Å of central residues in the DNA-binding domain (VERQP) in order to ascertain the potential of Crispene E to bind to the DNA binding domain of STAT3.

Finally, Crispene E was automatically positioned into the spheres with maximum number of conformations set at a high level (500) to explore a large amount of conformational space, thus producing a docked ligand:protein structure.

4.1.4 Evaluation of Ligand Binding

The ligand was evaluated based on two DOCK scoring functions, MMGBSA and GRID scoring. During the docking process, a grid was created around the receptor. The grid was then used to allow rapid score evaluation in DOCK. Prior to scoring, orientations of the ligand which exhibited significant steric interactions with the receptor molecule were discarded using the *bump* filter.

The grid-based scoring term is based on non-bonded parameters within the force-field (i.e. van der Waals forces, electrostatic forces). The Hawkins MMGBSA method is an adaptation of the original MMGBSA method and uses the pairwise solvation method (2, 3), and the interaction between ligand and receptor are represented by Lennard Jones and Coulombic potentials, coupled with the change in solvation (Δ GBSA), which in turn is represented by the following equation:

$$\Delta\text{GBSA} = \text{GBSA}_{\text{complex}} - (\text{GBSA}_{\text{receptor}} + \text{GBSA}_{\text{ligand}})$$

In evaluating ligand-protein interactions, two factors were considered crucial; shape-fit of the molecule to the protein receptor, and prevention of the interaction of the STAT3 homodimer B (particularly residues PYLKTKFI) with octapeptide binding site. The latter was considered particularly relevant as modulation of either Y705 (which is phosphorylated before dimerisation) or P704 (i.e. P and Y of **P**YLKTKFI) is known to disrupt dimerisation of the protein. The former was considered relevant as shape-fit is crucially important to the binding of other STAT3-binding molecules that target the SH2 domain of the protein.

Figure S1

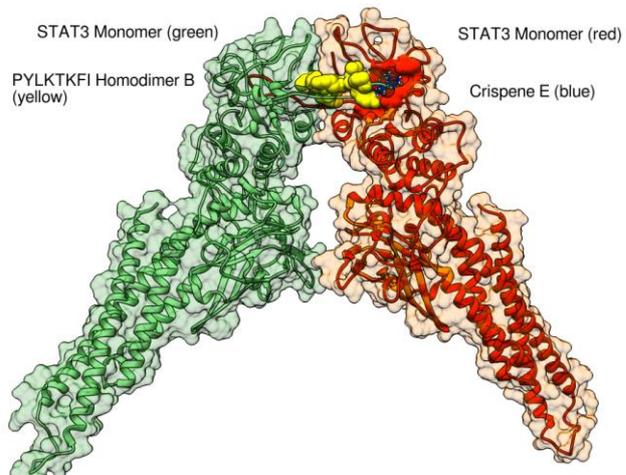


Figure S1: Model of the STAT3 dimer (PDB ID: 1BG1) illustrating the potential inhibition of binding of PYLKTKFI (yellow) of homodimer B (green) of the protein to its binding pocket due to the presence of Crispene E

Figure S2.

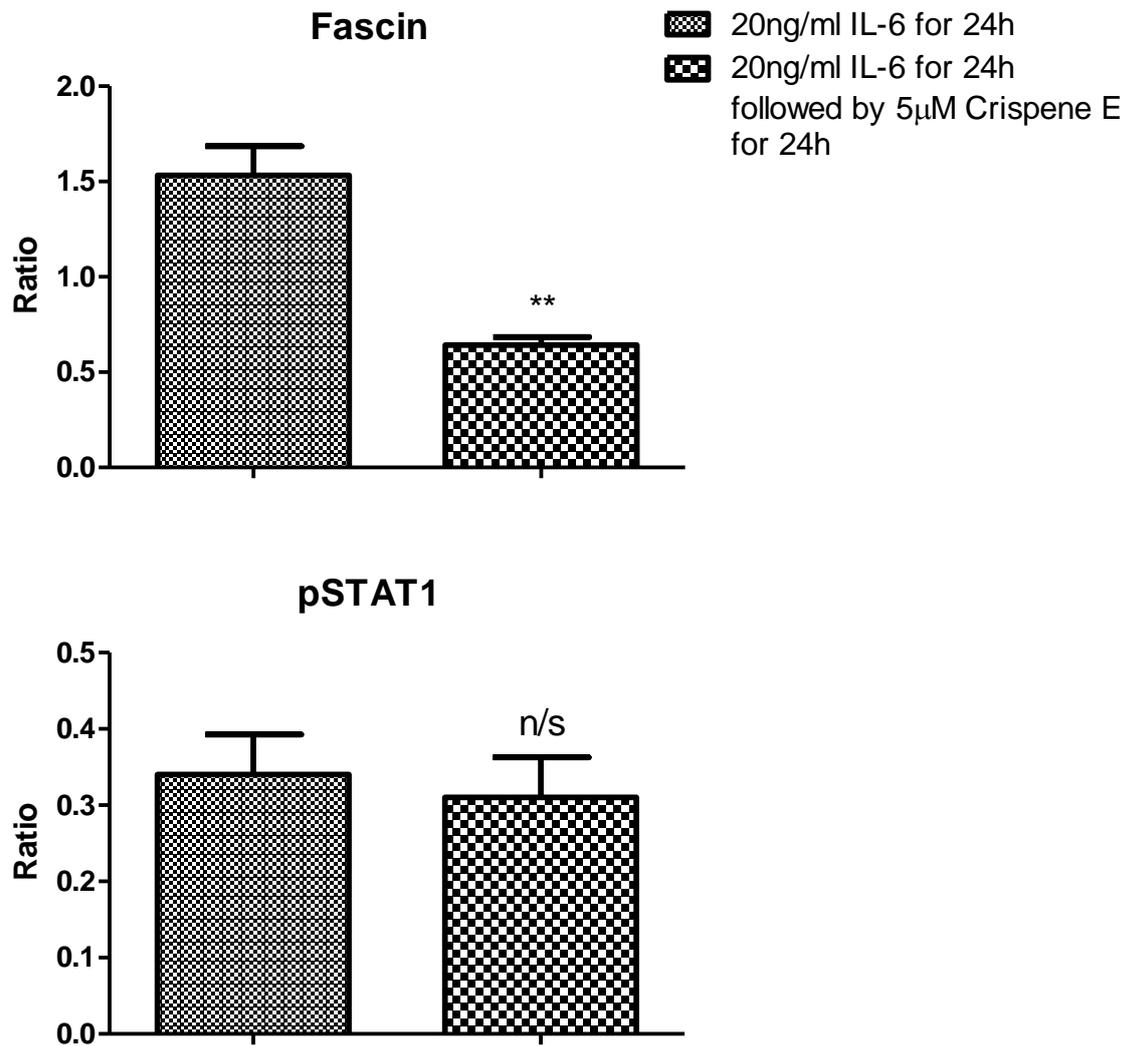


Figure S2: Statistic analysis on western blots of fascin protein expression in MDA-MB-231 showing the significance in fascin (** $P < 0.05$) down-regulation after treatment with 5µM Crispene E for 24 h. In contrast, pSTAT1 expression was insignificantly affected by the STAT3 inhibitor under the same conditions ($P > 0.05$). Error bars represent S.D, N/S = not significant.

5.0 Isolation and Characterization of Crispene E (1)

5.1 General

Column chromatography and vacuum liquid chromatography were performed on Sephadex (LH-20) and silica gel respectively. ^1H NMR, ^{13}C NMR and 2D NMR spectra were acquired at 300 °K using Bruker Advance NMR spectrometer at 400 MHz and 100 MHz. ESI-MS data were collected using a Waters Micromass ZQ instrument coupled to a Waters 2695 HPLC with a Waters 2996 PDA. Waters Micromass ZQ parameters used were: Capillary (kV), 3.38; Cone (V), 35; Extractor (V), 3.0; Source temperature (°C), 100; Desolvation Temperature (°C), 200; Cone flow rate (L/h), 50; Desolvation flow rate (L/h), 250.

5.2 Plant Material

Stems of *T. crispa* were collected from the Tangail district of Bangladesh, in the month of March 2009. The plant was identified by Mr. Sardar Nasir Uddin, Senior Scientific officer, Bangladesh National Herbarium, Dhaka, where a voucher specimen has been deposited (DACB accession number: 35291). The stems were sun dried for several days followed by oven dried for 24 h at 45°C. The dried stems were then ground into coarse powder using high capacity grinding machine in the Phytochemical Research Laboratory, Faculty of Pharmacy, University of Dhaka.

5.3 Extraction and isolation

The powdered stems (1.5 kg) were soaked in methanol (7 L) at room temperature for 14 days with occasional shaking and the extract was collected by filtration. The solvent was evaporated under reduced pressure in a rotary evaporator to obtain a solid residue (25 g). 5 g of the solid residue was subjected to fractionation by using the modified Kupchan partitioning method (VanWagenen *et al* 1993) into *n*-hexane, CCl_4 , CHCl_3 and aqueous soluble fractions. Evaporation of solvent afforded *n*-hexane (400 mg), CCl_4 (1.56 g), CHCl_3 (140 mg) and aqueous soluble fractions. The *n*-hexane soluble fraction was chromatographed over Sephadex (LH-20) and the column was eluted with *n*-hexane: CH_2Cl_2 :MeOH (2:5:1) followed by CH_2Cl_2 :MeOH (9:1) and MeOH (100%) in order to increase the polarities. The column fractions were then concentrated and subjected to TLC screening. The fractions with

satisfactory resolution of compounds were re-chromatographed over silica gel separately to obtain Crispene E (**1**).

5.4 Spectroscopic Characterization of Crispene E (1**)**

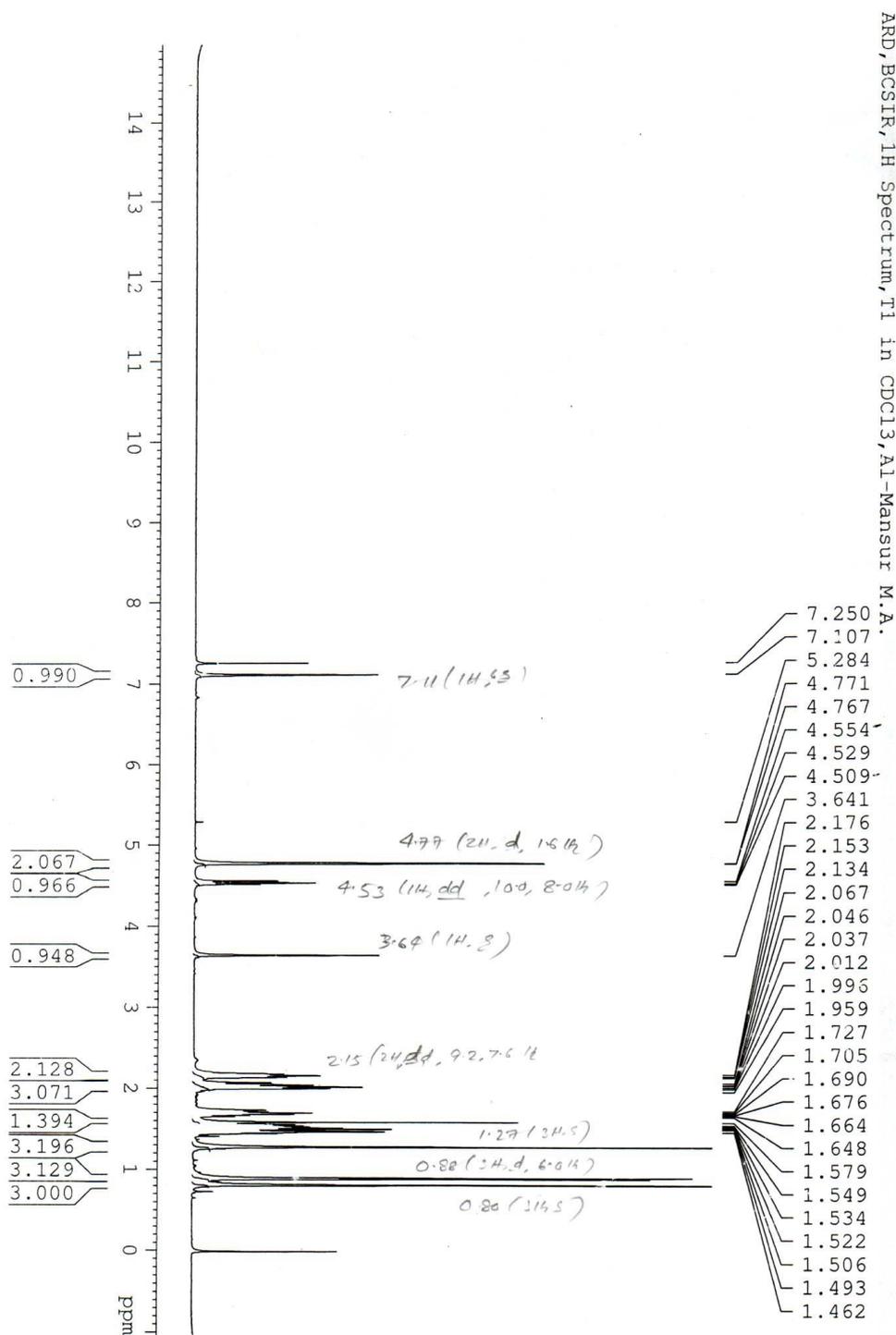
Colourless crystals; Electrospray ionization MS: m/z $[M + Na]^+$ 369.17, $C_{20}H_{26}O_5 + Na$; 1H and ^{13}C NMR: Table S1.

Table S2. ^1H NMR (400 MHz) and ^{13}C NMR (100 MHz) spectroscopic data and HMBC correlations of compound **6** in CDCl_3

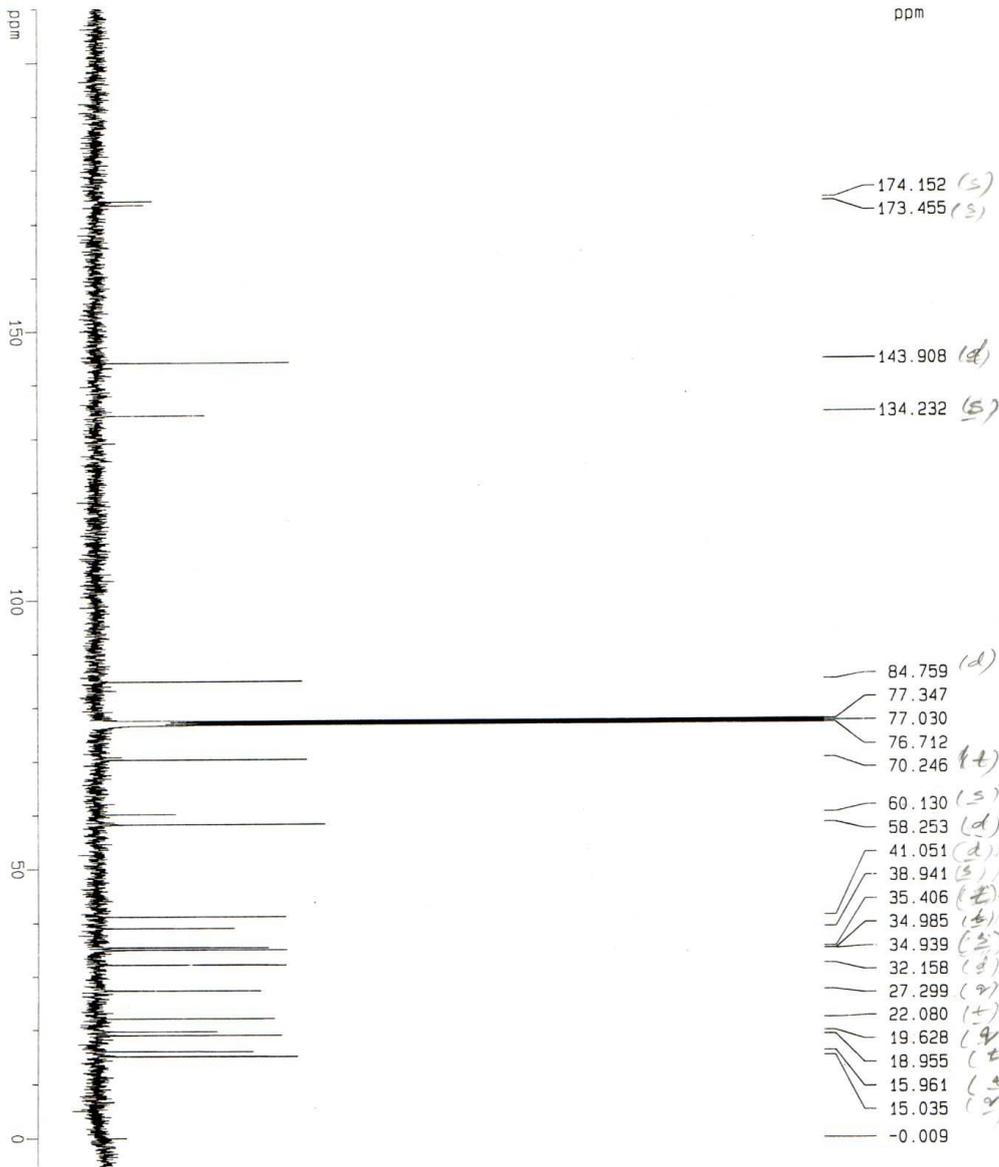
Position	δ_{H} (<i>J</i> in Hz)	δ_{C} (mult.)	HMBC
1	1.69 <i>m</i> *	15.96 <i>t</i>	3, 5, 9
2	2.00 <i>m</i> *	22.08 <i>t</i>	3, 10
3	3.64 <i>s</i>	58.25 <i>d</i>	1, 2, 4
4	-	60.13 <i>s</i>	-
5	-	34.95 <i>s</i>	-
6	4.53 <i>dd</i> (10.0, 8.0)	84.76 <i>d</i>	4, 18, 19
7	2.04 <i>m</i> , 1.43 <i>m</i>	34.99 <i>t</i>	8, 9, 17
8	1.53 <i>m</i>	32.16 <i>d</i>	7, 10, 17, 20
9	-	38.94 <i>s</i>	-
10	1.46 <i>m</i>	41.05 <i>d</i>	1, 2, 4, 5, 6, 9, 19, 20
11	2.00 <i>m</i> *, 1.69 <i>m</i> *	35.41 <i>t</i>	8, 9, 10, 12
12	2.15 <i>dd</i> (9.2, 7.6)	18.99 <i>t</i>	11, 13, 14, 16
13	-	134.23 <i>s</i>	-
14	7.11 <i>bs</i>	143.91 <i>d</i>	12, 13, 15, 16
15	4.77 <i>d</i> (1.6)	70.25 <i>t</i>	13, 15, 16
16	-	174.15 <i>s</i>	-
17	0.88 <i>d</i> (6.0)	15.04 <i>q</i>	7, 8, 9
18	-	173.46 <i>s</i>	-
19	1.27 <i>s</i>	27.30 <i>q</i>	4, 5, 6, 10
20	0.80 <i>s</i>	19.63 <i>q</i>	8, 9, 10, 11

* Signals overlapped in each column

¹HNMR, ¹³CNMR, DEPT, HSQC, HMBC, NOESY and ESI-MS Spectra of Crispene E



ABD, BCSIR, 13C Spectrum, T1 in CDCl3, Al-Mansur M.A.



Current Data Parameters
 NAME ASD72
 EXPNO 2
 PROCNO 1

F2 - Acquisition Parameters
 Date_ 20110420
 Time 12.56
 INSTRUM dpx400
 PROBD 5 mm Multinu
 PULPROG zgpg30
 TD 32768
 SOLVENT CDCl3
 NS 1429
 DS 2

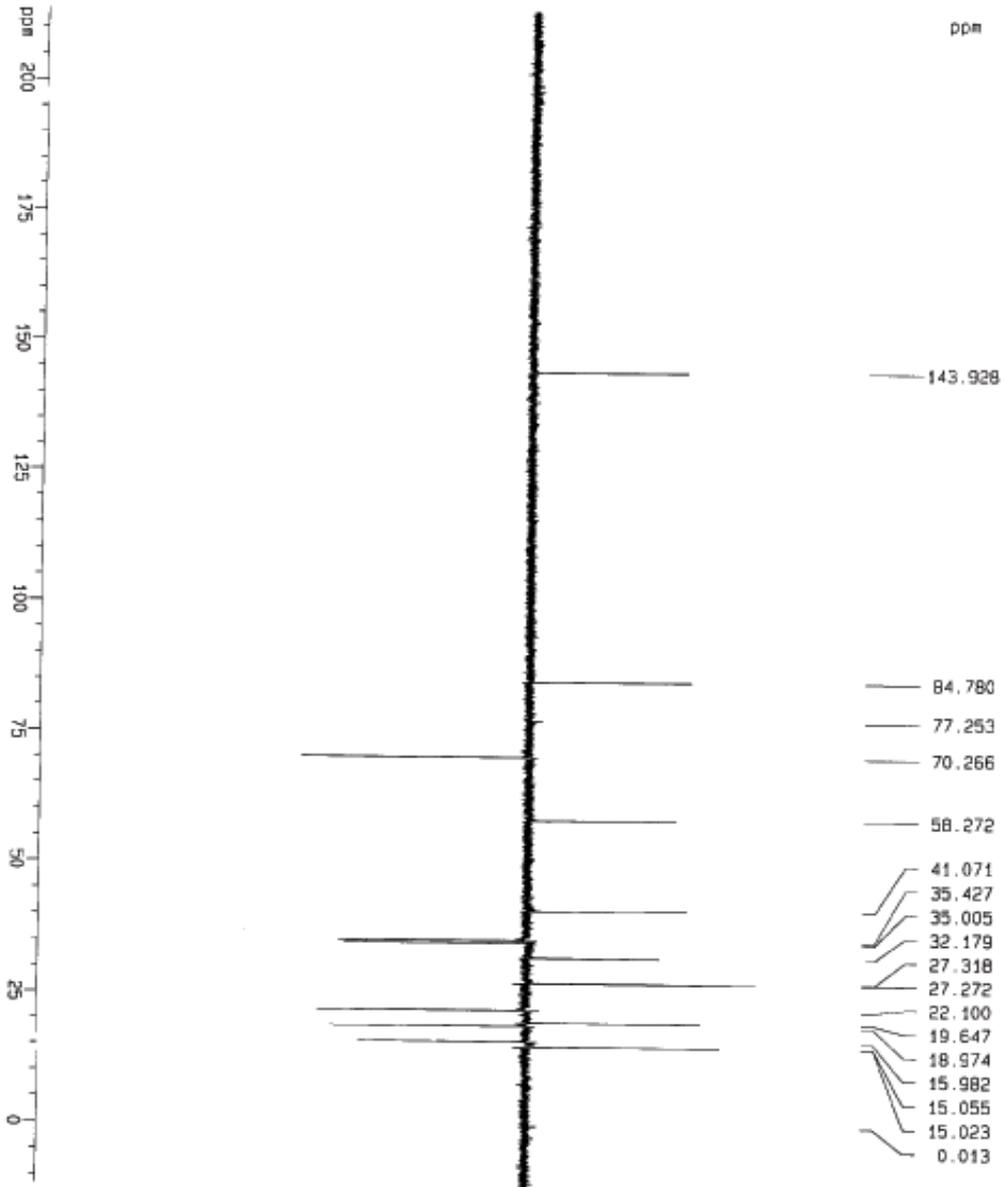
SWH 21645.021 Hz
 FIDRES 0.1680524 Hz
 AQ 0.7369308 sec
 RG 16384
 DW 23.100 usec
 DE 6.00 usec
 TE 300.0 K
 D1 1.50000000 sec
 d11 0.03000000 sec
 d12 0.00020000 sec

CHANNEL f1
 NUC1 13C
 P1 8.30 usec
 PL1 -6.00 dB
 SF01 100.6258980 MHz

CHANNEL f2
 CPUPRG2 MALT216
 NUC2 1H
 PCPRG2 80.00 usec
 PL2 -6.00 dB
 PL12 16.00 dB
 PL13 120.00 dB
 SF02 400.1400000 MHz

F2 - Processing parameters
 SI 32768
 SF 100.6158823 MHz
 MDW EM
 SSB 0
 LB 2.50 Hz
 GB 0
 PC 1.40

1D NMR Plot Parameters
 CX 20.00 cm
 F1P 210.000 ppm
 F1 21129.25 Hz
 F2P -5.126 ppm
 F2 -515.77 Hz
 BPWCK 10.75633 ppm/cm
 HZCK 1082.25110 Hz/cm



```

Current Data Parameters
NAME      A0672
EXPNO    3
PROCNO   1

F2 - Acquisition Parameters
Date_     20110420
Time      15:38
INSTRUM   00x400
PROBHD    5 mm MULTIC
PULPROG   zgpg30
TD         32768
SOLVENT   CDCl3
NS         1426
DS         8
SWH        24154.592 Hz
FIDRES    0.737140 Hz
AQ         0.0783476 sec
RG         13004
DM         20.700 usec
DE         5.00 usec
TE         300.0 K

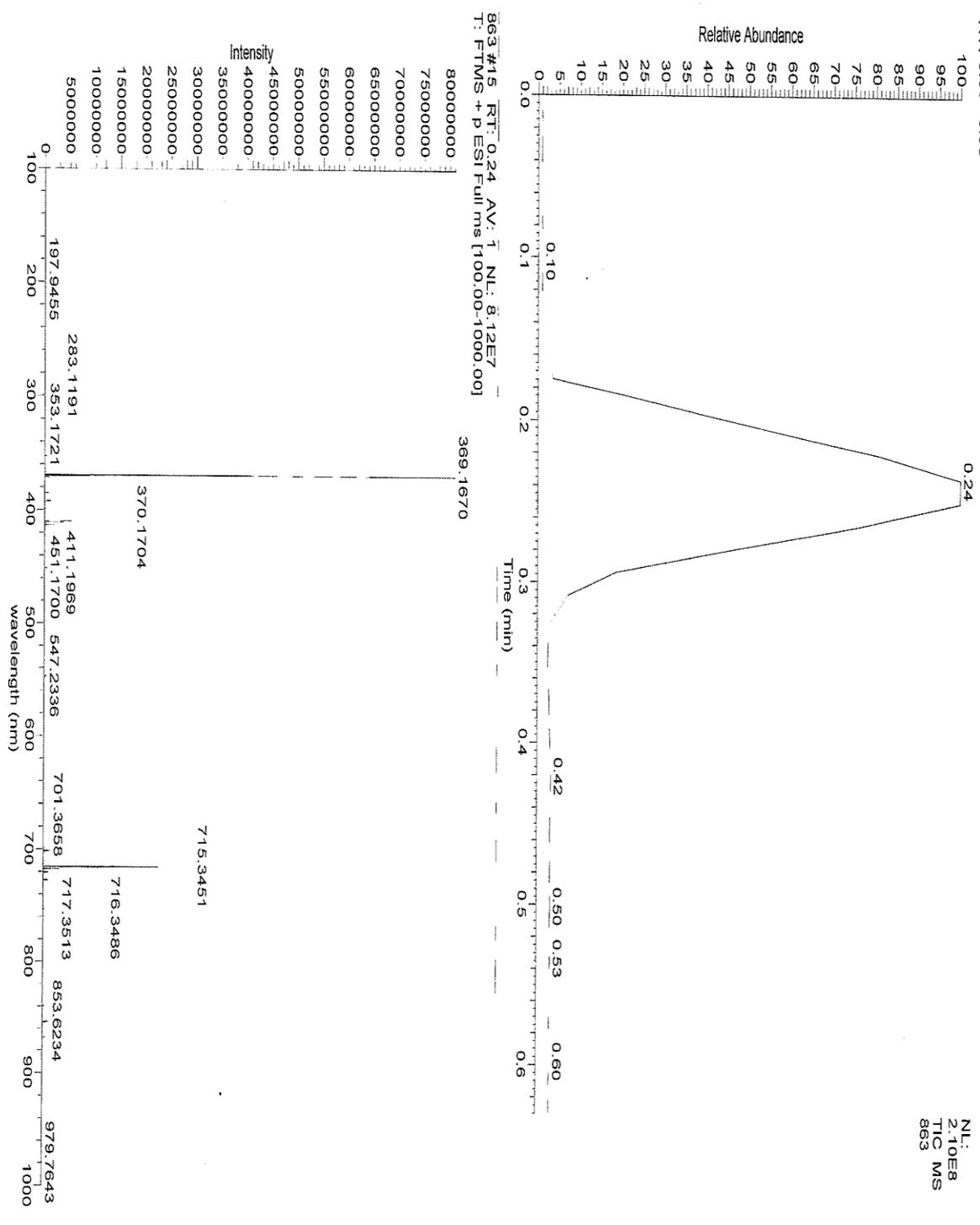
DELTA     140.000000 sec
D1         4.00000000 sec
d2         0.00344828 sec
d32        0.00020500 sec
DELTA     0.00000764 sec

***** CHANNEL f1 *****
NUC1       13C
P1         6.00 usec
P2         12.00 usec
PL1        -5.00 dB
SF01       100.625000 MHz

***** CHANNEL f2 *****
CPROG     zgpg30
NUC2       31P
P3         8.30 usec
P4         15.60 usec
PCPD2     60.00 usec
PL2        -5.00 dB
PL12       15.00 dB
SF02       400.142000 MHz

F2 - Processing parameters
SI         32768
SF         100.6152804 MHz
WDW        EM
SSB        0
LB         1.00 Hz
GB         0
PC         1.48

ID parameters
CX         20.00 cm
FSP        213.331 00M
FE         21464.38 Hz
FPP        -11.797 00M
F2         -1188.01 Hz
FREQON     11.25029 00M/CX
H2ON       1122.26482 Hz/CX
    
```

High resolution mass data - Exact Mass: 369.1678(M+ Na), observed mass 369.1670 (M+ Na)+, Error - 2.16 ppm