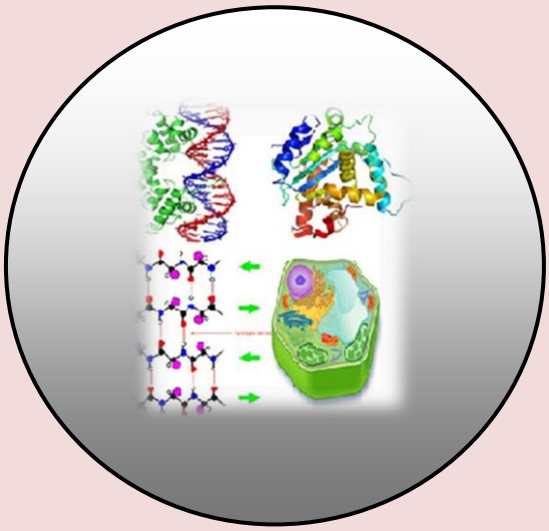
Isolation and Structure Elucidation of a Novel Nonasaccharide 'Tarose' from Bos Primigenius Taurus (Jarsi cow) Colostrum

By

Gunjan, Shivam Yadav, Manisha Shukla and Desh Deepak

ISSN 2319-3077 Online/Electronic
ISSN 0970-4973 Print

Index Copernicus International Value
IC Value of Journal 82.43 Poland, Europe (2016)
Journal Impact Factor: 4.275
Global Impact factor of Journal: 0.876
Scientific Journals Impact Factor: 3.285
InfoBase Impact Factor: 3.66



J. Biol. Chem. Research
Volume 39 (2), 2022 Pages No. 243-263

Journal of Biological and Chemical Research

An International Peer Reviewed / Referred Journal of Life Sciences and Chemistry

Indexed, Abstracted and Cited in various International and National Scientific Databases

Published by Society for Advancement of Sciences®

J. Biol. Chem. Research. Vol. 39, No. 2, 243-263, 2022
(An International Peer Reviewed / Refereed Journal of Life Sciences and Chemistry)
Ms 39/02/131/2022

ISSN 2319-3077 (Online/Electronic)
ISSN 0970-4973 (Print)

Dr. Desh Depak
http:// www.sasjournals.com
http:// www.jbcr.co.in
jbiolchemres@gmail.com

RESEARCH PAPER

Received: 20/11/2022 Revised: 30/12/2022 Accepted: 31/12/2022

Isolation and Structure Elucidation of a Novel Nonasaccharide 'Tarose' from Bos Primigenius Taurus (Jarsi cow) Colostrum

Gunjan, *Shivam Yadav, Manisha Shukla and Desh Deepak

Department of Chemistry, Lucknow University, Lucknow-226007, U.P., India
***Department of chemistry, Shri J.N. M.P.G College (K.K.C.), Lucknow226001, U.P., India**

ABSTRACT

The importance of cow's milk has been well defined in Ayurveda and ancient Indian literature as a panacea for all diseases. It is responsible for the development of the brain and skeletal system, and it is also responsible for the stimulation of the immune system. Cow's milk consists of protein, fats, and carbohydrates, which are in the form of lactose and oligosaccharides. The biological activities of cow's milk depend on the number of oligosaccharides present in it. In search for more novel oligosaccharides, the cow milk was re-investigated, and a novel nonasaccharide (TAROSE) was isolated from Jarsi cow's colostrum, and its structure was elucidated with the help of chemical degradation, chemical transformation, and physico-chemical techniques like ^{1}H, ^{13}C, and 2D NMR experiments (COSY, TOCSY, HSQC, and HMBC) and mass spectrometry, which is as under:

Gal–β–(1→4)
↑
Gal–β–(1→4)–Glc–α–(1→3)–Gal–β–(1→3)–GalNAc–β–(1→3)–GalNAc–β–(1→3)–GlcNAc–β–(1→3)–GalNAc–β–(1→3)Glc

TAROSE

Keywords: Structure Elucidation, Immune system, Biological activities, Chemical transformation and Physico-chemical techniques.

INTRODUCTION

The livelihood of any mammalian neonate depends on the biofluid, i.e., milk, provided to him by the mother. In the case of humans, the importance of mother's milk plays a pivotal role in the development of the infant.

Sometimes, it is observed that the mother is not able to provide the complete diet required by the newborn. In that case the cow's milk is the only replacement which is available with ease and in abundance. Further in ancient medicinal literature, the cow's milk has been defined as penicia, which is responsible for the development of the immune system, brain, and bones of the newly born. The cow's milk contains proteins, fats, glycoproteins, and carbohydrates, and each and every constituent has a definite role in the development of the neonate [Mayank Sharma et al., 2022].

The carbohydrate contents of the cow's milk are in the form of lactose and oligosaccharides. The recent research confirmed that the biological activity of the cow's milk is due to the various oligosaccharides present in the milk [Mayank Sharma et al., 2022]. These oligosaccharides are present in the form of straight or branched chains that contain glucose, galactose, glcNAc, and galNAc in different combinations and glycosidic linkages with α and β configurations at different ring positions of the monosaccharides, resulting in the formation of novel oligosaccharides responsible for varied biological activities. Cow milk has immunomodulating effects and plays a definite role in the development of the brain of the neonate [Mayank Sharma et al., 2022]. Besides the antioxidant and anti-inflammatory activities, it is well defined in ancient literature that it increases lactation in mothers, feeding babies. The colostrum of cows (i.e., early milk) contains more medicinal value, which may be due to a high oligosaccharide concentration. In our previous communications, we have isolated 26 novel oligosaccharides [Mayank Sharma et al., 2022], namely Aurose [Gunjan and Desh Deepak, 2019], Osose [Mayank Sharma et al., 2022], Tosose [Mayank Sharma et al., 2022], Ausose, Tausose, Rusose, Ureose, Taurose, etc. [Mayank Sharma et al., 2022], from different species of cow. In our endeavour to isolate more novel oligosaccharides from cow's milk or colostrum, we have collected 10 litres of Jarsi cow's colostrum and processed it by the method developed by Deepak et al. (the modified method of Kobata and Ginsberg) involving deproteination, filteration, gel filteration, lyophilization, and acetylation followed by silica gel column chromatography for the isolation and purification of chromatographically purified novel oligosaccharide. This whole process led to the isolation of a novel nonasaccharide 'Tarose' into its acetylated form i.e. acetylated Tarose 'b' while the natural oligosaccharide was designated as Tarose B. The structure of this novel nonasaccharide was elucidated with the help of ^{1}H, 13C, and 2D-NMR (COSY, TOCSY, HSQC, and HMBC) experiments, as well asmass spectrometry.

MATERIAL AND METHODS

General Procedure

General procedures were same as described in our previous articles [Desh Deepak et al., 2020]. Isolation of Jarsi Cow Colostrum Milk Oligosaccharides by the Modified Method of Kobata and Ginsburg [Ashish Kumar Singh and Desh Deepak, 2020]. The isolation of cow colostrum milk oligosaccharides was done by the method developed by Desh Deepak (modified method of Kobata and Ginsburg), which was described in our previous communication except for the isolation and structure elucidation of compounds. The isolation of oligosaccharide was processed from 10 litres of cow colustrum milk, and theyield of the oligosaccharide mixture was 315 g.

Acetylation of the Oligosaccharide Mixture

12.0 g of pooled fractions that gave positive results in the phenolsulfuric acid test were acetylated with pyridine (12 ml) and acetic anhydride (12 ml) at 60 °C, and the solution was stirred overnight. The mixture was evaporated under reduced pressure, and the viscous residue was taken in $CHCl_3$ (500 ml) and washed in sequence with 2N-HCl (I x 25 ml), ice- cold 2N-$NaHCO_3$ (2 x 25 ml), and finally with H_2O (2 x 25 ml). The organic layer was dried over anhydrous Na_2SO_4, filtered, and evaporated to dryness, yielding the acetylated mixture (13 gm).

Deacetylation of compound b, Tarose Acetate

Compound **'b'** (48 mg) was obtained from column chromatography 2 of acetylated oligosaccharide mixture. 40 mg of compound b was dissolved in acetone (3 ml), 3.5 ml of NH_4OH was added, and it was left overnight in a stoppered hydrolysis flask. After 24 hours, ammonia was removed under reduced pressure, the compound was washed three times with $CHCl_3$ (5 ml) (to remove acetamide), and the water layer was finally freeze dried giving the natural oligosaccharide **B** (34 mg).

Methyl glycosidation/acid hydrolysis of compound B [Kuldeep Kumar and Desh Deepak et al. (2022)]

Compound **B** (8 mg) was refluxed with absolute MeOH (2 ml) at 70 °C for 18 h in the presence of cation exchange IR-l20 (H) resin. The reaction mixture was filtered while hot, and the filtrate was concentrated. To this reaction mixture of methylglycoside B, 1,4-dioxane (1 ml), and 0.1N H_2S0_4 (1 ml), was added, and the solution was warmed for 30 minutes at 50°C. The hydrolysis was complete after 24 hrs. (TLC) The hydrolysate was neutralized with freshly prepared $BaCO_3$ filtered and concentrated under reduced pressure to afford α-and β- methylglucosides along with the Glc, Gal, GalNAc and GlcNAc. Their identification of compound B was confirmed by comparison with authentic samples (TLC, PC) of α-and β- methylglucosides along with the Glc, Gal, GalNAc and GlcNAc.

Killiani hydrolysis of compound B [Kuldeep Kumar and Desh Deepak et al. (2022)] Compound **B** (5 mg) was dissolved in a 2 ml Killiani mixture (AcOH-H_2O-HCl, 7:11:2) and heated at 100 °C for 1 h, followed by evaporation under reduced pressure. It was dissolved in 2 ml of H_2O and extracted twice with 3 ml of $CHCl_3$. The aqueous residual solution was made neutral by the addition of 1-2 drops of 2N NaOH and was evaporated under reduced pressure to afford Glc, Gal, GalNAc, and GlcNAc on comparison with authentic samples of Glc, Gal, GalNAc, and GlcNAc.

Description of Compound Tarose

Compound B, TAROSE

Compound **b** (48 mg) was obtained from fraction 109-114 of column chromatography 2. On deacetylation of 40 mg of substance b with NH_4OH/acetone, it afforded substance **B** (34 mg) $[\alpha]_D^{25}$ =+34º (*c* 1% H_2O). For experimental analysis, this compound was dried over P_2O_5 at 100°C and 0.1 mm pressure for 8 hr. It gave positive Phenol-sulphuric acid test, Feigl test and Morgan-Elson test.

$C_{62}H_{104}O_{46}N_4$		%C	%H	%N
	Calculated	45.36	6.34	3.41
	Found	45.37	6.34	3.41

^{1}H NMR of Compound-b, Tarose Acetate in $CDCl_3$ at 300 MHz.

δ6.09[d, 1H, J=3.9Hz, *α*-Glc(S-1) H-1], δ5.29[d, 1H, J=8.1Hz, *β*-Glc(S-1) H-1], δ5.28[d, 1H, J=3.9Hz, α-Glc(S-8) H-1], δ4.67[d, 1H, J=7.5Hz, *β*-GlcNAc(S-4) H-1], δ4.59[d, 1H, J=8.1Hz, *β*-Gal(S-7) H-1], δ4.50[d, 1H, J=8.1Hz, *β*-Gal(S-2) H-1], δ4.49[d, 1H, J=8.1Hz, *β*- Gal(S-9) H-1], δ4.43[d, 2H, J=7.5Hz, *β*-GalNAc(S-3) & (S-5) H-1], δ4.41[d, 1H, J=7.8Hz, *β*- GalNAc(S-6) H-1], δ3.75[m, *β*-GalNAc(S-2, S-3, S-4, S-6) & *β*-Glc(S-1) H-3], δ3.73[m, 2H, *α*-Gal(S-7) & *α*-Glc(S-8) H-3], δ3.49[m, 1H, *β*-Glc (S-1) H-4].

^{13}C NMR of Compound-b, Tarose Acetate in $CDCl_3$ at 300 MHz.

δ91.5[1C, *α*-Glc(S-1) C-1], δ90.43[1C, *β*-Glc(S-1) C-1], δ90.30[1C, *α*-Glc(S-8) C-1], δ95.45[1C, *β*-GlcNAc(S-4) C-1], δ95.53[1C, *β*-Gal(S-7) C-1], δ101.17[2C, *β*-GalNAc(S-3) & (S-5) C-1], δ101.26[1C, *β*-GalNAc(S-6) C-1], δ102.08[1C, *β*-Gal(S-2) C-1], δ102.16[1C, *β*-Gal(S-9) C-1].

^{1}H NMR of Compound-B, Tarose in D_2O at 300 MHz.

δ5.21[d, 2H, J=3.6Hz, *α*-Glc(S-1) & (S-8) H-1], δ4.66[d, 2H, J=7.8Hz, *β*-Glc(S-1) & *β*- Gal(S-7) H-1], δ4.54[d, 1H, J=8.1Hz, *β*-GlcNAc(S-4) H-1], δ4.51[d, 2H, J=8.1Hz, *β*-Gal(S- 2) & (S-9) H-1], δ4.44[d, 3H, J=7.5Hz, *β*-GalNAc(S-3) & (S-5) & (S-6) H-1], δ3.27[t, *β*- Glc(S-1), H-2], δ2.08[s, 3H, $NHCOCH_3$ *β*-GlcNAc(S-4)], δ1.98[s, 6H, $NHCOCH_3$ *β*- GalNAc(S-3) & (S-5)], δ1.90[s, 3H, $NHCOCH_3$ *β*-GalNAc(S-6)].

ES Mass

1640 [M^+], 1608, 1581, 1539, 1489, 1478, 1447, 1446, 1419, 1316, 1273, 1254, 1154, 1125, 1083, 1053, 951, 919, 875, 841, 824, 748, 706,705, 677, 646, 645, 545, 481, 467, 466, 465, 406, 342, 324, 180.

RESULTS AND DISCUSSION

Compound B, TAROSE

Compound B, Tarose, $C_{62}H_{104}O_{46}N_4$ $[\alpha]_{25}$ = +34º gave positive Phenol-sulphuric acid test [Dubois M. and Smith F., 1956], Fiegl test [Fiegel F., et al., 1954)] and Morgan-Elson test [Fiegel F., et al., 1954] showing the presence of normal and amino sugars moietie(s) in the compound B. The HSQC spectrum of acetylated Tarose showed the presence of nine cross peaks of ten anomeric protons and carbons in their respective region at δ91.5x6.09, δ90.30x5.28, δ90.43x5.29, δ95.45x4.67, δ95.53x4.59, δ101.17x4.43, δ101.26x4.41, δ102.08x4.50 and δ102.16x4.49, suggested the presence of ten anomeric protons and carbons in compound B. The reducing nature of Tarose on its methylglycosylation followed by its acid hydrolysis MeOH/H^+, led to the isolation of *α* and *β*-methylglucosides along with Glc, Gal, GlcNAc and

GalNAc suggested that compound was a reducing nonasaccharide in which glucose was present at its reducing end, for convenience all nine monosaccharides were denoted as S-1, S-2, S-3, S-4, S-5, S-6, S-7, S-8 and S-9. The monosaccharides constituents in Tarose were also confirmed by its Killiani hydrolysis under strong acidic condition, followed by its paper chromatography and TLC. In this hydrolysis four spots were found identical with the authentic samples of Glc, Gal, GlcNAc and GalNAc by co-chromatography. Thus Killiani hydrolysis confirmed that Tarose contained four types of monosaccharides units i.e. Glc, Gal, GlcNAc and GalNAc. Further the presence of nine doublets for ten anomeric protons at δ6.09(1H), δ5.29(1H), δ5.28(1H), δ4.67(1H), δ4.59(1H), δ4.50(1H), δ4.49(1H), δ4.43(2H) and δ4.41(1H) in the acetylated spectrum of Tarose in $CDCl_3$ at 300 MHz confirmed the presence of a nonasaccharide in its reducing form. Further the presence of nine anomeric carbon signals for ten anomeric carbon at δ90.30 (1C), δ90.43(1C), δ91.5(1C), δ95.45(1C), δ95.53(1C), δ101.17(2C), δ101.26(1C), δ102.08(1C) and δ102.16(1C) in the ^{13}C NMR spectrum of acetylated Tarose in $CDCl_3$ at 300 MHz. confirmed the presence of a nonasaccharide in its reducing form. The nonasaccharide nature of Tarose was further supported by the presence of five anomeric proton doublets for ten anomeric protons at δ5.21(2H), δ4.66(2H), δ4.54(1H), δ4.51(2H) and δ4.44 (3H) in ^1H NMR spectrum of Tarose in D_2O at 300 MHz. The ^1H and ^{13}C NMR spectra of Tarose justify the ten anomeric signals for nonasaccharide with total integral intensity of nine anomeric proton/carbons. The mass ion peak at 1640$[M]^+$ present in ES- MS of Tarose was in agreement with molecular formula $C_{62}H_{104}O_{46}N_4$.

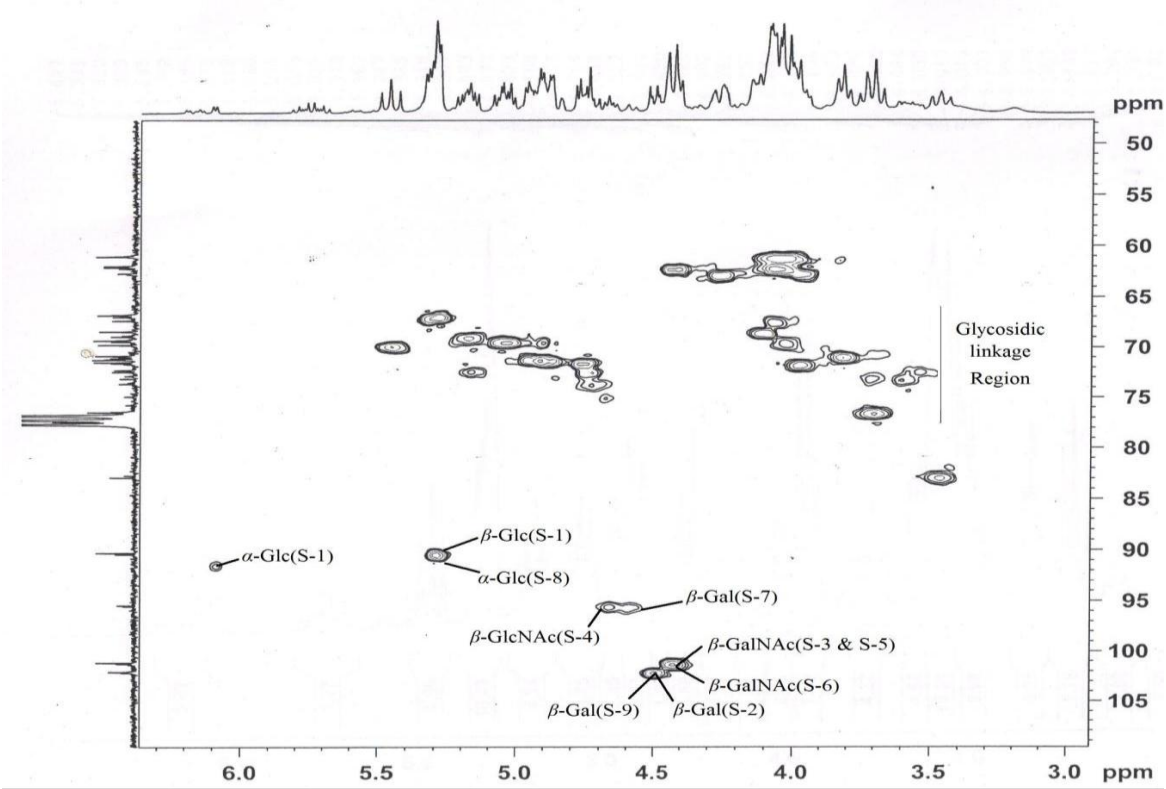


Figure: HSQC Spectrum of Tarose acetate in $CDCl_3$ at 300 MHz.

Table 1. ^{1}H NMR values of Compound B Tarose in D_2O and $CDCl_3$ at 300MHz.

Moieties	In D_2O		In $CDCl_3$	
	^{1}H NMR(δ)	Coupling constant(J)	^{1}H NMR(δ)	Coupling constant(J)
α-Glc(S-1)	5.21	3.6Hz	6.09	3.9Hz
β-Glc(S-1)	4.66	7.8Hz	5.29	8.1Hz
β-Gal(S-2)	4.51	8.1Hz	4.50	8.1Hz
β-GalNAc(S-3)	4.44	7.5Hz	4.43	7.5Hz
β-GlcNAc(S-4)	4.54	8.1Hz	4.67	7.5Hz
β-GalNAc(S-5)	4.44	7.5Hz	4.43	7.5Hz
β-GalNAc(S-6)	4.44	7.5Hz	4.41	7.8Hz
β-Gal(S-7)	4.66	7.8Hz	4.59	8.1Hz
α-Glc(S-8)	5.21	3.6Hz	5.28	3.9Hz
β -Gal(S-9)	4.51	8.1Hz	4.49	8.1Hz

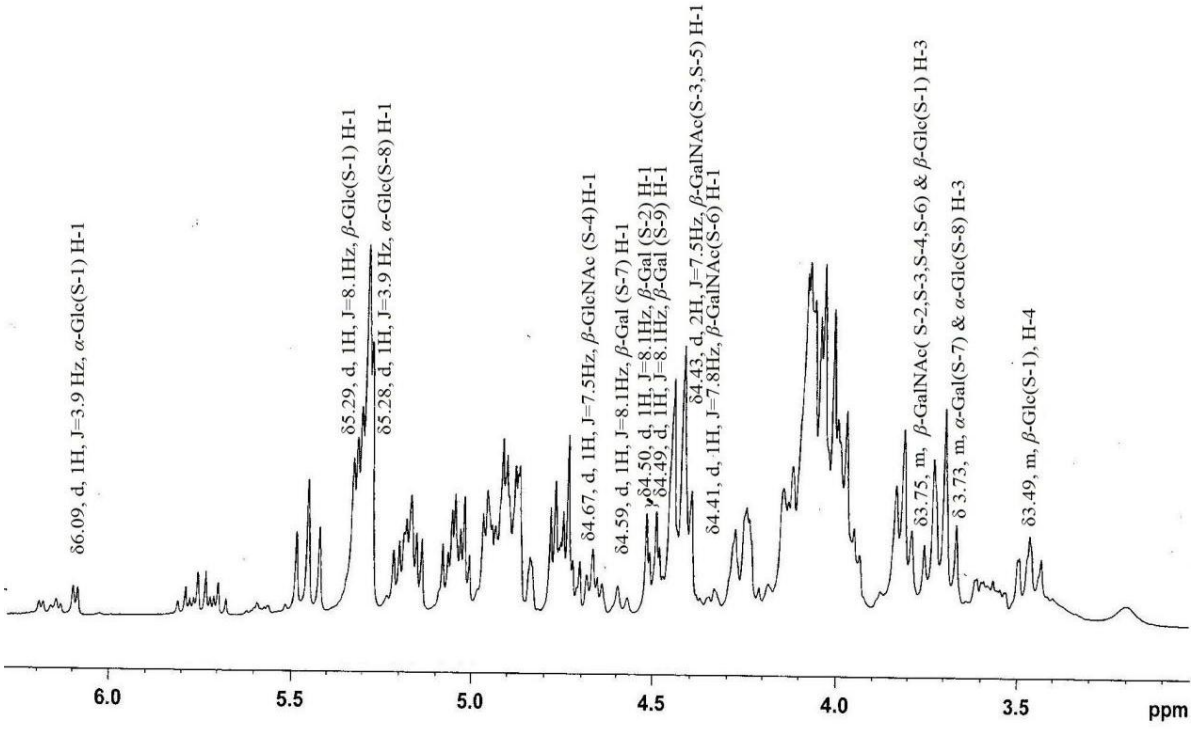


Figure: ^{1}H NMR Spectrum of Tarose acetate in $CDCl_3$ at 300 MHz.

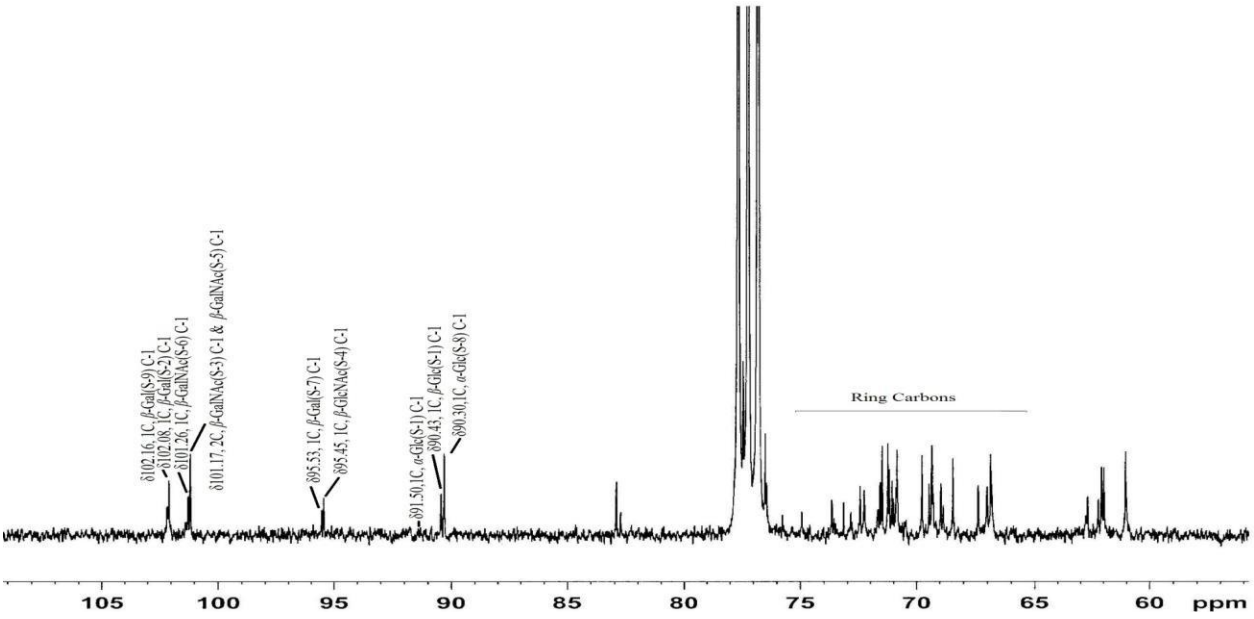
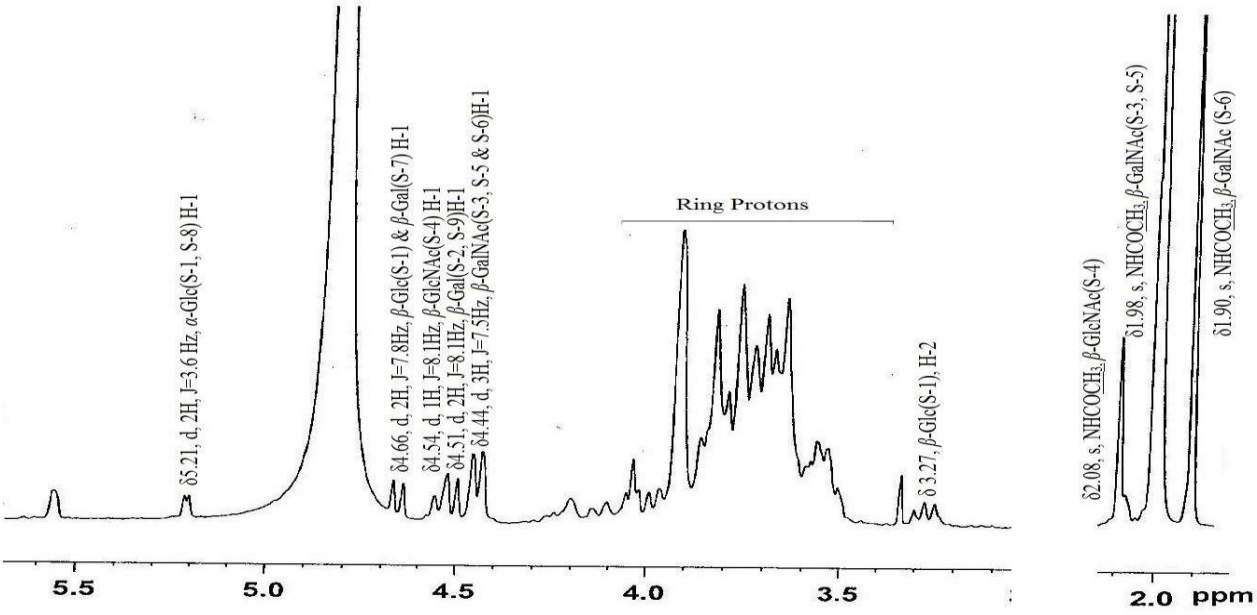
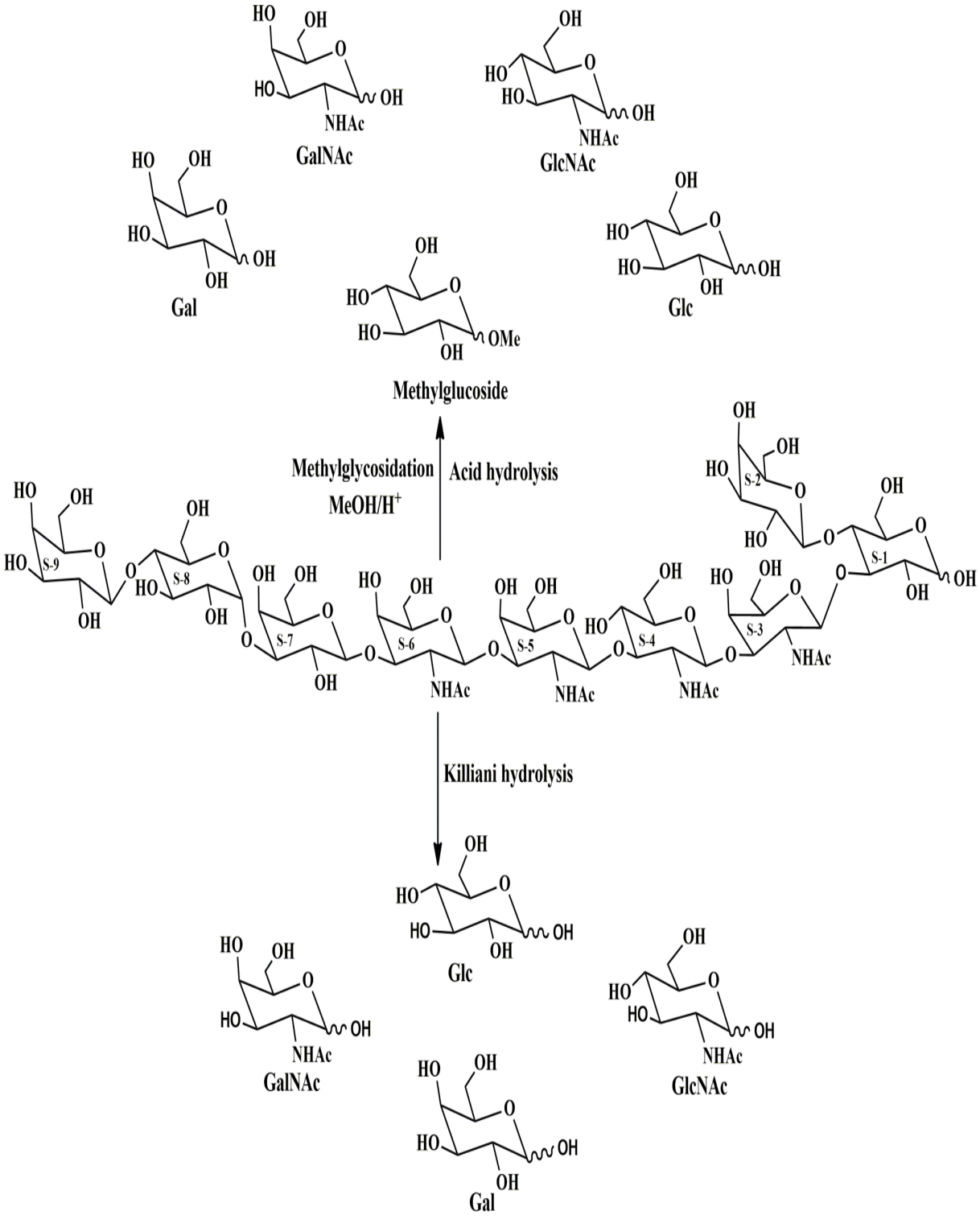


Figure: ^{13}C NMR Spectrum of Tarose acetate in $CDCl_3$ at 300 MHz.

Figure: ^{1}H NMR Spectrum of Tarose in D_2O at 300 MHz.



The ^{1}H NMR spectrum of Tarose in D_2O at 300 MHz contain two doublets at δ5.21 (J=3.6Hz) and δ4.66 (J=7.8Hz) confirmed the presence of glucose at the reducing end in the nonasaccharide [Dua, V.K. and Bush, C.A. (1983)].



Scheme 2.1: Methylglycosidation/Acid Hydrolysis and Killiani Hydrolysis of Tarose.

Further the presence of another anomeric doublet at δ4.51 (J=8.1 Hz) in ^{1}HNMR spectrum of D_2O suggested the presence of *β*-Gal (S- 2) residue as the next monosaccharide unit. In addition to anomeric signals of Glc and Gal, presence of a triplet at δ3.27 which was due to H-2 of *β*-Glc (S-1) suggested the presence of Lactose type of structure i.e. *β*-Gal(1-4)→Glc (structure reporter group)[Prasoon Chaturvedi and Chandra B. Sharma.,(1988)] at the reducing end of Tarose. Simultaneously ^{1}H NMR and ^{13}C NMR spectrum of Tarose acetate also showed downfield shifted *α* and *β* anomeric protonand carbon of reducing monosaccharide (S-1) i.e. Glc (S-1) at δ6.09 (J=3.9Hz), δ5.29 (J=8.1Hz) and δ91.5, δ90.43 respectively. The anomeric protons signal present at δ5.29 in TOCSY Spectrum of Tarose acetate assigned to *β*-Glc (S-1) gave three cross peaks at δ5.29x3.49, δ5.29x3.75 and δ5.29x4.75, which was later identified as H-4, H-3 and H-2 of reducing Glc respectively by COSY spectrum of Tarose acetate. The chemical shift of the cross peak at δ5.29x3.49 and δ5.29x3.75 suggested that in sugar S-1, two positions were available for glycosidic linkage by next monosaccharide units. The earlier suggested (1→4) linkage between *β*-Glc (S-1) and *β*-Gal (S-2) was further confirmed by HMBC spectrum of Tarose acetate at 300 MHz which contain the cross peak signal of H-4 of *β*-Glc (S-1) and anomeric carbon of next monosaccharide(S-2) i.e. *β*-Gal (S-2) at δ3.49x102.08.

The anomeric carbon of *β*-Gal (S-2) at δ102.08 gave its complimentary anomeric proton signal at δ4.50 in the HSQC spectrum of Tarose acetate. The chemical shift value of anomeric carbon and anomeric proton were having resemblance with literature value of anomeric chemical shift value of *β*-Gal (S-2) hence S-2 monosaccharide was confirmed as *β*- Gal (S-2). The anomeric proton signal present at δ4.51 for *β*-Gal (S-2) had a J value of 8.1Hz confirmed the *β* glycosidic linkage between S-2→S-1. The (1→4) linkage between *β*-Glc (S- 1) and *β*-Gal (S-2) was supported by the presence of H-4 signal of S-1 at δ3.49 in upfield region of ^{1}H NMR spectrum of Tarose acetate which was confirmed by the TOCSY and COSY spectrum. Further the (1→4) linkage between S-2 and S-1 was supported by the presence of cross peak at δ3.49x82.73 in glycosidic region of HSQC spectrum of Tarose acetate in $CDCl_3$. The anomeric proton signal at δ4.50 in ^{1}H NMR spectrum of Tarose acetate in $CDCl_3$ assigned for *β*-Gal (S-2) gave three cross peaks at δ4.50x4.97, δ4.50x5.15 and δ4.50x5.28 in its TOCSY spectrum since this anomeric proton does not showed anycross peak in the linkage region, hence confirmed that *β*-Gal (S-2) was present at non- reducing end and none of its -OH group were involved in glycosidic linkage. Since it was ascertained by the COSY and TOCSY spectrum of Tarose acetate that the *β*-Glc (S-1) has two vacant position i.e. H-3 and H-4 and it was already confirmed that H-4 of S-1 was linkedwith *β*-Gal (S-2) whereas the left over H-3 position of *β*-Glc (S-1) at δ3.75 showed a long range coupling with anomeric carbon i.e. C-1 of next monosaccharide (S-3) at δ101.17 in HMBC spectrum of Tarose acetate in $CDCl_3$ at 300 MHz. The anomeric carbon signal at δ101.17 showed its complimentary signal at δ4.43 in HSQC spectrum of Tarose acetate in $CDCl_3$ at 300 MHz. The chemical shift value of anomeric carbon and anomeric proton were having resemblance with literature value of anomeric chemical shift value of *β*-GalNAc (S-3)hence S-3 monosaccharide was confirmed as *β*-GalNAc (S-3). Further the presence of *β*- GalNAc (S-3) as next monosaccharide in Tarose was supported by appearance of anomeric proton signal at δ4.44 (J=7.5 Hz) along with a singlet of three proton of amide methyl at δ1.98 in ^{1}H NMR

spectrum of Tarose in D_2O at 300 MHz[Uemura Y., Asakuma S. et al.(2005)] The coupling constant of anomeric signal *β*-GalNAc (S-3) with larger value of 7.5 Hz showed that *β* configuration of the *β*-GalNAc (S-3). The (1→3) linkage between *β*-Glc (S-1) and *β*-GalNAc (S-3) was supported by the presence of H-3 signal of S-1 at δ3.75 in upfield region of ^1H NMR spectrum of Tarose acetate which was confirmed by the TOCSY and COSY spectrum. Further the (1→3) linkage between S-3 and S-1 was supported by the presence of cross peak at δ3.75x76.48 in glycosidic region of HSQC spectrum of Tarose acetate in $CDCl_3$.The anomeric proton signal at δ4.43 ^1H NMR spectrum of Tarose acetate assigned for *β*-GalNAc (S-3) gave four cross peaks at δ4.43x3.75, δ4.43x4.12, δ4.43x5.05 and δ4.43x5.30 in its TOCSY spectrum in $CDCl_3$ at 300 MHz, out of which proton signal arised at δ4.12 corresponded to H-2 position of *β*-GalNAc (S-3) and another proton signal arised at δ3.75 represented the linkage methine proton of *β*-GalNAc (S-3) which was later identified as H-3 of *β*-GalNAc (S-3) by COSY spectrum of Tarose acetate which was available for (1→3) glycosidic linkages by the next monosaccharide unit (S-4). The next anomeric proton signal which appeared as doublet at δ4.67 in ^1H NMR spectrum of Tarose acetate in $CDCl_3$ at 300 MHz. was due to the presence of *β*-GlcNAc (S-4) moiety [Singh M. and Deepak D. et al., 2016]. This anomeric proton signal present at δ4.67 gave its complimentary signal at δ95.45 in its HSQC spectrum in $CDCl_3$ at 300 MHz. The chemical shift value of anomeric carbon and anomeric proton were having resemblance with literature value of anomeric chemical shift value of *β*-GlcNAc (S-4) hence S-4 monosaccharide was confirmed as *β*-GlcNAc (S-4). Further the presence of *β*-GlcNAc (S-4) as next monosaccharide in Tarose was supported by appearance of anomeric proton signal at δ4.54 along with a singlet of three protons of amide methyl at δ2.08 in ^1H NMR spectrum of Tarose in D_2O [Singh K. A. and Deepak D. et al., 2015]. Since it was ascertained by COSY and TOCSY spectrum of Tarose Acetate that H-3 of S-3 was available for glycosidic linkage by the next monosaccharide unit (S-4), hence *β*-GlcNAc (S-4) must be linked to H-3 of *β*- GalNAc (S-3). Further the (1→3) linkage between S-4 and S-3 was supported by the presence of cross peak at δ3.75x76.48 in glycosidic region of HSQC spectrum of Tarose acetate in $CDCl_3$. The (1→3) linkage between *β*-GlcNAc (S-4) and *β*-GalNAc (S-3) was further supported by the ^1H NMR spectrum of acetylated Tarose in which the signal for H-3 of □-GalNAc (S-3) appeared at δ3.75 which was later confirmed by COSY, TOCSY and HSQC spectrum of acetylated Tarose at 300 MHz in $CDCl_3$. The coupling constant of anomeric signal (S-4) with J value 8.1Hz confirmed the *β*-glycosidic linkage between S-4 and S-3 moiety. The anomeric protons signal present at δ4.67 assigned for *β*-GlcNAc (S-4) in the ^1H NMR spectrum of Tarose acetate contains three cross peaks at δ4.67x3.71, δ4.67x4.73 and δ4.67x5.18 in its TOCSY Spectrum in $CDCl_3$ at 300MHz. The chemical shift of the cross peak at δ4.67x3.71 suggested that in sugar S-4 only one position was available for glycosidic linkage by the next monosaccharide unit. Further the double doublet present at δ3.71 in ^1H NMR spectrum of Tarose acetate was identified as H-3 of *β*-GlcNAc (S-4) by COSY spectrum of Tarose acetate confirming that the H-3 of *β*-GlcNAc (S-4) was available for glycosidation by the next monosaccharide moiety (S-5). Further, H-3 of *β*-GlcNAc (S-4) gave a long range coupling with anomeric carbon of next monosaccharide (S-5) at δ3.71x101.17 in HMBC confirmed the (1→3) linkage between S-5 and S-4.

The anomeric carbon signal present at δ101.17 had its complimentary signal at δ4.43 in HSQC spectrum of Tarose acetate in $CDCl_3$ at 300 MHz. The chemical shift value of anomeric carbon and anomeric proton were having resemblance with literature value of anomeric chemical shift value of *β*-GalNAc (S-5) hence S-5 monosaccharide was confirmed as *β*-GalNAc (S-5). Further the presence of *β*-GalNAc (S-5) as next monosaccharide in Tarose was supported by appearance of anomeric proton signal at δ4.44 (J=7.5 Hz) along with a singlet of three proton of amide methyl at δ1.98 in ^1H NMR spectrum of Tarose in D_2O at 300 MHz. The coupling constant of anomeric signal *β*-GalNAc (S-5) with larger value of 7.5 Hz showed the *β* configuration of the *β*-GalNAc (S-5). The (1→3) linkage between S-5 and S-4 was supported by the presence of glycosidic linkage at δ3.71x71.48 in glycosidic region of HSQC spectrum of Tarose acetate in $CDCl_3$. The (1→3) linkage between *β*-GalNAc (S-5) and *β*-GlcNAc (S- 4) was further supported by the ^1H NMR spectrum of acetylated Tarose in which the signal for H-3 of *β*-GlcNAc (S-4) appeared at δ3.71 which was later confirmed by COSY, TOCSY and HSQC spectrum of Tarose acetate.

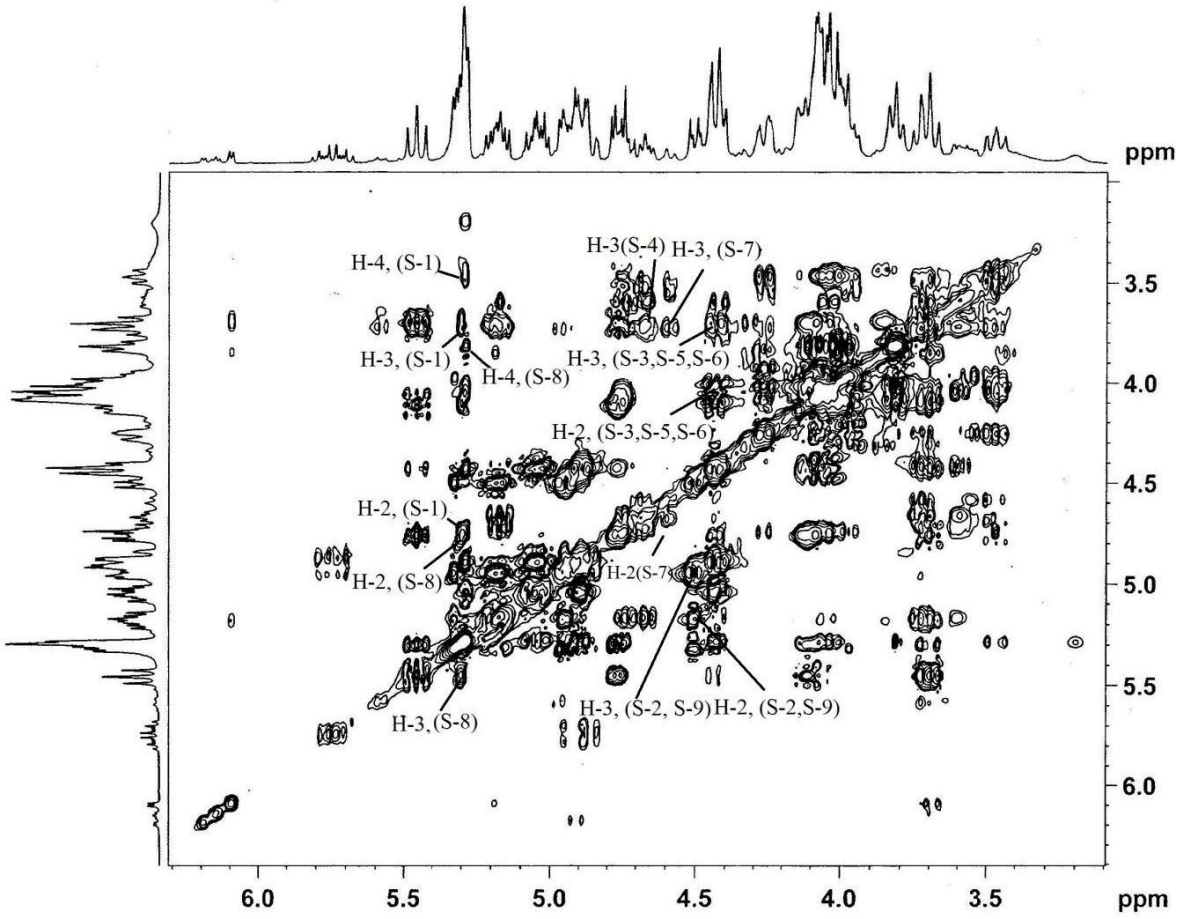


Figure: TOCSY Spectrum of Tarose acetate in $CDCl_3$ at 300 MHz.

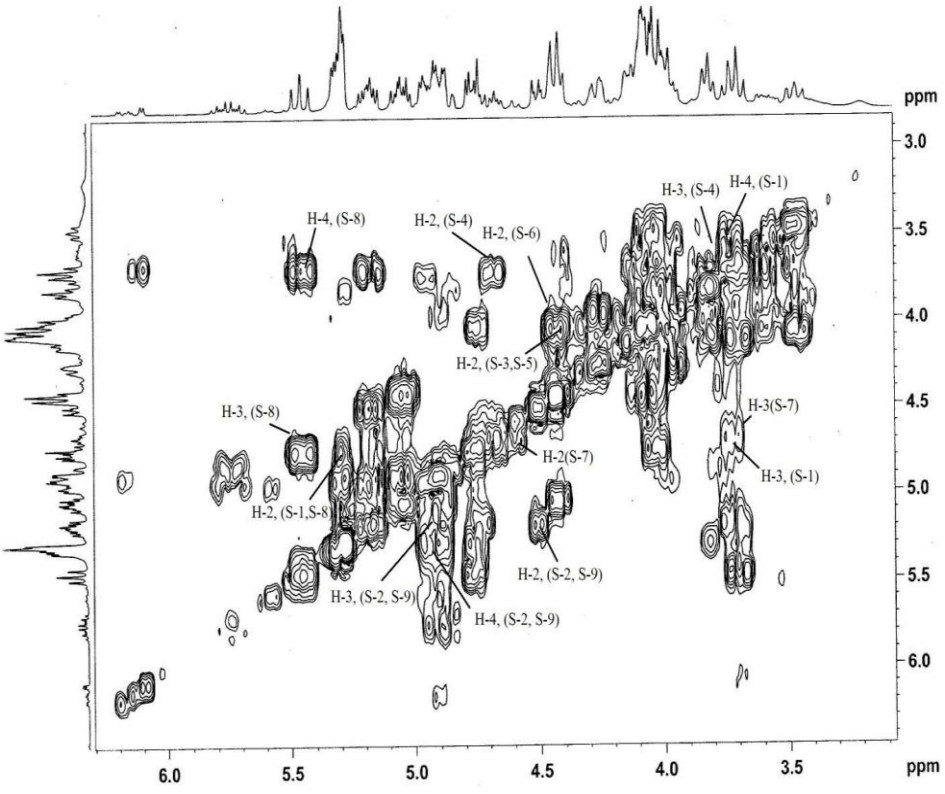


Figure: COSY Spectrum of Tarose acetate in $CDCl_3$ at 300 MHz.

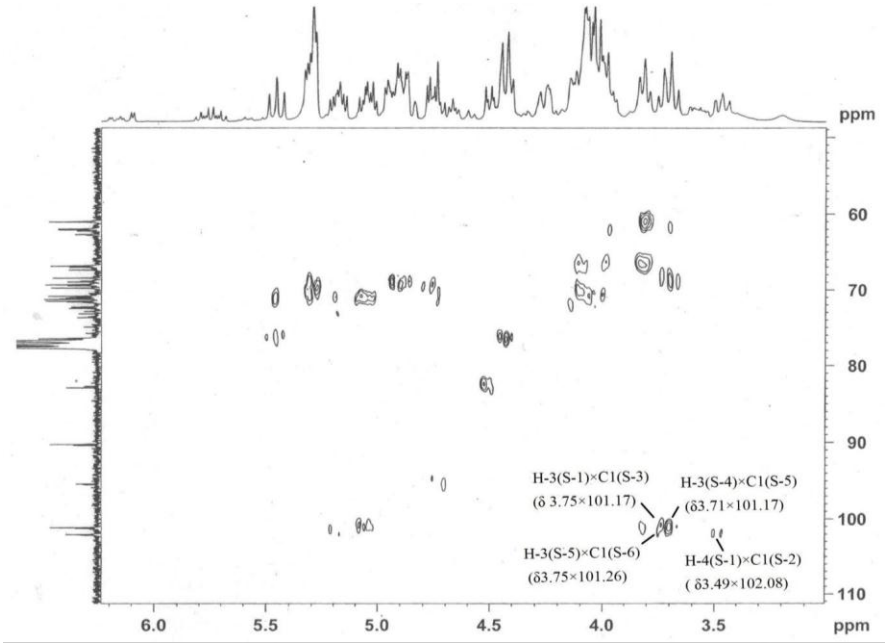


Figure: HMBC Spectrum of Tarose acetate in $CDCl_3$ at 300 MHz.

The anomeric proton signal present at δ4.43 in 1H NMR spectrum of Tarose acetate in CDCl3 at 300 MHz. assigned for β-GalNAc (S-5), contain four cross peaks at δ4.43x3.75, δ4.43x4.12, δ4.43x5.05 and δ4.43x5.35 in its TOCSY spectrum, out of which one cross peak arised at δ4.43x4.12 corresponded to H-2 position of β-GalNAc (S-5) and another signal arised at δ4.43x3.75 represented the linkage region of β- GalNAc (S-5) which was later defined as H-3 of β-GalNAc (S-5) by COSY spectrum of Tarose acetate which was available for (1→3) glycosidic linkage by the next monosaccharide unit (S-6). Further, HMBC spectrum of Tarose acetate showed long range coupling between H-3 of β-GalNAc (S-5) and anomeric carbon of next monosaccharide (S-6) at δ3.75x101.26 confirmed the (1→3) linkage between S-6 and S-5. This (1→3) linkage between S-6 and S-5 was supported by the presence of H-3 signal of β-GalNAc (S-5) at δ3.75 in upfield region of 1H NMR spectrum of Tarose acetate which was confirmed by the TOCSY and COSY spectrum. The anomeric carbon signal present at δ101.26 gave its complimentary signal at δ4.41 in HSQC spectrum of Tarose acetate in CDCl3 at 300 MHz. The chemical shift value of anomeric carbon and anomeric proton were having resemblance with literature value of anomeric chemical shift value of β-GalNAc (S-6) hence S-6 monosaccharide was confirmed as β-GalNAc (S-6). Further the presence of β-GalNAc (S-6) as next monosaccharide in Tarose was supported by appearance of anomeric proton signal at δ4.44 along with a singlet of three protons of amide methyl at δ1.90 in 1H NMR spectrum of Tarose in D2O.

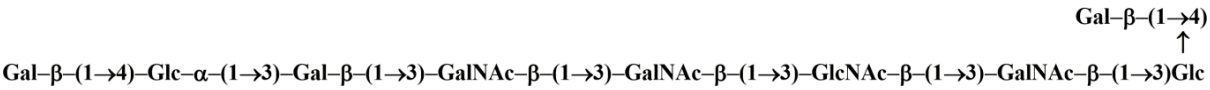
The coupling constant of anomeric signal β-GalNAc (S-6) with J value of 7.5 Hz confirmed the β glycosidic linkage between S-6 and S-5. The (1→3) linkage between S-6 and S-5 was supported by the presence of cross peak at δ3.75x76.48 in glycosidic region of HSQC spectrum of Tarose acetate in CDCl3. The (1→3) linkage between β-GalNAc (S-6) and β- GalNAc (S-5) was further confirmed by the 1H NMR spectrum of acetylated Tarose in which the signal for H-3 of ?-GalNAc (S-5) appeared at δ3.75 which was later confirmed by COSY, TOCSY and HSQC spectrum of acetylated Tarose at 300 MHz in CDCl3. The anomeric proton signal present at δ4.41 in 1H NMR spectrum of Tarose acetate in CDCl3 at 300 MHz. assigned for β-GalNAc (S-6), contain four cross peaks at δ4.41x3.75, δ4.41x4.12, δ4.41x5.05 and δ4.41x5.35 in its TOCSY spectrum, out of which one cross peak arised at δ4.41x4.12 corresponded to H-2 position of β-GalNAc (S-6) while the other signal arised at δ4.41x3.75 corresponded to linkage region of β-GalNAc (S-6) which was later identified as H-3 of β-GalNAc (S-6) by COSY spectrum of Tarose acetate which was available for (1→3) glycosidic linkage by the next monosaccharide unit (S-7). Further anomeric proton doublet appeared at δ4.59 in1H NMR spectrum of Tarose acetate in CDCl3 gave its complimentary anomeric carbon signal at δ95.53 in the HSQC spectrum of Tarose acetate. The chemical shift value of anomeric carbon and anomeric proton were having resemblance with literature value of anomeric chemical shift value of β-Gal (S-7) hence S-7 monosaccharide was confirmed as β-Gal (S-7).Further the presence of β-Gal (S-7) as next monosaccharide in Tarose was supported by appearance of anomeric proton signal at δ4.66 (J=7.8Hz) in 1H NMR spectrum of Tarose in D2O at 300 MHz. The anomeric proton signal present at δ4.66 for β-Gal (S-7) had a J value of 7.8Hz confirmed the β glycosidic linkage between S-7→S- 6.Since it was ascertained by the COSY and TOCSY spectrum of Tarose acetate that the β- GalNAc (S-6) has two vacant position for substitution i.e. H-2 and H-3 and it was already confirmed that H-2 of S-

6 was occupied with –NHAc of S-6 hence the left over H-3 position of S-6 at δ3.75 must be linked to β-Gal (S-7. The (1→3) linkage between β-GalNAc (S-6) and β-Gal (S-7) was supported by the presence of H-3 signal of S-6 at δ3.75 in upfield region of 1H NMR spectrum of Tarose acetate which was confirmed by the TOCSY and COSY spectrum. Further the (1→3) linkage between S-7 and S-6 was supported by the presence of cross peak at δ3.75x76.48 in glycosidic region of HSQC spectrum of Tarose acetate in $CDCl_3$. The anomeric proton signal at δ4.59 assigned for β-Gal (S-7), gave two consequent complimentary anomeric signal at δ4.59x3.73 and δ4.59x4.68 in its TOCSY spectrum, out of which cross peak arised at δ4.59x4.68 corresponded to H-2 position of β-Gal (S-7), however the cross peak arised at δ4.59x3.73 corresponded to linkage region of β-Gal (S-7), which was later defined as H-3 of β-Gal (S-7), by COSY spectrum of Tarose acetate which was available for (1→3) glycosidic linkage by the next monosaccharide unit (S-8). The next anomeric proton doublet which appeared at δ5.28 in the 1H NMR spectrum of Tarose acetate in $CDCl_3$ gave its complimentary signal δ90.30 in HSQC spectrum of Tarose acetate. The chemical shift value of anomeric proton and carbon were having resemblance with literature value of anomeric chemical shift value of α-Glc.

Further the presence of *α*-Glc (S-8) as next monosaccharide was supported by anomeric proton doublet at δ5.21 (J=3.6Hz) in the ^1H NMR spectrum of Tarose in D_2O at 300 MHz. Since it was ascertained by COSY and TOCSY spectrum of Tarose acetate that H-3 of S-7 was available for glycosidic linkage by next monosaccharide i.e. (S-8) hence S-8 (α-Glc) must be linked to H-3 of *β*-Gal (S-7). The (1→3) linkage between S-8 and S-7 was confirmed by the presence of cross peak at δ3.75x76.42 in glycosidic region of HSQC spectrum of Tarose acetate in $CDCl_3$. Further (1→3) linkage between S-8 and S-7 was supported by presence of downfield H-3 signal of S-7 at δ3.75 in ^1H NMR spectrum of Tarose acetate in $CDCl_3$. The coupling constant of anomeric signal (S-8) with J value 3.6Hz confirmed the *α*-configuration of the S-8 moiety with *β*-Gal (S-7). The anomeric proton signal at δ5.28 assigned for *α*-Glc (S-8), gave four cross peaks at δ5.28x3.73, δ5.28x4.75, δ5.28x4.9 and δ5.28x5.46 in its TOCSY spectrum, out of which one cross peak arised at δ5.28x3.73 corresponded to linkage region of *α*-Glc (S- 8), which was identified as H-4 of *α*-Glc (S-8) by COSY spectrum of Tarose acetate which was available for (1→4) glycosidic linkage by the next monosaccharide unit (S-9). The next anomeric proton signal which appeared as doublet at δ4.49 in ^1H NMR spectrum of Tarose acetate in $CDCl_3$ gave its complimentary signal at δ102.16 in HSQC spectrum of Tarose acetate. The chemical shift value of anomeric proton and carbon were having resemblance with literature value of anomeric chemical shift value of *β*-Gal (S-9). Hence (S-9) monosaccharide was confirmed as *β*-Gal. Further the presence of *β*-Gal (S-9) as next monosaccharide was supported by anomeric proton doublet at δ4.51 (J=8.1Hz) in the ^1H NMR spectrum of Tarose in D_2O at 300 MHz. Since it was ascertained by COSY and TOCSY spectrum of Tarose acetate that H-4 of S-8 was available for glycosidic linkage by next monosaccharide i.e. (S-9) hence S-9 *β*-Gal must be linked to H-4 of *α*-Glc (S-8).The (1→4) linkage between *α*-Glc (S-8) and *β*-Gal (S-9) was further supported by the ^1HNMR spectrum of Tarose acetate in which the signal for H-4 of *α*-Glc (S-8) appeared at δ3.73 which was already confirmed by COSY, TOCSY and HSQC spectrum of acetylated Tarose at 300 MHz in $CDCl_3$.

The (1→4) linkage between S-9 and S-8 was further confirmed by the presence of cross peak at δ3.73x76.48 in glycosidic region of HSQC spectrum of Tarose acetate in $CDCl_3$. The anomeric proton signal at δ4.49 in ^1H NMR spectrum of Tarose acetate in $CDCl_3$ assigned for *β*-Gal (S-9) gave three cross peaks at δ4.49x4.97, δ4.49x5.15 and δ4.49x5.28 in its TOCSY spectrum since this anomeric proton does not showed any cross peak in the linkage region, hence confirmed that *β*-Gal (S-9) was present at non-reducing end and none of its -OH group were involved in glycosidic linkage, which was confirmed by the TOCSY and COSY spectrum of Tarose acetate in $CDCl_3$ at 300 MHz. All the ^1H NMR assignments for ring protons of monosaccharide units of Tarose were confirmed by COSY and TOCSY experiments. The positions of glycosidation in the oligosaccharide were confirmed by position of anomeric signals, S.R.G. and comparing the signals in ^1H and ^{13}C NMR of acetylated and deacetylated oligosaccharide. The glycosidic linkages in Tarose were assigned by the cross peaks for glycosidically linked carbons with their protons in the HSQC and HMBC spectrum of acetylated Tarose. All signals obtained in ^1H and ^{13}C NMR of compound Tarose were in confirmity with the assigned structure and their position were confirmed by 2D NMR viz. COSY, TOCSY, HSQC and HMBC experiments. Thus based on the pattern of chemical shifts of ^1H NMR, ^{13}CNMR, COSY, TOCSY, HSQC and HMBC experiments it was interpreted that the compound was a nonasaccharide having followingstructure as:

TAROSE



The Electronspray Mass Spectrometry data of Tarose not only confirmed the derived structure but also supported the sequence of monosaccharide in Tarose. The highest mass ion peaks were recorded at m/z 1640 which was due to $[M^+]$ confirming the molecular weight of Tarose was 1640 and in agreement with its molecular formula. Further the mass fragments were formed by repeated H transfer in the oligosaccharide and was accompanied by the elimination of terminal sugar less water. The nonasaccharide m/z 1640 (I) fragmented to give mass ion at m/z 1478 (II) [1640-S-9], this fragment was arised due to the loss of terminal *β* - Gal (S-9) moiety from nonasaccharide indicating the presence of *β*-Gal (S-9) at the non- reducing end. It further fragmented to give mass ion peak at m/z 1316 (III) [1478-S-8] which was due to loss of *α*-Glc (S-8) moiety from octasaccharide. This fragment of 1316 further fragmented to give mass ion peak at m/z 1154(IV) [1316-S-7] which was due to loss of *β*-Gal (S-7) moiety from the heptasaccharide. This fragment of m/z 1154(IV) further fragmented to give mass ion peak at m/z 951(V) [1154-S-6], which was due to loss of *β*-GalNAc (S-6) moiety from hexasaccharide. This pentasaccharide fragment of m/z 951 further fragmented to give mass ion peak at m/z 748(VI) [951-S-5] which was due to loss of *β*-GalNAc (S-5) moiety. The tetrasaccharide unit of m/z 748(VI) again fragmented to give mass ion peak at m/z 545(VII) [748-S-4], which was due to loss of *β*-GlcNAc (S-4) moiety**.** This fragment of m/z 545 (VII) further fragmented to give mass ion at m/z 342 (VIII) [545-S-3],

this fragment was arised due to the loss of *β*-GalNAc (S-3) moiety from trisaccharide indicating the presence of *β*-GalNAc (S-3) It further fragmented to give mass ion peak at m/z 180(IX) [342-S-2] which was due to loss of *β*-Gal (S-2) moiety from disaccharide. The other fragmentation pathway in ES Mass spectrum of compound B m/z 1640 shows the mass ion peak at 1608 [1640-CH_3OH], 1581 [1640-CH_2OCHO], 1539 [1581-CH_2CO], 1489 [1539-CH_3OH,- H_2O], 1447 [1489-CH_2CO], 1478 [1640-S-9], 1446 [1478-CH_3OH], 1419 [1478-CH_2OCHO], 1316 [1478-S-8], 1273 [1316-CH_2CHO], 1254 [1273-H_3O^+], 1154 [1316-S-7], 1125 [1154-CHO], 1083 [1125-CH_2CO], 1053 [1154-CH_2OHCHO], 951 [1154-S-6], 919 [951-CH_3OH], 875 [951-CH_2OCHO,OH], 841 [875-2OH], 824 [841-OH], 748 [951-S-5], 706 [748-CH_2CO], 705 [748-CH_2CHO], 677 [706-CHO], 646 [677-CH_2OH], 645 [646-H^+], 545 [748-S-4], 481 [545-CHO,-OH,-H_2O], 467 [545-CH_2OHCHO,-H_2O], 466 [545-HCHO,CH_2OH,H_2O], 465 [466-H^+], 406 [466-2HCHO], 342 [545-S-3], 324 [342-H_2O] and 180 [342-S-2].

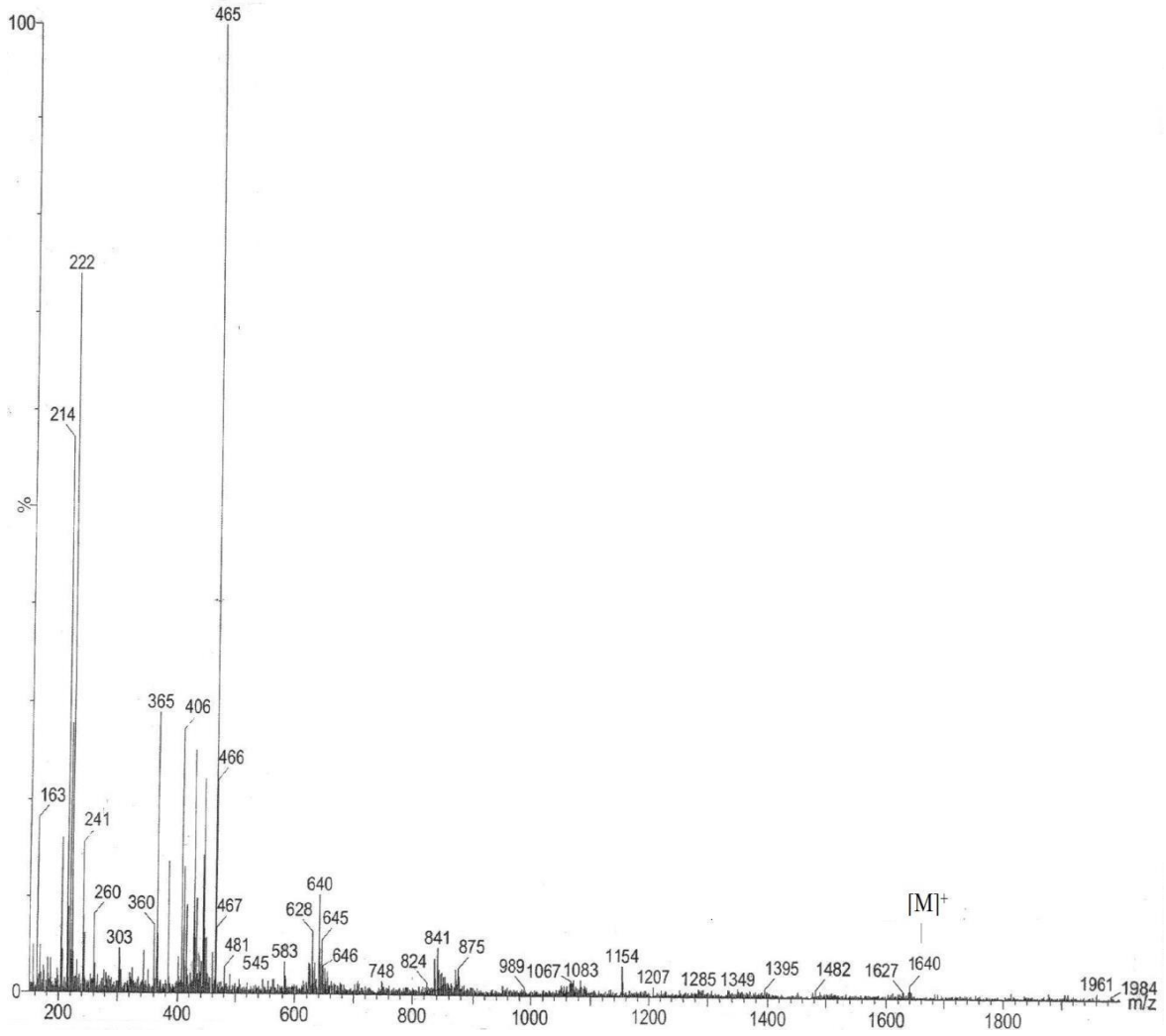
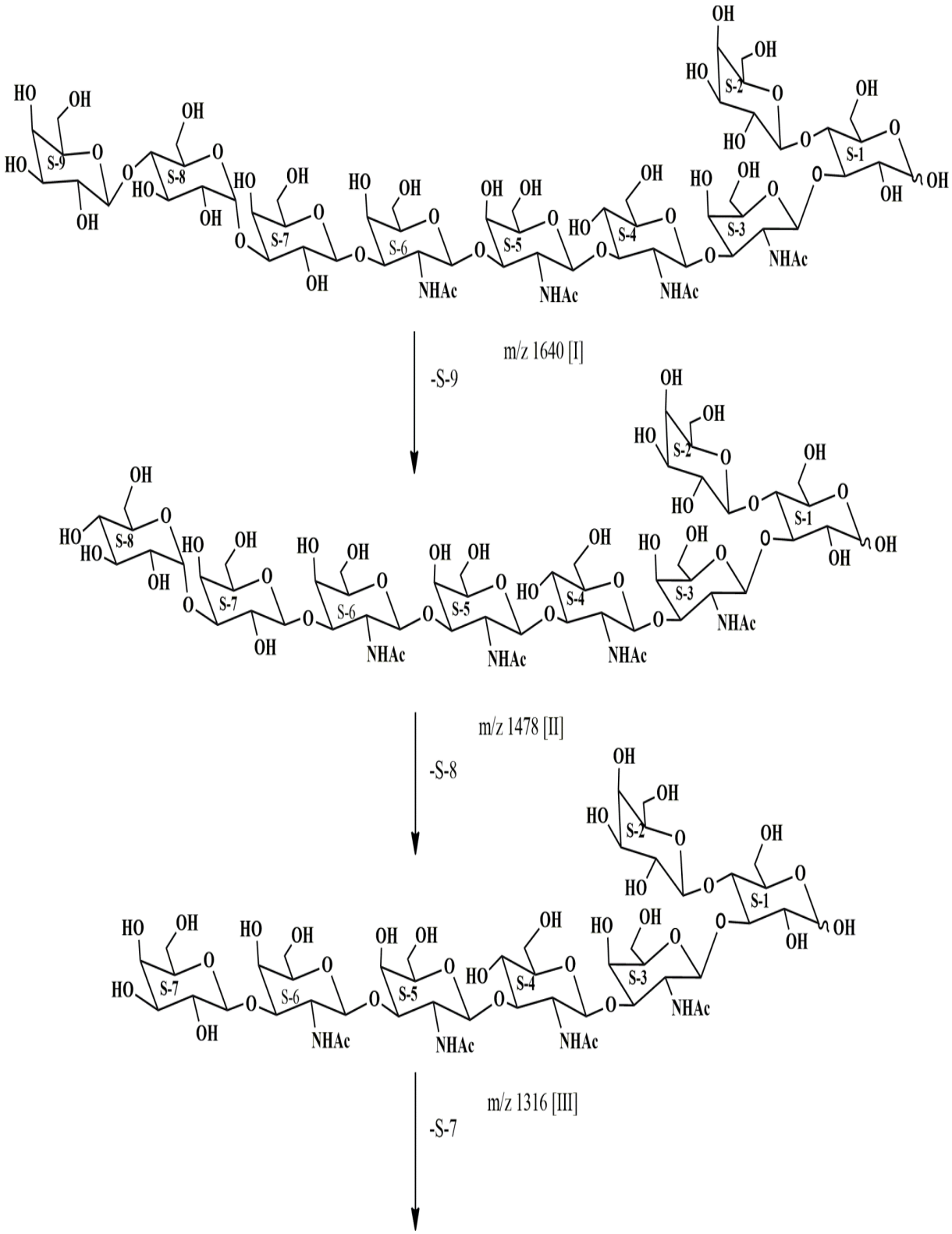
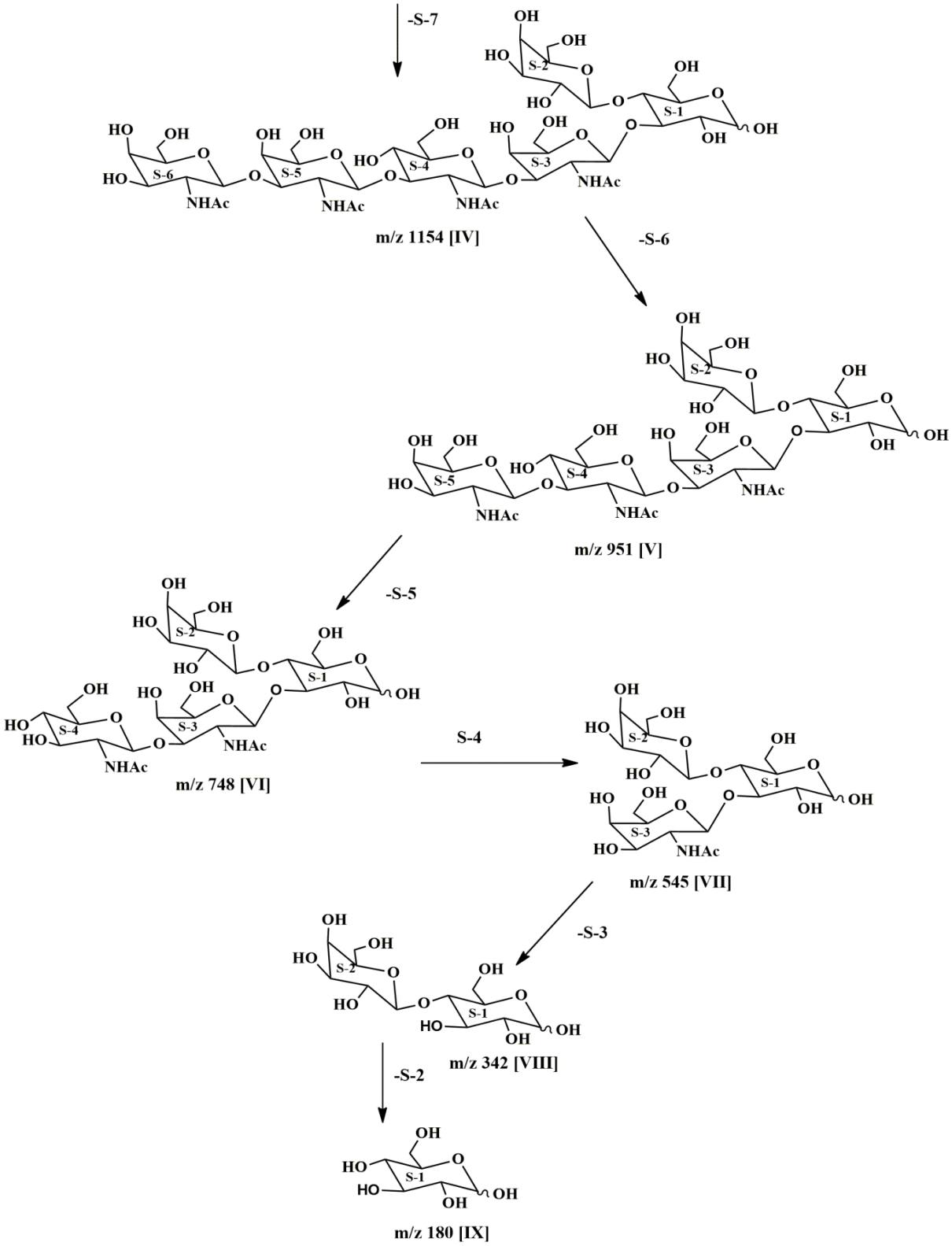


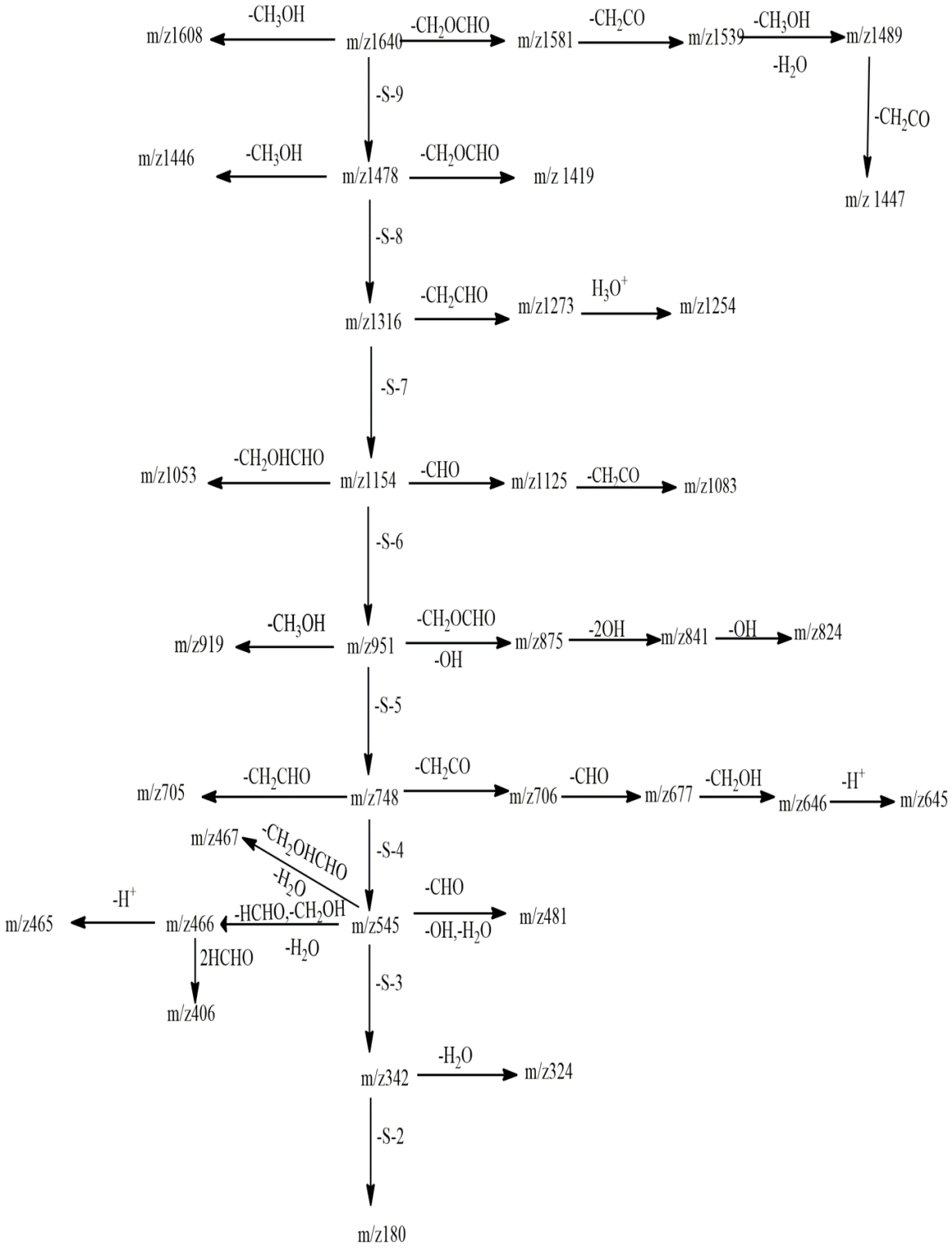
Figure: ES-MS Spectrum of compound B, Tarose.



Scheme 2.2: Mass fragmentation of compound B, Tarose.

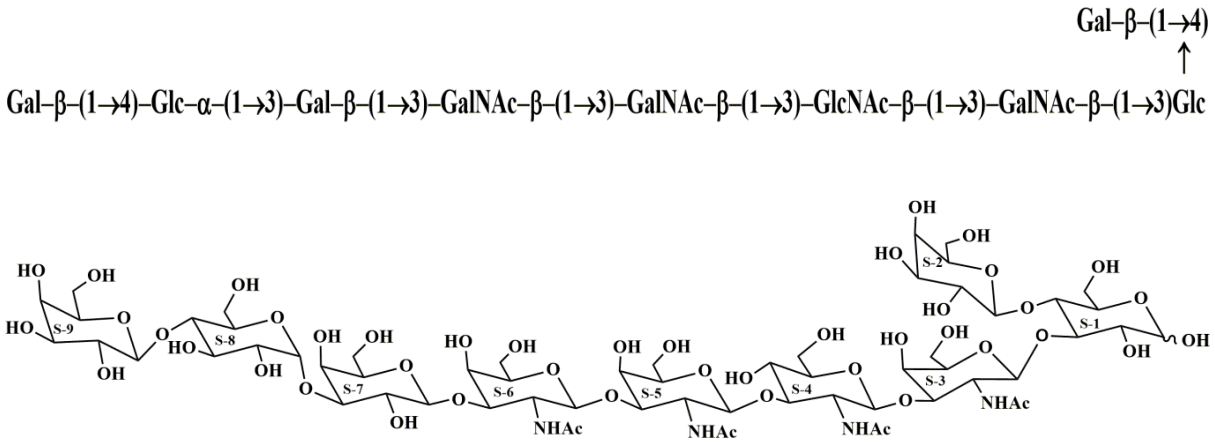


Scheme 2.2: Mass fragmentation of compound B, Tarose.



Scheme 2.3: ES-MS fragmentation of compound B, Tarose

Based on result obtained from chemical degradation/acid hydrolysis, Chemical transformation, Electro spray mass spectrometry and 1D NMR viz. ^{1}H NMR, ^{13}C NMR and 2D NMR viz. COSY, TOCSY, HMBC and HSQC spectra of Tarose acetate and Tarose, the structure and sequence of isolated Novel oligosaccharide Tarose structure was deduced as:



TAROSE

CONCLUSION

A novel nonasaccharide Tarose were isolated from cow colostrum and its structure was elucidated with the help of chemical degredation, chemical transformation and physico chemical techniques ie. 1H, 13C, and 2D NMR experiments (COSY, TOCSY, HSQC, and HMBC) and Mass Spectrometry.

ACKNOWLEDGMENTS

The authors are thankful to the Head, Department of Chemistry, University of Lucknow, for providing lab, NMR, and computational facilities.

REFERENCES

Sharma, M., Shukla M. and Deepak D. (2022). Structures of biologically active milk oligosaccharides isolated from various indigenous cow species. *JBCR*, Vol 39 (2):146-157.

Gunjan, Kumar K., and Deepak D. (2019). Structural characterization of novel milk oligosaccharide Aurose from cow colostrum. *Journal of molecular structure*, 1176: 394-401.

Deepak, D., Gangwar, L. and Sharma, M. (2020). Isolation and Structure Elucidation of a Novel Tetrasaccharide Asose from Shyamadhenu's Milk. *Trends in carbohydrate chemistry*, vol 12 (4): 64-76.

Singh, A.K. and Deepak, D. (2020). Isolation and Structure Elucidation of Novel Yak Milk Oligosaccharide 'Nienose'. *Trends in carbohydrate chemistry*, vol 12 (4): 15-26.

Kumar, K., Verma, A., Shukla, M., Mishra, A. and Deepak, D. (2022). Isolation and structure elucidation of novel oligosaccharide (gagriose) from goat milk by NMR and Mass spectrometry. *JBCR*, vol 39(2): 99-111.

Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A. and Smith, F. (1956). Colorimetric Method for Determination of Sugars and Related Substances. *Anal. Chem* 28: 350-356.

Dua, V.K. and Bush, C.A. (1983). Identification and fractionation of human milk oligosaccharides by proton-nuclear magnetic resonance spectroscopy and reverse-phase high- performance liquid chromatography. *Anal. Biochem.,* 133: 1-8.

Chaturvedi, P. and Sharma, C.B. (1988). Goat milk oligosaccharides: purification and characterization by HPLC and high-field 1H-NMR spectroscopy. *Biochemica et Biophysica Acta*, 967: 115-121.

Uemura, Y., Asakuma, S., Nakamura, T., Arai, I., Taki, M. and Urashima, T. (2005). Occurrence of a unique sialyl tetrasaccharide in colostrum of a bottlenose dolphin (*Tursiops truncatus*). *Biochim. Biophys. Acta,* 1725: 290-297.

Singh, M., Kumar, A., Srivastava, G., Deepak, D., Singh, M.P.V.V., Singh, M. and Deepak, D. (2016). Isolation, structure elucidation and DFT study on two novel oligosaccharides from yak milk. *Journal of Molecular structure*, 1117: 69-78.

Singh, A.K., Ranjan, A.K., Srivastava, G., and Deepak, D. (2016). Structure elucidation of two novel yak milk oligosaccharides and their DFT studies. *Journal of molecular structure*, vol 1108: 87-91.

Corresponding author: Dr. Desh Deepak, University of Lucknow, Lucknow Uttar Pradesh, 226007, U.P., India
Email: deshdeepakraju@rediffmail.com